



جامعة أبو بكر بلقايد - تلمسان

Université Abou Bakr Belkaïd de Tlemcen

Faculté de Technologie

Département de Génie Biomédical

**MEMOIRE DE PROJET DE FIN D'ETUDES**

pour l'obtention du Diplôme de

**MASTER en GENIE BIOMEDICAL**

*Spécialité : Instrumentation Biomédicale*

Présenté par : **NAIM Marwa et DJADOUDI Muhammad Sobir**

---

**Réalisation d'un dispositif d'acquisition,  
d'analyse et classification de pleurs de bébés**

---

**Présenté devant le jury**

M.	BEREKSIREGUIG Fethi	<i>Prof</i>	Université de Tlemcen	Président
Melle	HABIBES Naïma	<i>MAA</i>	Université de Tlemcen	Encadreur
Melle	MEZIANI Fadia	<i>MCA</i>	Université de Tlemcen	Examineur

**Année universitaire 2020-2021**

## Dédicace

Je dédié ce modeste travail

A **mes chers parents** pour tous les sacrifices consentis pour me voir devenir ce que je suis aujourd'hui.

**Ma mère**, l'école de la vie, la source de tendresse et d'amour, la chose la plus précieuse de l'univers qui m'a bénie avec ses prières tout au long de mes études.

**Mon père**, l'épaule solide, mon premier exemple et mon guide qui éclaire mon chemin, celui qui m'a appris à résister aux vagues déchaînées de la mer, à celui qui m'a donné et continué de me donner sans limites.

Puisse Dieu, le tout puissant, vous protège de tout mal et vous accorde santé, longue vie et bonheur.

À **mes chères sœurs**, pour leurs précieux conseils, leurs soutiens et encouragements ; Je vous souhaite tout le bonheur, la santé et la réussite.

A l'âme de **ma grand-mère**, que dieu l'accueille dans son vaste paradis.

A toute **ma famille**.

A tous **mes collègues** et **mes amis** de près ou de loin.

À **mon partenaire** de projet Muhammad pour tout le soutien, la motivation et pour son éthique de travail et son inspiration.

**Marwa NAIM**

## Dédicace

À mes cher parents et toute ma famille

À tous mes amis

À mon partenaire de projet

**Merci pour tous.**

**Muhammad Sobir DJADOUDI**

## Remerciement

Avant tout, on remercie Allah le tout puissant et miséricordieux, qui nous a donné la force et la patience pour accomplir ce projet.

Nous tenons à remercier Mme **HABIBES NAIMA** d'avoir accepté de nous encadrer, pour la qualité de son encadrement, sa patience, ses précieux conseils et ses encouragements qui ont contribué efficacement à l'avancement de ce travail.

Nous exprimons nos sincères remerciements à **Mr BRIKSI REGUIG. F** pour l'honneur qu'il nous fait en acceptant de présider le jury de soutenance.

Nous remercions également Mme **MEZIANI Fadia** pour l'intérêt qu'il a porté à ce travail en acceptant de l'examiner.

Nous remercions également tous les professeurs et les enseignants qui nous ont apporté leurs connaissances et leurs soutiens durant toute notre formation.

Enfin, nous ne manquerons pas d'adresser nos remerciements à toutes les personnes qui ont contribué de près ou de loin à l'exécution de ce modeste projet.

## Table des matières

<b>Remerciement.....</b>	<b>3</b>
<b>Liste des figures.....</b>	<b>7</b>
<b>Liste des tableaux.....</b>	<b>8</b>
<b>Abréviations.....</b>	<b>9</b>
<b>Résumé.....</b>	<b>12</b>
<b>Introduction générale .....</b>	<b>14</b>
<b>Chapitre I : Etat d’art .....</b>	<b>15</b>
1. Introduction .....	15
2. Les travaux effectués .....	15
2.1. Détection des pleurs de bébé dans un environnement domestique grâce à l'apprentissage profond.....	15
2.2. Chatter Baby .....	17
2.3. Les méthodes de reconnaissance des pleurs de bébé et surveillance intelligente basée sur DNN-GMM-HMM .....	19
2.4. Berceau intelligent pour bébé utilisant le module FN-M16P.....	20
2.5. Système intelligent de détection des pleurs des bébés avec Raspberry Pi .....	22
3. Problématique.....	24
4. Méthode .....	24
5. Conclusion .....	25
<b>Chapitre II : Généralité.....</b>	<b>26</b>
1. Introduction .....	26
2. Définition du son .....	26
3. Les caractéristiques du son.....	27
3.1. La période .....	28
3.2. La fréquence.....	28
3.3. La longueur d’onde .....	28
3.4. L’amplitude.....	29
3.5. La phase .....	29
3.6. La célérité.....	30

3.7. Le spectre .....	30
4. Les types de son.....	30
4.1. Les sons purs ou sons simples.....	30
4.2. Les sons complexes.....	31
5. Les propriétés du son .....	31
5.1. L'impédance (Z) .....	31
5.2. La pression acoustique (P).....	32
5.3. La puissance acoustique (W) .....	32
5.4. Intensité sonore (I) .....	32
6. La voix humaine .....	32
7. Les pleurs de bébé .....	33
8. L'origine et le rôle du pleur du bébé .....	33
9. Les caractéristiques de signal pleurs des bébés .....	34
10. Les différences entre le cri d'un bébé et un enfant .....	34
11. Anatomie de l'oreille .....	35
11.1. L'oreille externe.....	35
11.2. Oreille moyenne.....	35
11.3. Oreille interne .....	35
12. Le mécanisme de l'audition.....	36
13. La surdité .....	36
13.1. Les types de la surdité.....	37
14. Impact de la perte auditive.....	37
14.1. Impact fonctionnel .....	37
14.2. Impact social et émotionnel .....	38
14.3. Impact économique .....	38
15. Principaux faits .....	38
16. Conclusion .....	39
<b>Chapitre III : méthodologie .....</b>	<b>40</b>
1. Introduction .....	40
2. Hardware .....	41
3. Logiciels utilisés (Software).....	41
3.1. Anaconda Navigator .....	41

3.2. Langage python.....	43
3.3. App Inventor .....	44
3.4. Windows on Raspberry.....	44
4. Mel-Frequency Cepstral Coefficients (MFCC).....	45
5. Apprentissage profond (Deep learning) .....	48
6. Les réseaux de neurones convolutifs .....	48
6.1. Les couches convolutives.....	49
6.2. La fonction d'activation ReLU .....	50
6.3. La couche pooling.....	50
6.4. La couche d'aplatissement.....	50
6.5. La couche entièrement connectée .....	50
7. La précision .....	51
8. L'optimisation Adam.....	51
<b>Chapitre IV : réalisation et implémentation .....</b>	<b>53</b>
1. Introduction .....	53
2. Architecture du back-end de logiciel.....	53
3. Première partie de projet.....	54
3.1. Etape 01 : Préparation de base de données .....	54
3.2. Etape 02 : Création d'un modèle CNN .....	56
3.3. Etape 03 : l'apprentissage pour le model .....	57
3.4. Etape 04 : Enregistrement de son.....	58
3.5. Etape 05 : Extraction des paramètres MFCC et conversion en jsonfile.....	59
3.6. Etape 06 : Prédiction et envoi du résultat de la classification au smartphone.....	60
4. Deuxième partie du projet .....	61
4.1. Initialisation et configuration du Raspberry pi.....	61
4.2. Application « Baby Crying Classifier » .....	62
5. Conclusion .....	64
<b>Conclusion générale .....</b>	<b>65</b>
<b>Références .....</b>	<b>66</b>

## Liste des figures

Figure 1: L'architecture CNN et un schéma fonctionnel de l'algorithme de régression logistique (3).....	17
Figure 2: logo de Chatter Baby (4) .....	18
Figure 3: Organigramme de l'algorithme de reconnaissance des pleurs (5) .....	20
Figure 4: schéma bloc (6) .....	21
Figure 5: organigramme du projet (7).....	23
Figure 6: schéma bloc .....	25
Figure 7: diagramme d'onde sonore.....	27
Figure 8: onde périodique (9) .....	28
Figure 9: un signal représentant la longueur d'onde (13).....	29
Figure 10: un signal représentant le son pur (16) .....	30
Figure 11: un signal représentant le son complexe (16) .....	31
Figure 12: Parole d'adulte vs. Pleurs de bébé dans le domaine temporelle et fréquentielle (22)..	35
Figure 13: Schéma des principales parties du système auditif périphérique .....	36
Figure 14: Schéma bloc générale.....	40
Figure 15: interface de logiciel ANACONDA NAVIGATOR.....	42
Figure 16: logo de python .....	43
Figure 17: logo de app Inventor.....	44
Figure 18: interface de logiciel "Windows on Raspberry" .....	45
Figure 19: diagramme représente le processus MFCC .....	45
Figure 20: diagramme de réseaux de neurones convolutif .....	48
Figure 21 : représentation matricielle du produit de convolution (49) .....	49
Figure 22 : graphe de filtre d'activation ReLU (52) .....	50
Figure 23: organigramme de back-end de "Baby Crying Classifier" .....	53
Figure 24: fichier json.....	54
Figure 25: paramétrage du modèle de CNN .....	56
Figure 26: graphe d'erreur et de précision .....	58
Figure 27: représentation de signal enregistré .....	59
Figure 28: spectrogramme de signal enregistré .....	60
Figure 29: le système hardware du projet.....	61
Figure 30: interface principale de l'application mobile.....	62
Figure 31: deuxième interface de l'application mobile .....	63



## Liste des tableaux

Tableau 1: définition des bibliothèques utilisés.....	44
Tableau 2: la précision avant et après l'optimisation .....	52
Tableau 3 : présentation de spectrogramme et MFCC pour chaque classe .....	56
Tableau 4: définition des termes de paramétrage .....	57

## Abréviations

**CAE** : conduit auditif externe

**CNN** : réseaux neuronal convolutif

**CWT** : transformée en Ondelette Continu

**DCT** : transformée en cosinus discrète

**DFT** : transformée de Fouries discrète

**DNN-GMM-HMM** : Deep Neural Network Gaussian Mixture Model Hidden Markov Model

**FFT** : transformation de Fourier rapide

**GSM** : système mondial de communication avec les mobiles

**GUI** : interface graphique

**HDMI** : interface multimédia haute définition

**IA** : intelligence artificielle

**LCD** : Affichage à Cristaux liquides

**LMFB** : logarithmique de banque de filtres de Mel

**LOT** : internet des objets

**MFCC** : Mel Frequency Cepstral Coefficients

**NN** : réseau neuronal

**OMS** : Organisation Mondiale de la santé

**OS** : système d'exploitation

**PCA** : L'analyse en composantes principales

**RAM** : mémoire vive, random access memory

**ROM** : mémoire à lecture seule ou mémoire mort

**SD** : Secure Digital

**SVM** : machine à vecteurs de support

**UCLA** : Université de Californie à Los Angeles

**VDA** : détection d'activité vocale

**Wifi** : fidélité sans fil



## Résumé

Le nourrisson exprime sa condition aux parents en pleurant. Ces pleurs se produisent à cause de différents problèmes : de la douleur, de l'inconfort, de la faim, etc.

L'objectif de ce travail est de réaliser un système intelligent permettant de faciliter la communication entre les nouveaux nés et leurs parents afin d'avoir une vie saine.

L'étape principale de notre réalisation de ce projet est l'acquisition des signaux de pleurs de bébé et ceci dans différents cas de pleures. Cette tâche est réalisée à l'aide d'un microphone USB connecté à une carte Raspberry PI 3. Cette étape est suivie d'un traitement du signal par la méthode MFCC (Mel-Frequency Cepstral Coefficients) pour extraire les paramètres utiles et les coefficients caractérisant les pleurs de bébés, ces caractéristiques aident à différencier les types des signaux enregistrés (cause de pleure de bébés). Par la suite, l'étape de classification par les réseaux neuronaux convolutifs (CNN) qui permet de classer les pleurs de bébé en quatre catégories : la douleur, l'inconfort, la faim et le silence.

Les résultats obtenus sont envoyés par SMS en utilisant un compte Twilio et affichés via l'application « Baby Crying Classifier » installée sur le téléphone portable, afin de notifier les parents ou les baby-sitters sur l'état de son bébé.

**Les mots clés :** pleurs de bébé, Raspberry pi, traitement du signal, apprentissage profond, python, MFCC, réseau neuronal convolutif (CNN).

## Abstract

The infant expresses his condition to their parent by crying. This crying occurs because of different problems: pain, discomfort, hunger, etc.

The goal of this work is to realize an intelligent system to facilitate communication between newborns and their parents in order to have a healthy life.

The main step in our realization of this project is the acquisition of baby crying signals and this in different cases of crying. This task is performed using a USB microphone connected to a RaspberryPI 3 board. This step is followed by a signal processing step using MFCC method (Mel-Frequency Cepstral Coefficients) to extract the useful parameters and the coefficients characterizing the crying of babies, these characteristics help to differentiate the types of the recorded signals (cause of crying from babies). Next, the Convolutional Neural Network (CNN) classification step classifies baby crying into four categories: pain, discomfort, hunger and silence.

The obtained results are sent by SMS using a Twilio account and displayed via the "Baby Crying Classifier" application installed into the mobile phone, in order to notify parents or babysitters about the condition of the baby.

**Keywords:** baby crying, Raspberry pi, signal processing, deep learning, python, MFCC, convolutional neural network (CNN).

## ملخص

يقوم الرضيع بالتعبير عن حالته لوالديه بالبكاء. يحدث هذا البكاء بسبب مشاكل مختلفة: الألم، عدم الراحة، الجوع وما إلى ذلك.

الهدف من هذا العمل هو تحقيق نظام ذكي لتسهيل التواصل بين الأطفال حديثي الولادة و ابائهم من أجل التمتع بحياة صحية.

الخطوة الرئيسية في تحقيقنا لهذا المشروع هي الحصول على إشارات بكاء الأطفال وهذا في حالات مختلفة من البكاء. يتم تنفيذ هذه المهمة باستخدام ميكروفون USB متصل ببطاقة Raspberry PI 3. تتبع هذه الخطوة معالجة الإشارات بواسطة طريقة (MFCC) لاستخراج الخصائص المفيدة والمعاملات التي تميز بكاء الأطفال، وتساعد هذه الخصائص على التمييز بين أنواع الإشارات المسجلة (سبب بكاء الأطفال). بعد ذلك، نستعمل الشبكة العصبية التلافيفية (CNN) التي تصنف بكاء الطفل إلى أربع فئات وهي الألم وعدم الراحة والجوع والصمت.

يتم إرسال النتائج التي تم الحصول عليها عن طريق الرسائل القصيرة باستخدام حساب Twilio ويتم عرضها عبر تطبيق "Baby Crying Classifier" في الهاتف المحمول، وذلك لإخبار الآباء أو المربيات بحالة طفلهم.

**الكلمات المفتاحية:** بكاء الأطفال، Raspberry pi، التعلم العميق، بايثون، MFCC، الشبكة العصبية التلافيفية (CNN).

## Introduction générale

L'homme est confronté tout au long de sa vie à des problèmes qui l'obligent à trouver une solution. Les pleurs d'un enfant sont l'un de ces problèmes qui empêchent les parents de répondre à leurs besoins. Ces pleurs sont considérés comme un langage pour communiquer avec le monde.

Le cri d'un nourrisson est une sirène biologique pour alerter l'environnement de soins sur ses besoins pour motiver l'auditeur à réagir. La plupart du temps, les baby-sitters suivent des routines strictes pour former l'enfant à des habitudes régulières d'alimentation, d'éveil et de sommeil sans tenir compte de ses besoins émotionnels et physiques. Les chercheurs ont découvert que les bébés dont les pleurs sont généralement ignorés ne développeront pas de saines compétences intellectuelles et sociales. D'un autre côté, laisser un bébé en détresse pleurer régulièrement pourrait nuire au développement du cerveau. [1]

Au cours de la dernière décennie, les chercheurs ont développé des algorithmes basés sur l'analyse spectrale du signal qui permettent l'identification et la classification des pleurs des nourrissons à partir d'une base de données. Et ces dernières années, afin d'améliorer les résultats et l'efficacité, les chercheurs ont utilisé d'autres techniques plus avancées telle que le "Deep Learning" pour le traitement des signaux.

Le but de ce projet est de réaliser un système intelligent permettant de détecter les pleurs de bébé et prédire la cause de ces pleurs.

Nous avons organisé notre travail en quatre chapitres :

Dans le premier chapitre, nous avons présenté l'état de l'art des recherches effectuées dans le même domaine, la problématique et la solution proposée.

Le deuxième chapitre s'intéresse à mettre en évidence les notions de base du son, l'anatomie auditive et physiologie.

Les deux derniers chapitres sont consacrés à la description de notre conception proposée. Le troisième chapitre expose la méthodologie et les moyens utilisés dans le coté software. Le quatrième chapitre démontre l'implémentation des méthodes utilisées pour réaliser notre système.

Nous avons terminé par une conclusion générale et des perspectives.

# Chapitre I : Etat d'art

## 1. Introduction

Les pleurs d'un enfant sont l'un des problèmes les plus difficiles auxquels baby-sitter peut être confrontée lorsqu'elle s'occupe d'un enfant, surtout avant l'âge de 24 mois.

Vers 2 à 3 semaines, le bébé commence à pleurer davantage. Ses pleurs atteignent ensuite un sommet vers 6 à 8 semaines et puis diminuent peu à peu jusqu'à ce que l'enfant ait 12 semaines. Pendant ces trois premiers mois, les pleurs commencent et arrêtent souvent soudainement, sans que l'on sache pourquoi, particulièrement en fin d'après-midi ou en début de soirée. Ainsi, il peut arriver qu'un parent ne parvienne pas à calmer son bébé qui pleure, même en le nourrissant, en s'assurant que sa couche est propre ou en le cajolant. [2]

C'est ce qui a suscité l'intérêt des scientifiques et des chercheurs, qui ont mené de nombreuses recherches et des études (que ce soit au niveau de l'acquisition du signal ou des méthodes de sa caractérisation et de sa classification) qui améliorent la communication entre les parents et leurs enfants.

Le but de ce chapitre est de découvrir quelques-uns des travaux qui ont été mis en œuvre ces dernières années et le problème le plus courant auquel les parents sont confrontés. Enfin proposant les méthodes utilisées pour la mise en place de ce projet.

## 2. Les travaux effectués

Les travaux effectués sont basés principalement sur l'acquisition du signal vocal, le traitement du signal et les méthodes de classification.

### 2.1. Détection des pleurs de bébé dans un environnement domestique grâce à l'apprentissage profond

En 2016, Yizhar Lavner, Rami Coheny, Dima Ruinskiy et Hans IJzerman ont réalisé un système intelligent utilisant l'apprentissage profond. [3]



L'évaluation des performances des algorithmes dans cette recherche est réalisée à l'aide d'une base de données annotée contenant de plusieurs dizaines d'heures d'enregistrements audio réalisés par des parents de bébés aux Pays-Bas. Les bébés étaient dans leurs 6 premiers mois de vie, et les enregistrements ont été réalisés 24h/24 et 7j/7 dans un environnement domestique. Les enregistrements contiennent divers types de sons, tels que des pleurs, des paroles de parents, l'ouverture de porte, etc. La base de données a été collectée dans le cadre d'une étude pilote visant à étudier la « formation d'attachement » (formant le lien entre le soignant et l'enfant). Trois heures d'enregistrements ont été entièrement annotées, au milliseconde, avec environ 50 types d'événements différents. La fréquence d'échantillonnage des enregistrements est  $F_s = 44100$  Hz. [3]

Dans cet article, ils ont proposé deux algorithmes d'apprentissage automatique pour la détection automatique des pleurs de bébé. [3]

Le premier algorithme est un classificateur de régression logistique de faible complexité, utilisé comme référence. Différents descripteurs ont été utilisés pour extraire des caractéristiques telles que les coefficients de cepstre de fréquence Mel, la hauteur et les formants des enregistrements. [3]

Un schéma fonctionnel de l'algorithme basé sur la régression logistique est représenté sur la figure 1. Les données d'entrée sont divisées en segments consécutifs de 4096 échantillons. Pour chaque segment, un vecteur de caractéristiques à 50 dimensions est calculé. La régression logistique régularisée entraînée est ensuite appliquée sur chaque vecteur de caractéristiques, et la fonction d'hypothèse  $h(x)$  est obtenue. Si la fonction  $h(x)$  est supérieure d'une valeur seuil, le résultat est 1, cela signifie que le bébé est pleuré. Et si le résultat est 0, cela signifie que le bébé ne pleure pas. [3]

Le deuxième algorithme utilise un réseau de neurones convolutifs (CNN) dédié, fonctionnant sur une représentation de banque de filtres de Mel. L'architecture de cet algorithme est présentée dans la figure 1. [3]

Pour travailler avec un classificateur CNN, le signal audio est divisé en segments consécutifs de 4096 échantillons. Chaque segment est ensuite divisé en trames de 512 échantillons, avec une taille de pas de 128 échantillons. La contribution des bandes hautes fréquences à la détection des signaux cri étant limitée, un filtre passe-bas à 11025 Hz est appliqué. Une représentation logarithmique de banque de filtres de Mel (LMFB) est ensuite produite pour chaque trame, à l'aide de 40 filtres répartis selon l'échelle de Mel dans la gamme de fréquences [0 ; 11025] Hz. Étant donné des segments de 4096 échantillons et une taille de pas de 128 échantillons, cela conduit à une représentation « image ». [3]

Le CNN utilise des couches de convolution avec des filtres « tall » :  $10 \times 2$ ,  $6 \times 2$  et  $3 \times 2$ , pour atteindre une résolution haute fréquence par rapport à une basse résolution temporelle. [3]

Chaque couche convolutée est suivie par une couche ReLU, une couche max pooling, une couche entièrement connectée et un Softmax. [3]

En raison de l'échelle Mel, chaque « pixel » dans le LMFB représente une plage de fréquences. [3]

Le classificateur CNN donne des résultats considérablement meilleurs par rapport au classificateur de régression logistique, démontrant la puissance de l'apprentissage en profondeur lorsqu'il est appliqué au traitement audio. [3]

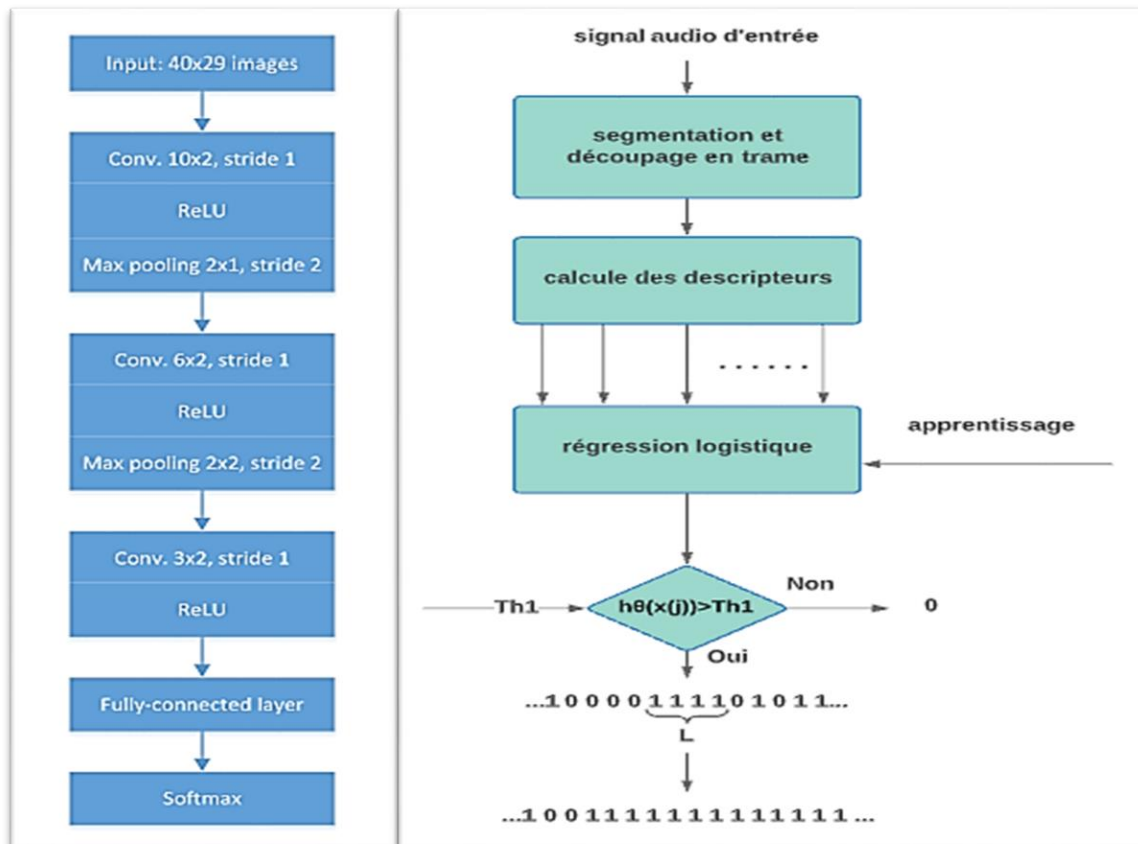


Figure 1: L'architecture CNN et un schéma fonctionnel de l'algorithme de régression logistique [3]

## 2.2. Chatter Baby

Chatterbaby est une application pour smartphone de l'UCLA (Université de Californie à Los Angeles) peut traduire le sens des pleurs d'un bébé. [4]

Initialement, cette application était destinée aux parents sourds pour identifier quand leur bébé pleure à distance, mais le problème ici est qu'elle ne fait pas la distinction entre un cri de détresse et un bébé qui parle fort. [4]

Le chercheur Arianna Anderson, et en tant que mère, a pensé à créer un algorithme qui fait ce que les oreilles d'un parent peuvent faire automatiquement c'est à dire découvrir ce que votre bébé essaie de vous dire en pleurant. Pour cela, l'équipe de l'UCLA, dirigée par ce chercheur a entrepris de développer un algorithme qui pourrait classer les différents cris en trois catégories : la douleur, la faim et l'anxiété. [4]

Après avoir testé une série d'algorithmes différents sur plus de 2 000 échantillons de cris de bébé, le système a finalement été en mesure d'identifier correctement les cris de douleur plus de 90 %, bien que le système ait plus de difficulté à faire la différence entre les cris d'agitation et les cris de faim. [4]

Tous les bébés ne communiquent pas exactement de la même manière, chaque bébé a un algorithme spécial qui lui permet de traduire ses cris. [4]

Des clips audios des pleurs d'un bébé peuvent être enregistrés, étiquetés et téléchargés sur le serveur de l'UCLA pour optimiser la précision des prédictions futures de l'application. Les chercheurs soulignent que ces données audios sont anonymisées de toute caractéristique d'identification individuelle, mais qu'elles sont stockées en permanence dans le cadre d'une étude de recherche plus vaste conçue pour améliorer lentement la précision globale de l'algorithme. [4]

L'équipe mène également des recherches pour diagnostiquer les troubles neurodéveloppementaux par le biais d'un discours anormal. [4]

Chatterbaby fait partie d'un projet de recherche plus vaste et gratuit à utiliser et ne nécessite que le consentement de l'utilisateur. Les chercheurs espèrent développer des algorithmes et élargir des catégories telles que l'autisme. [4]



Figure 2: logo de Chatter Baby [4]

### 2.3. Les méthodes de reconnaissance des pleurs de bébé et surveillance intelligente basée sur DNN-GMM-HMM

Cette recherche est faite en 2018 à l'université d'Anhui (Chine) par Chaoyuan Pan et al. [5]

Dans cette recherche le modèle de reconnaissance des pleurs de bébé est basé sur le modèle DNN-GMM-HMM (Deep Neural Network Gaussian Mixture Model Hidden Markov Model). [5]

Comme première étape, Avant de poursuivre le traitement des signaux collectés, cette équipe de recherche s'est basé sur le prétraitement des pleurs de bébé pour améliorer la précision de reconnaissance du système par la détection d'activité vocale (VDA) et la préaccentuation. [5]

La détection d'activité vocale (VDA) est souvent appelée détection de point de terminaison vocale. La préaccentuation fait référence à une méthode de traitement du signal qui compense la composante haute fréquence du signal d'entrée à l'extrémité émettrice. [5]

Afin d'améliorer encore la précision de la reconnaissance des pleurs de bébé, une approche «Bottleneck» devrait également être utilisée pour étendre le vecteur d'entrée dans le système GMM-HMM qui utilise des caractéristiques dérivées d'un ou plusieurs réseaux de neurones. [5]

L'analyse en composantes principales (PCA) peut être utilisée avant d'appliquer les fonctionnalités du système GMM-HMM pour réduire le surdimensionnement et pour compresser et réduire les fonctionnalités combinées pour identifier les pleurs de bébé. [5]

Afin de parvenir à terme à la reconnaissance des pleurs de bébé, le modèle acoustique établi par GMM-HMM est nécessaire pour décoder les caractéristiques compressées après réduction de la dimensionnalité. [5]

L'organigramme de la figure 3 résume les différentes étapes de l'algorithme de reconnaissance des pleurs.

Du côté pratique, cette recherche est basé principalement sur la technologie STM32 (une série de microcontrôleurs 32-bits en circuits intégrés réalisés par la société Franco-Italienne STMicroelectronics), DSP (un microprocesseur optimisé pour exécuter des applications de traitement numérique du signal le plus rapidement possible) , un module d'acquisition de fréquence cardiaque, un module d'identification des informations qu'est lié à l'importation du module de contrôle principal et consiste principalement en une acquisition, un traitement et une classification des pleurs ainsi qu'une unité d'entrée et de sortie, et le module de communication Bluetooth transmettra les données au module pilote d'affichage selon le format et le protocole de transfert. Enfin, l'émotion ou les besoins correspondants sont affichés. [5]

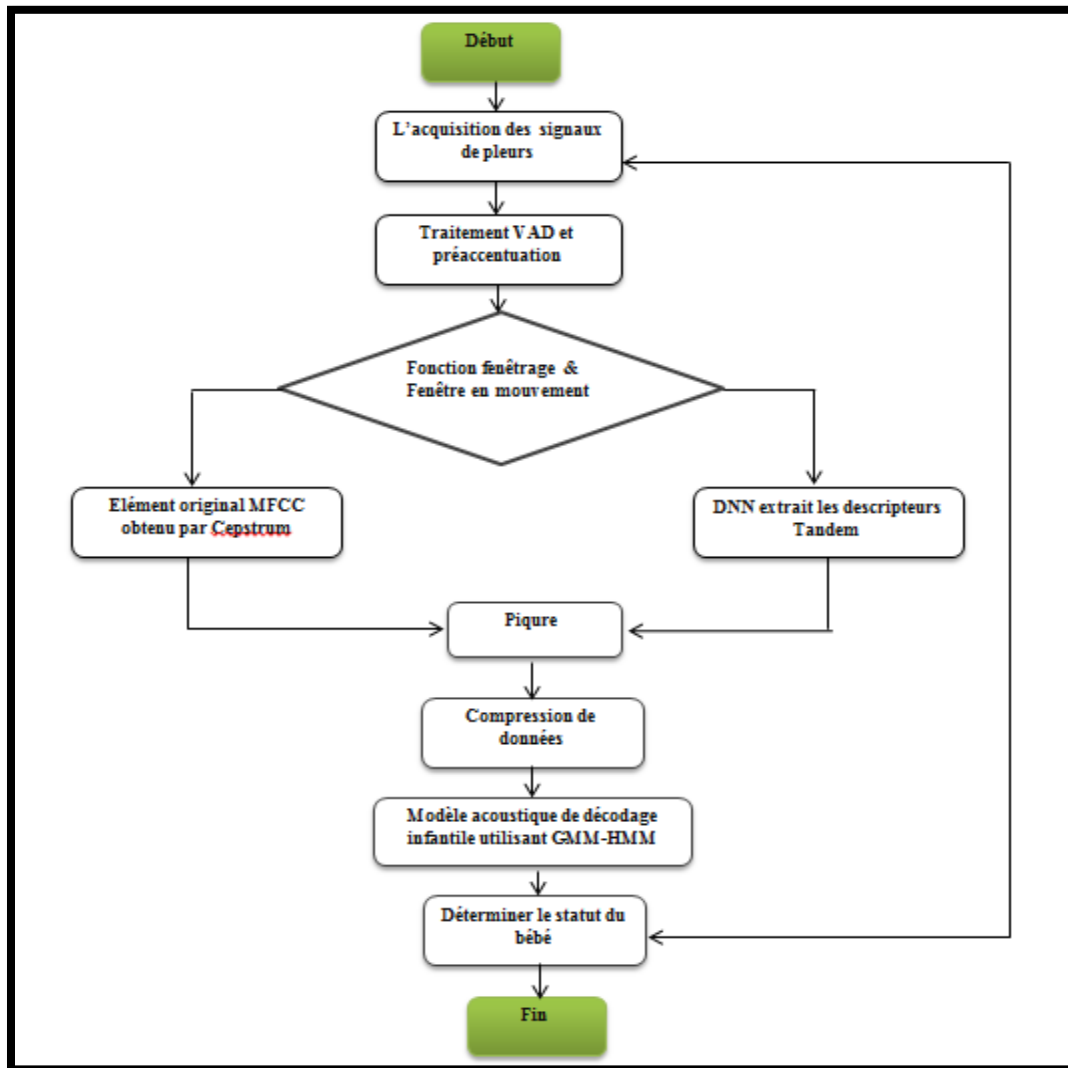


Figure 3: Organigramme de l'algorithme de reconnaissance des pleurs [5]

#### 2.4. Berceau intelligent pour bébé utilisant le module FN-M16P

En 2019 à l'institut de technologie Sambhram en Inde. Deepali Bhiwapurkar et d'autres ont mis en place un projet de berceau intelligent afin de réduire le temps de garde et de supervision au cas où les parents seraient occupés car ils ne peuvent pas s'occuper constamment de leurs enfants pendant une journée entière. [6]

Ce projet se compose de modules, dont un pour les pleurs du nourrisson, l'autre pour ses mouvements, et le dernier pour la sécurité des nourrissons. [6]

Des chercheurs dans ce domaine ont développé un microcontrôleur (Renesas) qui reçoit des informations de toutes les unités et traite les données pour d'autres usages. Lorsqu'un enfant rencontre un problème, le berceau fournira des alertes au parent concernant les problèmes cités ci-dessus. [6]

Le schéma bloc de ce projet est illustré sur la figure 4. [6]

Le Renesas est caractérisé par une ROM de 512 Ko, une RAM de 32 Ko et une mémoire flash de données de 8 Ko. [6]

Au début, ils ont utilisé un capteur sonore M213 basé sur l'ECM est utilisé pour détecter le cri et envoie les informations au microcontrôleur. Après une analyse de l'entrée, le berceau se fait basculer automatiquement à l'aide d'un moteur après un intervalle de temps de 3 comptes, si les pleurs du bébé continuent, la voix de la mère sera jouée après 5 comptes. Si le cri ne s'arrête pas, une alerte sera envoyée aux parents. [6]

Deux signaux physiologiques ont été utilisés par la suite. Le mouvement respiratoire dont un capteur à ultrasons est utilisé pour détecter la respiration du nourrisson. Lorsque le bébé souffre d'apnée, des informations seront envoyées aux parents par le microcontrôleur via GSM. Et le mouvement non respiratoire dans lequel l'accéléromètre est utilisé pour se fixer au poignet ou à la cheville du nourrisson, ce qui aidera à détecter le mouvement du nourrisson par rapport aux axes x et y. Lorsqu'il n'y aura pas de mouvement détecté une alerte sera envoyée. [6]

La surveillance de présence du bébé est basée aussi sur le module GSM. Lorsque l'enfant sort du berceau une notification est transmise à la mère via GSM. [6]

Un écran LCD est utilisé pour afficher toutes les informations. [6]

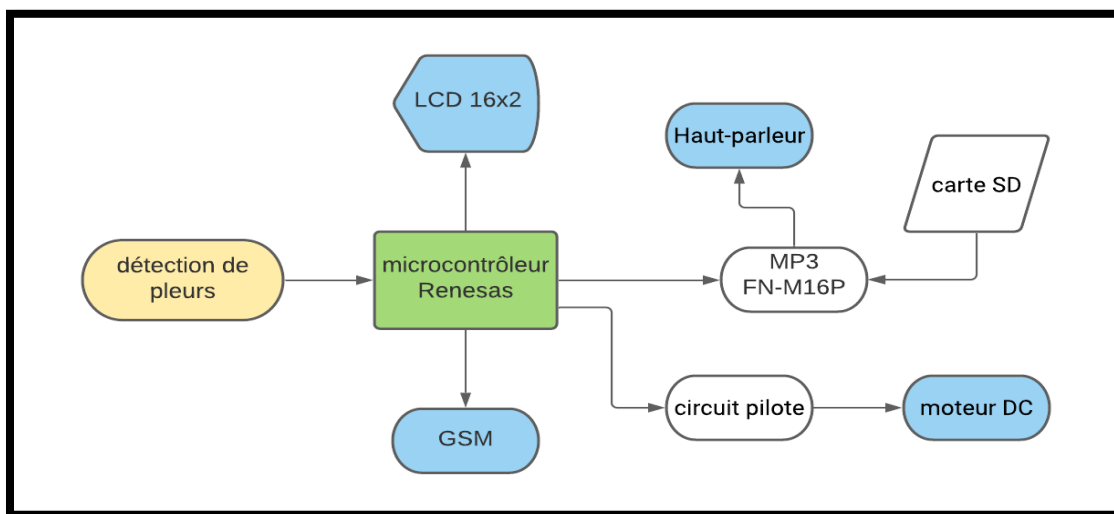


Figure 4: schéma bloc [6]

## 2.5. Système intelligent de détection des pleurs des bébés avec Raspberry Pi

Ce projet a été réalisé en 2020 à l'université de Tlemcen par les deux étudiants MEZOUAR Bochra Nour ElHouda et BOUREGA Mohammed. [7]

Dans un premier temps, ils ont fait un circuit global pour tester le microphone à condensateur. [7]

Le circuit global est composé d'un capteur de son, d'un circuit de polarisation, de filtrage, un Arduino pour la transmission du son et un écran d'ordinateur utilisé pour la visualisation. [7]

Ils ont utilisé un capteur de son qu'est un microphone standard à condensateur à électret. Ce capteur a une sensibilité moyenne pour détecter le son uniquement à courte distance et faible sensibilité. [7]

Pour l'étape de polarisation, ils ont testé environ cinq circuits différents pour trouver le plus compatible avec l'application et celui qui peut leur donner les meilleurs résultats pour la polarisation du microphone. Afin d'éliminer les bruits, ils ont utilisé un filtre passe-haut, un filtre passe-bas et un filtre passe-bande tous au format passif, les résultats étaient négatifs. Puis ils ont utilisé les filtres actifs utilisant l'amplificateur TL081 et l'amplificateur LM741 à la recherche de la meilleure acquisition du signal sonore. [7]

Le signal obtenu à la sortie du circuit de filtrage est injecté à une carte Arduino pour la conversion analogique numérique et la visualisation du signal sur l'ordinateur. [7]

Ils ont testé aussi un autre capteur sonore KY-083 qu'il a donné des résultats plus optimaux pour l'onde signal. [7]

Au lieu d'Arduino Uno, ils ont utilisé le Raspberry Pi connecté à un convertisseur analogique numérique (ADS) et un capteur de son KY-083. Ils ont constaté que ce dernier sera bon pour leur application en particulier pour la fréquence d'échantillonnage, le nombre de bits et la vitesse de calcul. [7]

Les données utilisées dans ce projet ont été collectées à l'hôpital Pranaam, Madinaguda, Hyderabad, sous la supervision du Dr Manish Gour (MBBS, DCH) et du Dr Nizam (MBBS). La tranche d'âge des nourrissons était restreinte entre 3 mois et 2 ans. [7]

Ils ont été utilisés plusieurs méthodes telles que l'analyse prédictive linéaire, la transformée en Ondelette Continu CWT, la transformée en Ondelette Discrète DWT et le MFCC pour analyser les pleurs du nourrisson et en extraire les caractéristiques énergétiques et les coefficients caractérisant les pleurs de bébés. [7]

Les pleurs de bébé ont été classés en deux catégories telles que la douleur et l'inconfort en utilisant plusieurs méthodes, telles que les réseaux neuronaux (NN), les réseaux neuronaux convolutifs (CNN) et la machine à vecteurs de support (SVM). [7]

Les meilleurs résultats ont été obtenu par l'association de la technique des Réseaux de Neurones (NN) avec la méthode MFCC (Mel-Frequency Cepstral Coefficients sont des coefficients cepstraux calculés par une transformée en cosinus discrète appliquée au spectre de puissance d'un signal). [7]

Le schéma de la figure 5 présente les différentes méthodes citées au-dessous. [7]

Les résultats obtenus sont directement affichés via la visionneuse « VNC viewer » installée dans le téléphone. [7]

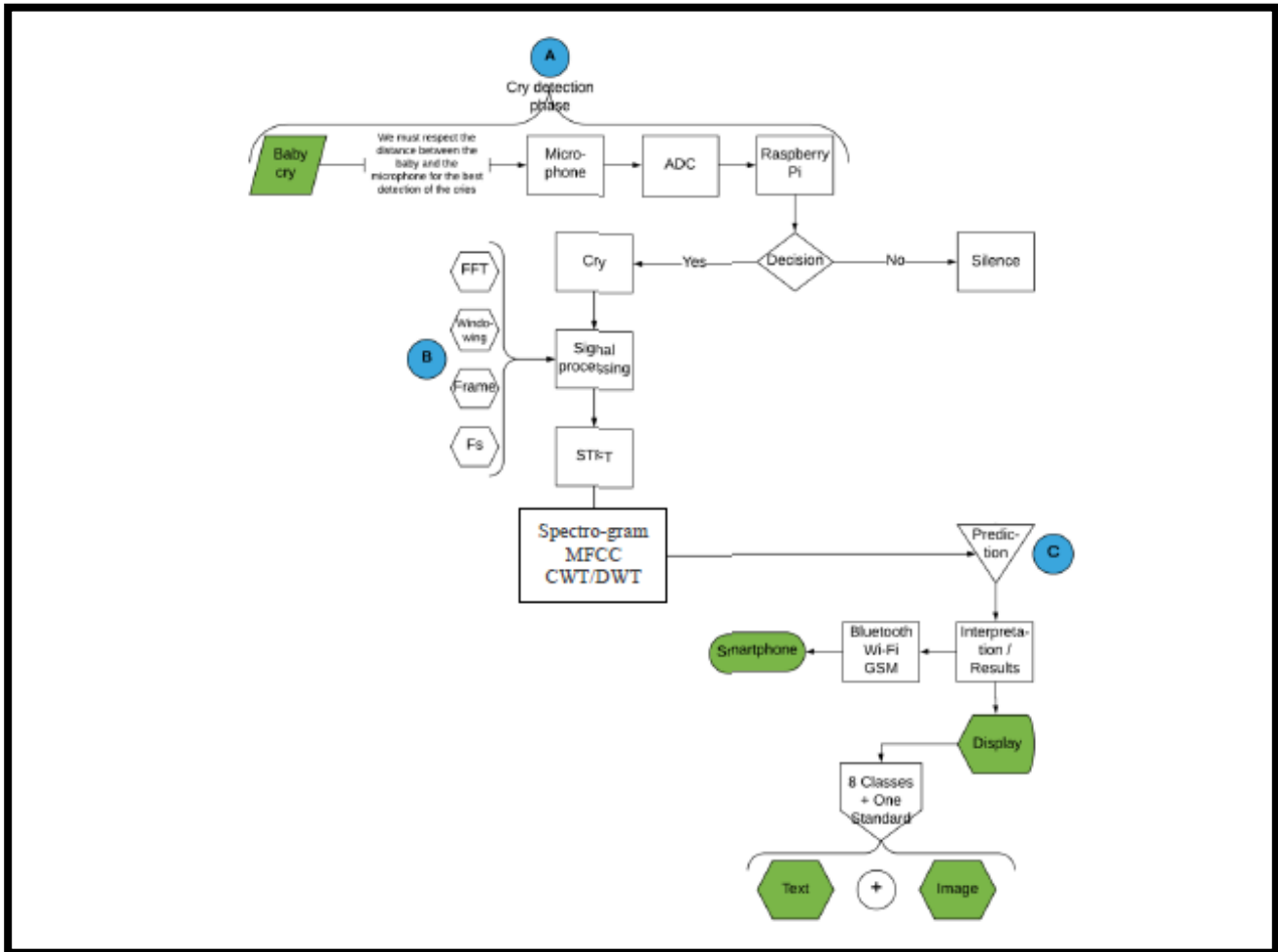


Figure 5: organigramme du projet [7]



### 3. Problématique

Nous avons vu au cours de cet état de l'art que de nombreux systèmes existent pour aider les parents à s'occuper leur enfant. Mais l'une des difficultés auxquelles sont confrontés la plupart des nouveaux parents est d'essayer de deviner ce que veut leur nouveau-né : a-t-il à faim, froid, soif, s'ennuie-t-il, veut-il un câlin ?

Les premiers jours avec le bébé peuvent donc être déroutants, vous n'êtes pas encore sûre de ce que veut votre enfant. Cela peut même être une source d'inquiétude parentale : est-ce que mon bébé va bien?

Une autre raison qui nous a poussés à réfléchir à une solution c'est pour aider les mamans sourdes pour connaître quand et pourquoi leur bébés pleurs (plus de 5% de la population mondiale, soit 389 millions de personnes, souffre de déficience auditive incapacitante [8])

Il y a aussi le problème du prix élevé de l'appareil, qui constitue un obstacle pour les familles aux revenus limités.

Notre objectif est de réaliser un système intelligent facile à utiliser qui aide autant que possible les parents à comprendre pourquoi leur bébé pleure.

### 4. Méthode

Portons comme solution à la problématique expliquée précédemment, nous avons mis en place un système qui aide les parents à comprendre la cause de la pleure du bébé. Ce système aura une utilité indispensable aux parents qui ont un déficit auditif, aussi, facilitera la tâche aux parents normaux.

La figure 6 présente un schéma bloc générale de notre système.

La première partie consiste à faire une analyse des données pour la préparation du système d'apprentissage automatique.

La deuxième partie est d'acquérir le signal utile (pleurs de bébé). Le composant principal est un microphone de haute qualité.

Le signal capté est transmis vers le pc pour faire un traitement utilisant le langage PYTHON.

La troisième partie est d'utiliser la méthode MFCC pour extraire les paramètres nécessaires pour les comparer avec la base de données. La quatrième partie est de faire une classification utilisant la méthode CNN.

La dernière partie consiste à transmettre les résultats obtenus de la classification vers un smartphone (système de l'affichage) à l'aide d'une application Android réaliser sous « app Inventor ».

Nous avons choisi la méthode MFCC associée à la méthode CNN utilisée dans le mémoire de fin d'étude de MEZOUAR et BOURGA, en raison de sa précision.

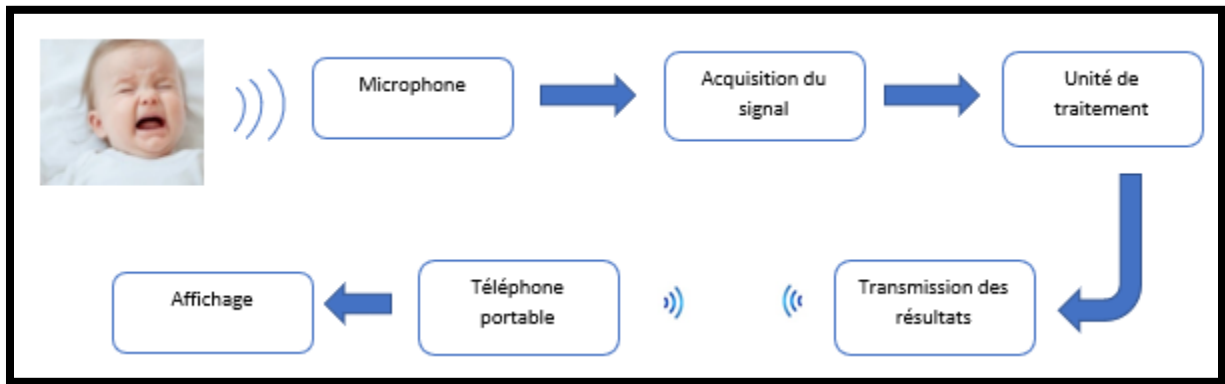


Figure 6: schéma bloc

## 5. Conclusion

Ce chapitre a été consacré à découvrir les travaux déjà effectués. Dans chaque recherche nous avons vu les différents moyens d'acquisition des pleurs des bébés, les différentes méthodes de traitement et de classification du signal vocale du bébé.

Le chapitre suivant est consacré à une présentation générale du signal vocale et principalement le signal vocal du bébé.

## Chapitre II : Généralités

### 1. Introduction

Les façons dont nous comprenons et percevons le monde qui nous entoure en tant qu'êtres humains sont connues par les sens.

Le sens de l'ouïe est l'un des cinq sens de l'homme, qui aide à recevoir des informations de l'environnement environnant à travers le système auditif dans l'oreille, sous la forme d'ondes acoustiques et de les transformer en impulsions électriques qui seront transmises au cerveau pour les identifier et les classer.

Dans le cas où ce son est celui d'un nourrisson qui pleure, il sera difficile pour les parents de comprendre ce que veut leur enfant, et cela empire s'ils ont une déficience auditive.

Ce son a des caractéristiques et des propriétés qui diffèrent de l'enfant à l'adulte. C'est ce que nous aborderons dans ce chapitre. Pour cela, nous allons voir la différence entre le son pure et le son complexe, l'origine du son, le mécanisme d'audition et nous présentons quelque notion de la physiologie de surdité et ses effets.

### 2. Définition du son

Le son est une sensation auditive causée par les perturbations d'un milieu physique matériel élastique (air, eau...) et engendrée par la stimulation des éléments sensoriels de l'oreille interne, le plus souvent par les ondes acoustiques. [9]

Le son est associé au mouvement oscillatoire d'un système vibrant (source sonore). Ce phénomène crée une onde acoustique qui se propage dans toutes les directions à une vitesse de 340 m/s (célérité du son dans l'air). [9]

D'un point de vue physiologique, c'est un signal perçu par le sens de l'ouïe, d'un point de vue physique, c'est une variation de pression locale qui se propage dans un milieu matériel élastique. [9]

Seules les ondes acoustiques dont les fréquences se situent entre environ 20 Hz et 20 kHz suscitent une perception auditive chez l'homme (voir le schéma de la figure 7). Les ondes sonores supérieures à

20 kHz sont appelées ultrasons et ne sont pas audibles par les humains. Les ondes sonores inférieures à 20 Hz sont appelées infrasons. Différentes espèces animales ont des plages d'audition variables. [10]

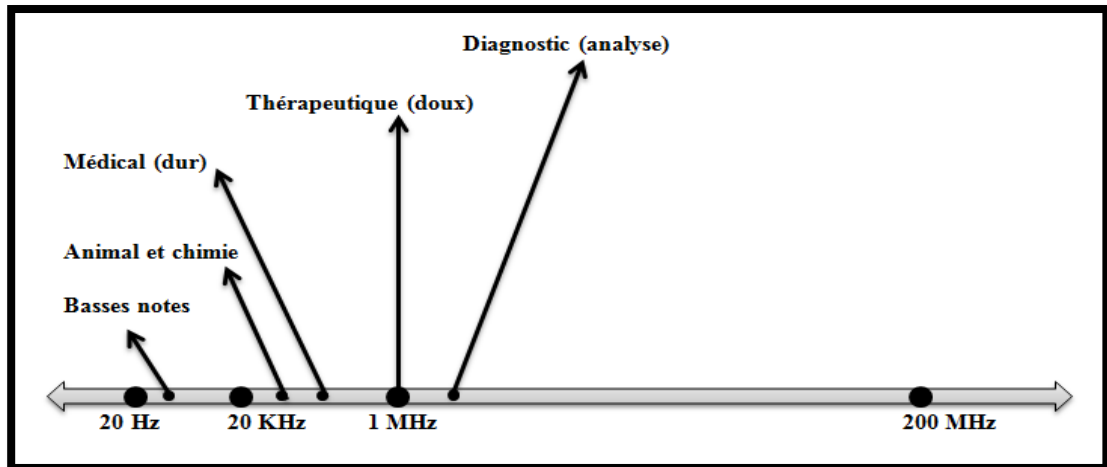


Figure 7: diagramme d'onde sonore

### 3. Les caractéristiques du son

La courbe mathématique d'une onde permet de décrire l'évolution de la perturbation du milieu de propagation : elle représente la variation de la grandeur physique considérée en fonction du temps. Cette variation est associée à un apport d'énergie. [9]

L'onde véhicule ainsi une information et de l'énergie en provenance de la source, modifiées par le(s) milieu(x) de transmission entre la source et le récepteur. Cette information est de nature différente selon le type d'onde. [9]

Lorsque la vibration reste la même, la perturbation se répète périodiquement comme une série d'ondes continues. Les variations de l'onde périodique continue peuvent être représentées par une ligne sinusoïdale.

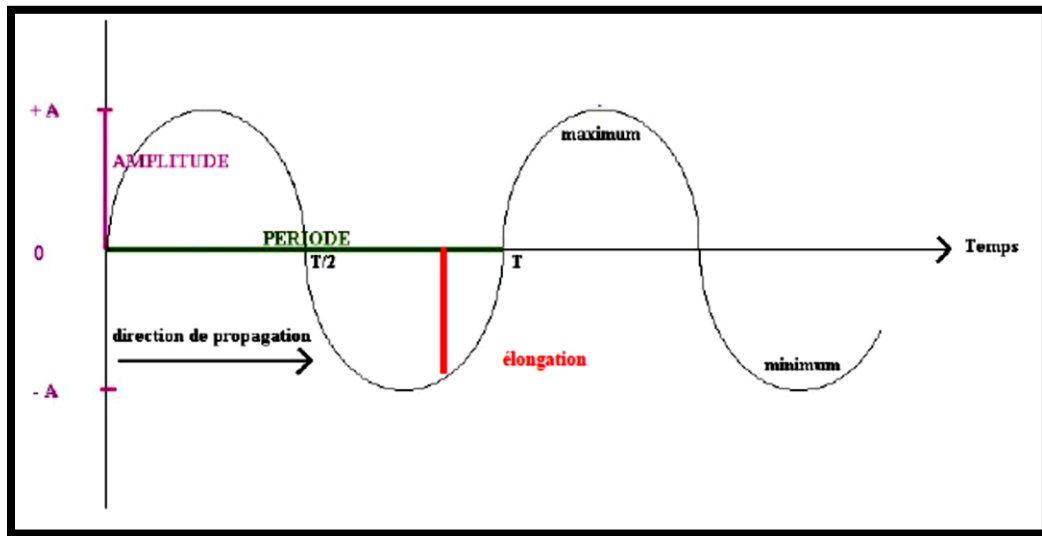


Figure 8: onde périodique [9]

L'onde sonore peut être décrite par les caractéristiques suivantes : la période ( $T$ ), la fréquence ( $\nu$ ), la longueur d'onde ( $\lambda$ ), l'amplitude ( $A$ ), la phase, la célérité ( $c$ ) et le spectre.

### 3.1. La période

La période d'un signal est la plus petite durée au bout de laquelle le motif élémentaire se reproduit identique à lui-même. Elle correspond donc à la durée séparant deux points de la courbe de même valeur et de même sens de variation. [11]

Elle est désignée par la lettre  $T$ .

### 3.2. La fréquence

La fréquence mesure le nombre de fois qu'un phénomène se répète par unité de temps. La fréquence des phénomènes ondulatoires tels que le son, les ondes électromagnétiques (comme les ondes radio ou la lumière), c.-à-d. le nombre de fois que la forme d'onde se répète en une seconde. [12]

### 3.3. La longueur d'onde

La longueur d'onde est l'une des caractéristiques propres à chaque onde, quelle que soit sa nature. Elle se note à l'aide de la lettre grecque lambda :  $\lambda$ . Elle représente la périodicité spatiale des oscillations, c'est-à-dire la distance entre deux maxima de l'oscillation, par exemple. La longueur d'onde est aussi la

distance parcourue par l'onde pendant une période d'oscillation. Ainsi, elle est inversement proportionnelle à la fréquence et s'exprime en mètre. [13]

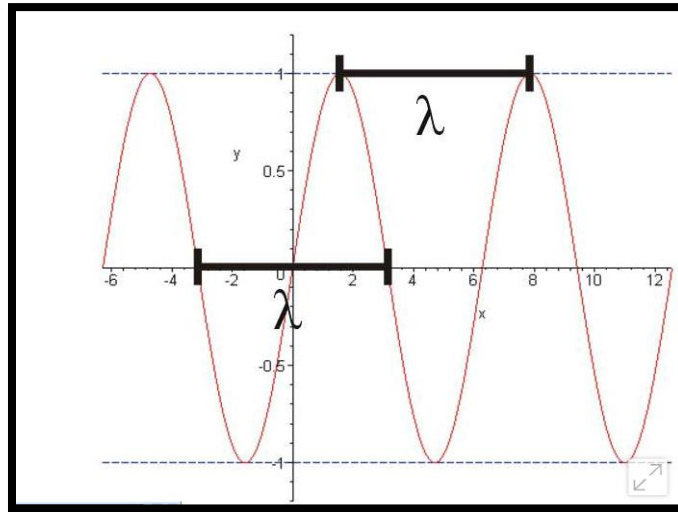


Figure 9: un signal représentant la longueur d'onde [13]

Notez que la longueur d'onde dépend de la vitesse à laquelle l'onde se propage dans le milieu. Ainsi, lorsqu'une onde passe d'un milieu à un autre en changeant de vitesse, sa longueur d'onde varie, même si sa fréquence reste la même. Le tout selon la relation suivante :  $\lambda = c.T = c/f$ , où  $c$  correspond à la célérité de l'onde,  $T$  à sa période temporelle et  $f$  à sa fréquence. [13]

### 3.4. L'amplitude

Lorsqu'une onde traverse un milieu, le déplacement maximal des particules de ce milieu par rapport à leurs positions d'origine non perturbées, est appelé amplitude de l'onde. [14]

L'unité S.I de mesure d'amplitude est le mètre (m). [14]

### 3.5. La phase

$\Phi$  est la phase à l'origine, ou déphasage. Et elle se mesure en radians (ou degrés). [15]

### 3.6. La célérité

C'est la vitesse à laquelle l'onde se propage dans sa direction, et elle dépend des caractéristiques du milieu (densité, compressibilité).

Elle est égale à 340 m/s dans l'air, 1450 m/s dans l'eau, 1540 m/s dans les tissus et dans l'os 3300 m/s (donc plus le milieu est dense plus la vitesse est grande). [9]

### 3.7. Le spectre

La décomposition du son en une série d'harmoniques constitue le spectre du son. [9]

Les harmoniques sont issus de la décomposition d'un son périodique en une succession de sons dont les fréquences sont toutes des multiples d'une même fréquence  $f$  appelée fondamental :  $f$ ,  $2f$ ,  $3f$ ,  $4f$ , etc. [9]

## 4. Les types de son

Il existe différents types de son :

### 4.1. Les sons purs ou sons simples

Les sons purs sont caractérisés par une fréquence unique et une amplitude. Dans la nature les sons purs sont très rares. On trouve par exemple le « La » d'un diapason, qui a une fréquence de 440 Hz. [16]

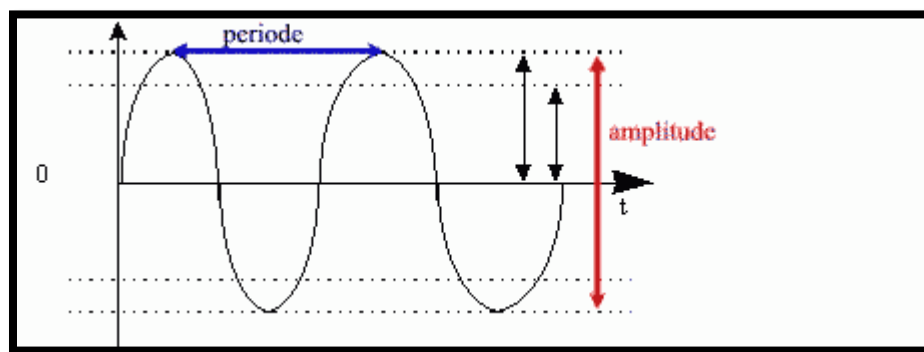


Figure 10: un signal représentant le son pur [16]

## 4.2. Les sons complexes

Les sons complexes sont formés d'une superposition de sons simples ayant des amplitudes et des fréquences différentes. Ils possèdent une fréquence fondamentale, qui est la fréquence la plus basse d'un son complexe, et d'autres fréquences appelées harmoniques qui sont des multiples entiers de la fréquence fondamentale. [16]

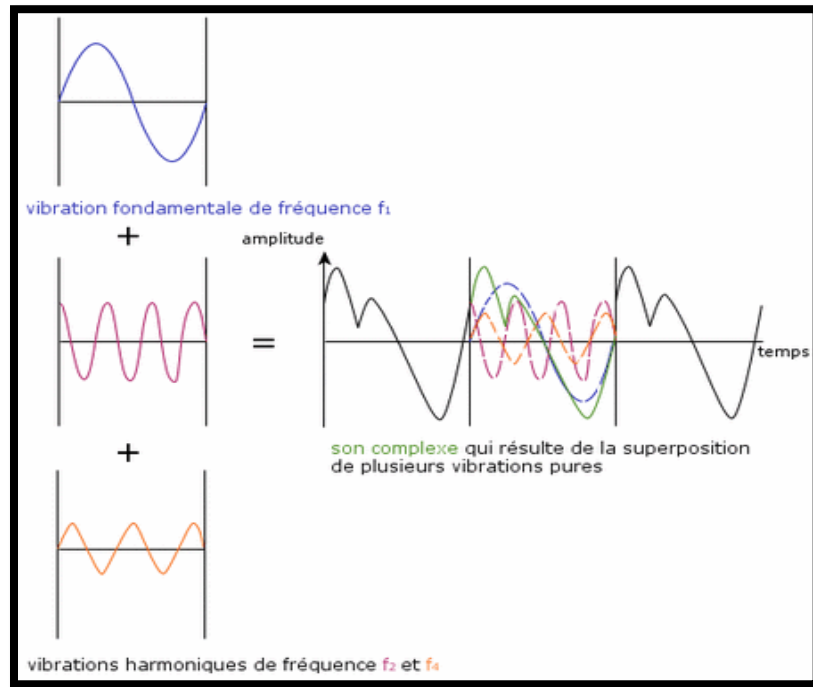


Figure 11: un signal représentant le son complexe [16]

## 5. Les propriétés du son

### 5.1. L'impédance (Z)

Elle caractérise la résistance d'un milieu au passage du son. [9]

L'impédance acoustique d'un milieu pour une onde acoustique est le rapport de la pression acoustique ( $p$ ) et de la vitesse de la particule associée du milieu ( $v$ ). [17]

$$Z = \frac{P(\text{Pa})}{v\left(\frac{\text{m}}{\text{s}}\right)}$$



## 5.2. La pression acoustique (P)

La surpression positive ou négative ajoutée par l'onde sonore à la pression de repos (pression atmosphérique = 105 Pa). [9]

$$P = v \times \rho \times c \quad [9]$$

$v$  : vitesse des particules mises en mouvement.

$\rho$  : La masse volumique spécifique du fluide dans lequel se propage le son.

$c$  : la célérité du son dans ce milieu.

## 5.3. La puissance acoustique (W)

L'onde sonore correspond à un transport d'énergie.

La puissance acoustique est la quantité d'énergie rapportée à l'unité de temps.

L'unité de La puissance acoustique est le **W**.

$$W = P \times v \quad [9]$$

Lorsque  $v$  n'est pas connu.

$$W = \frac{P^2}{\rho \times c} \quad [9]$$

## 5.4. Intensité sonore (I)

C'est l'énergie transportée par l'onde par unité de temps et de surface, c'est donc la valeur efficace de la puissance acoustique par unité de surface. [9]

Comme valeur standard l'American Standard Association a choisi le son de référence de puissance acoustique égale à  $10^{-12} \text{ W.m}^{-2}$ , correspondant à un son pur de 1000 Hz juste audible. [9]

## 6. La voix humaine

Parmi les voyelles produites par la voix humaine, le groupe harmonique accentué (appelé "consonnes") joue un rôle essentiel.

La voyelle provient de la résonance dans les cordes vocales. La longueur moyenne des cordes vocales humaines est d'environ 17,5 cm, avec l'extrémité inférieure aux cordes vocales et l'extrémité supérieure aux lèvres. Comme une anche ou comme une lèvre à l'embouchure d'un instrument à vent, la corde vocale joue un rôle acoustiquement fermé, de sorte que la colonne vocale est un résonateur tubulaire fermé avec des fréquences de résonance d'environ 500 Hz, 1500 Hz, 2500 Hz, 3500 Hz, etc. [18]

La fréquence de vibration des cordes vocales, déterminée par la tension des cordes, détermine la fréquence du son vocal. Lorsqu'un son est produit, toutes les harmoniques sont présentes dans le spectre, mais celles proches des fréquences de résonance de la colonne vocale sont augmentées en amplitude. Ces régions de fréquence accentuées sont les formants vocaux. En changeant la forme de la gorge, de la bouche et des lèvres, les fréquences des formants sont variées, créant les différents sons de voyelles. [18]

## **7. Les pleurs de bébé**

Le cri d'un nouveau-né est une condition particulière de la parole humaine, et comme la parole humaine, le cri est un signal stationnaire à court terme. Étant donné que les nouveau-nés n'ont pas le contrôle total du conduit vocal, les signaux de cris sont considérés comme plus stationnaires que les signaux de parole. [19]

## **8. L'origine et le rôle du pleur du bébé**

Les pleurs vocaux sont l'une des premières formes de communication pour interagir avec les mamans et les baby-sitters.

Les pleurs chez les nourrissons humains sont provoqués par des transitions rythmiques entre l'inspiration et l'expiration, en raison d'une vibration des cordes vocales qui produit des impulsions d'air périodiques. La période de ces impulsions est appelée fréquence fondamentale (hauteur tonale) et ses valeurs typiques chez les bébés en bonne santé sont de 250 à 600 Hz. Le signal de cri est façonné par le tractus vocal, conduisant à des fréquences de résonance appelées formants. Les deux premiers formants se produisent typiquement autour de 1100 Hz et 3300 Hz, respectivement. [3]

Certaines études soulignent que les humains tôt dans la vie présentent déjà diverses formes de pleurs: des pleurs de protestation (dans lesquels le bébé fait face à la perte, comme être laissé dans le berceau, et veut annuler la perte), des pleurs tristes de désespoir (un gémissement faible signifiant l'acceptation de

la perte) et des pleurs détachés et inhibés (généralement une absence de pleurs vers l'extérieur, associée à une séparation potentiellement mortelle de l'aidant). Parfois, des types supplémentaires, tels que la faim et les pleurs de douleur, sont considérés. [3]

## 9. Les caractéristiques de signal pleurs des bébés

Des études ont montré qu'il existe différents types de cris selon les besoins du nouveau-né, tels que la faim, la douleur, l'inconfort, etc. Un néonatalogiste ou un pédiatre peut faire la distinction entre différents types de cris et peut trouver des modèles dans chaque type de cri. [20]

Les nouveau-nés à haut risque souffrent à une irritabilité élevée et une faible stabilité physiologique parmi d'autres anomalies. [21]

Les nourrissons de deux mois n'ont pas de contrôle phonatoire en raison d'une immaturité neurologique à un âge précoce. Alors que l'âge augmente, la capacité d'expiration augmente également, ce qui se reflète dans la durée. [21]

## 10. Les différences entre le cri d'un bébé et un enfant

Le signal des pleurs du nourrisson diffère de la parole des adultes. La figure 12 donne une comparaison des spectrogrammes entre le son du nourrisson et la parole adulte. En général, les pleurs du nourrisson sont une combinaison de vocalisation, de silence, de toux, d'étouffement et d'interruptions, qui comprend une diversité d'informations acoustiques et prosodiques à différents niveaux. C'est le seul moyen pour les bébés de communiquer avec le monde. [22]

Il est montré que le cri du nourrisson est composé de quatre types de sons : un venant de la phase d'expiration, une brève pause et un son venant de la phase d'inspiration suivi d'une autre pause. Les variations d'intensité, de fréquence fondamentale ( $F_0$ ), de formants et de durée sont des indices acoustiques typiques qui véhiculent des informations prosodiques sur les pleurs et la parole du nourrisson. Les caractéristiques prosodiques citées ci-dessus sont efficaces pour identifier les types de pleurs du nourrisson. [22]

La fréquence fondamentale  $F_0$  de l'adulte varie entre 85 et 200 Hz, tandis que celle du signal vocal de la pleure du nourrisson se caractérise par une fréquence fondamentale  $F_0$  élevé de 250 à 700 Hz.  $F_0$  est généralement calculé en utilisant une méthode basée sur l'autocorrélation fournie par Praat. [22]

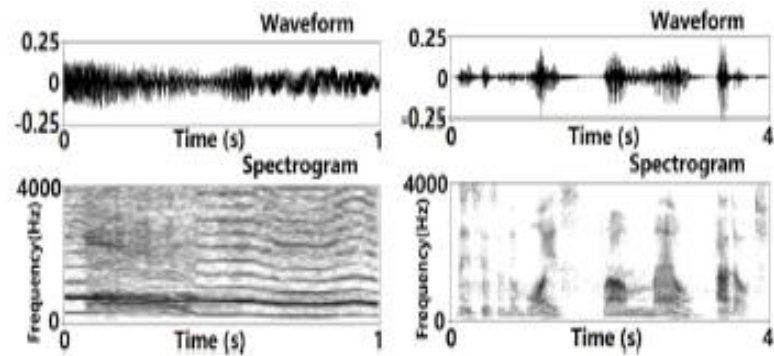


Figure 12: Parole d'adulte vs. Pleurs de bébé dans le domaine temporelle et fréquentielle [22]

## 11. Anatomie de l'oreille

L'oreille est un organe neurosensorielle pair. On distingue trois parties : l'oreille externe, l'oreille moyenne et l'oreille interne. [23]

### 11.1. L'oreille externe

L'oreille externe se compose du pavillon et du conduit auditif externe. Le pavillon est la partie externe et visible de l'oreille. Il sert à identifier des individus. [24]

Le conduit auditif externe (CAE) s'étend du fond de la conque au tympan. [25]

### 11.2. Oreille moyenne

L'oreille moyenne contient les trois plus petits os du corps humain : le marteau, l'enclume et l'étrier. Ces os forment une chaîne d'osselets qui transmet les vibrations du tympan à l'oreille interne. [24]

### 11.3. Oreille interne

L'oreille interne est une structure complexe appelée labyrinthe osseux. C'est en réalité une cavité qui comprend trois parties : la cochlée, le vestibule et les canaux semi-circulaires. La cochlée est l'organe de l'audition, en forme de spirale. [24]

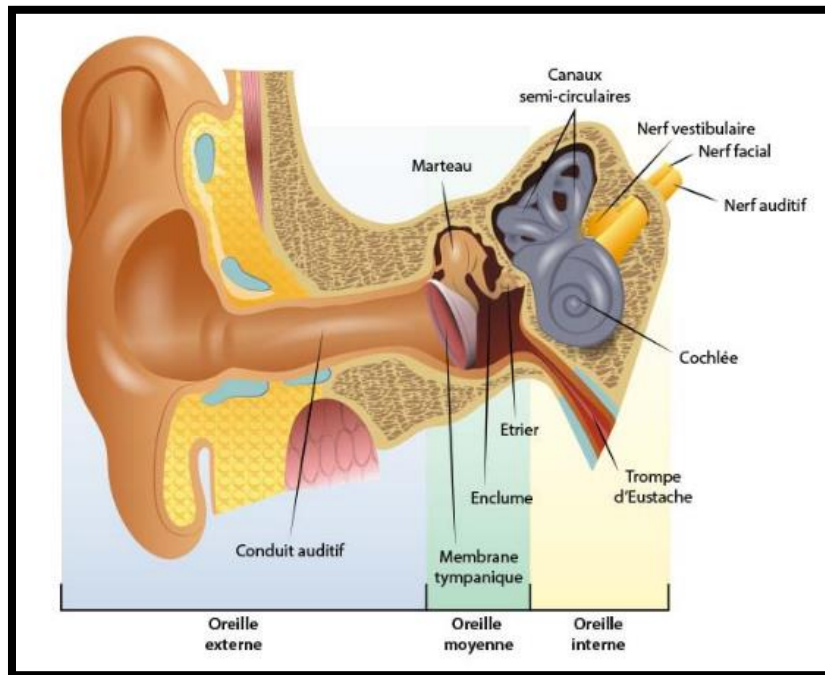


Figure 13: Schéma des principales parties du système auditif périphérique

## 12. Le mécanisme de l'audition

Lorsqu'une onde sonore entre dans le conduit auditif externe, elle frappe le tympan qui vibre. Ces vibrations sont transmises à la chaîne des osselets, qui à leur tour, propagent le mouvement aux liquides de l'oreille interne. Ces vibrations parviennent à la cochlée qui renferme l'organe de Corti : c'est le récepteur des vibrations. Ces cellules traduisent les mouvements sonores en influx nerveux. Ces messages sont transmis par le nerf cochléaire au cerveau où il est décodé et interprété. [24]

## 13. La surdité

La surdité est un état pathologique de l'audition caractérisé par une perte partielle ou totale de la perception des sons. Lorsque la perte est totale, il s'agit d'une cophose, parfois appelée anacousie. Le terme hypoacousie est quasi-synonyme de surdité, étant parfois réservé aux cas où la perte de l'audition est partielle. [26]

### **13.1. Les types de la surdité**

Parce que le traitement de la surdité dépend de la localisation du type de lésion ainsi que du dysfonctionnement de l'appareil auditif, il est important de connaître et de comprendre les différents types de surdité ainsi que leur gravité. [27]

#### **13.1.1. La surdité de transmission**

La surdité de transmission concerne une déficience ou une atteinte du conduit auditif, du tympan ou de la chaîne des osselets au niveau de l'oreille moyenne ou de l'oreille externe. [27]

#### **13.1.2. La surdité de perception**

La surdité de perception désigne toutes les formes de pertes auditives qui touchent l'oreille interne ou les voies nerveuses situées en arrière de la cochlée. Ce qui implique dans tous les cas une mauvaise transmission des informations au cerveau. [27]

#### **13.1.3. La surdité mixte :**

La surdité mixte regroupe la surdité de transmission et la surdité de perception c'est-à-dire qu'elle allie une mauvaise transmission de l'oreille externe ou de l'oreille moyenne à un problème de perception de l'oreille interne ou du nerf auditif. [27]

## **14. Impact de la perte auditive**

### **14.1. Impact fonctionnel**

L'un des principaux impacts de la perte auditive est la capacité de l'individu à communiquer avec les autres. [28]

La perte auditive non traitée et les maladies de l'oreille telles que l'otite moyenne peuvent avoir un effet défavorable significatif sur les résultats scolaires des enfants. Ils ont souvent des taux élevés d'échec scolaire et un plus grand besoin d'aide à l'éducation. L'accès à des aménagements appropriés est important pour des expériences d'apprentissage optimales, mais ils ne sont pas toujours disponibles. [28]

### 14.2. Impact social et émotionnel

Les déficiences auditives chez les parents conduisent à la solitude, l'isolement et la dépression. Ce qui les pousse à trouver un programme de garderie ou une baby-sitter qui peut communiquer avec eux et leurs enfants. [29]

### 14.3. Impact économique

L'OMS estime que la perte auditive non traitée coûte au monde 750 milliards de dollars par an, y compris les coûts pour le secteur de la santé (à l'exclusion du coût des appareils auditifs), les coûts de soutien éducatif, la perte de productivité et les coûts sociaux. [28]

## 15. Principaux faits

Au niveau mondial, 1,5 milliard de personnes sont atteintes d'une déficience auditive plus ou moins prononcée. Parmi elles, 430 millions ont besoin de services de réadaptation. [30]

En 2050, près de 2,5 milliards de personnes seront, d'après les projections, atteintes d'une déficience auditive plus ou moins prononcée et au moins 700 millions de personnes auront besoin de services de réadaptation. [30]

Chez l'enfant, près de 60 % des cas de déficience auditive sont dus à des causes pouvant être prévenues par des mesures de santé publique [30]

Plus d'un milliard de jeunes adultes risquent une déficience auditive permanente évitable car ils ont des pratiques d'écoute non sûres. [30]

Seuls 17 % des personnes pour qui une aide auditive serait utile en utilisent effectivement une. La part des personnes qui n'utilisent pas ces dispositifs alors qu'elles le devraient est élevée partout dans le monde ; elle va de 77 % à 83 % dans les différentes Régions de l'OMS et de 74 % à 90 % selon les niveaux de revenus. [30]

Plus de 5 % de la population mondiale, soit 430 millions de personnes, ont besoin de services de réadaptation en raison d'une déficience auditive incapacitante (432 millions d'adultes et 34 millions d'enfants). D'après les estimations, d'ici à 2050, plus de 700 millions de personnes (soit une personne sur dix) en seront atteintes. [30]

Une « déficience auditive incapacitante » désigne une perte auditive supérieure à 35 décibels (dB) dans la meilleure oreille. Près de 80 % des personnes qui en sont atteintes vivent dans des pays à revenu faible ou intermédiaire. La prévalence de la déficience auditive augmente avec l'âge : plus de 25 % des sujets âgés de plus de 60 ans présentent une déficience auditive incapacitante. [30]

## 16. Conclusion

Dans ce chapitre nous avons présenté les notions de base concernant le son et les problèmes auditifs en particulier pour les personnes malentendantes et ce à quoi elles sont confrontées dans le monde.

Grace à l'étude que nous avons menée durant ce chapitre, on a acquis des connaissances sur l'origine et le rôle du pleur du bébé.

Notre recherche de cette connaissance nous a poussés à réaliser un système simple permettant de prédire les pleurs de bébé.



## Chapitre III : méthodologie

### 1. Introduction

Dans ce chapitre on va présenter pas a pas les différents composants de logiciels utilisés lors de la partie programmation avec de notion sur la méthode, les bibliothèques et aussi les spécifications de la machine utiliser pour construire le model.

Notre modèle est basé sur un modèle CNN (voir les travaux effectués), ce réseau à donner les milliers résultats de la classification des signaux de pleurs de bébé.

Le langage python a été utilisé afin de réaliser le modèle de classification.

L’affichage des résultats de notification de la maman est fait à l’aide d’une application Android réaliser sous « App Inventor ».

L’organigramme suivant représente un schéma bloc générale de travail.

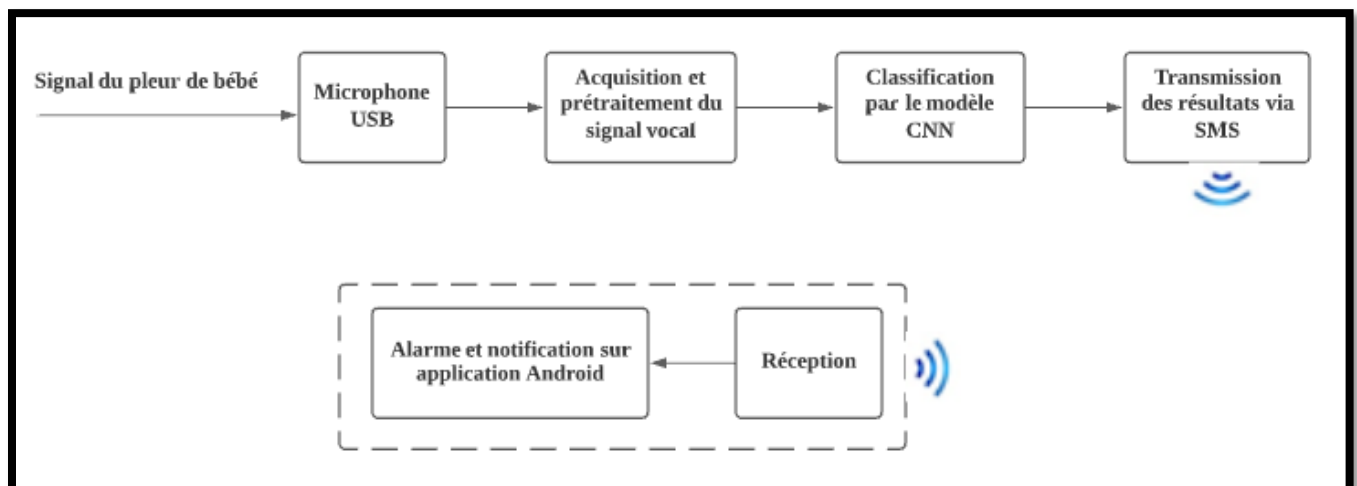


Figure 14: Schéma bloc générale

Notre travail est partitionné en deux parties majeures : hardware et software.

## 2. Hardware

Spécification	Hp pro book 4530s	Redmi note 8 pro	Raspberry pi 3 (modèle B)
<b>Systèmes d'exploitation</b>	Windows 10 professionnel 64 bits	Android Version : 9PPR1	Windows 10 professionnel 64 bits
<b>Processeurs</b>	I5 2eme génération 2.3 GHz	OCTA CORE MAX 2.05 GHz	ARM Cortex-A53 1.4 GHz
<b>Stockage</b>	1 Téra	128 GB	32 GB
<b>Ram</b>	4.00 GB	6.00 GB	1 GB

Tableau 2 : les matériels utilisés

**Raspberry pi** : est un nano ordinateur de la taille d'une carte de crédit que l'on peut brancher à un écran est utilisé comme un ordinateur standard. Sa petite taille, et son prix intéressant fait du Raspberry pi un produit idéal afin de tester différentes choses. Évidemment, pour sa taille il ne faut pas s'attendre à des performances incroyables, mais pour mettre en ligne des projets à montrer au client ou expérimenter avec Linux, c'est largement suffisant. [31]

**Les autres matériels utilisés** : système d'affichage (data show ou écran), une câble HDMI, un clavier, une souris, un modem Wifi, un microphone USB, adaptateur 5V.

## 3. Logiciels utilisés (Software)

### 3.1. Anaconda Navigator

Le Navigateur Anaconda est une interface graphique (GUI) incluse dans la distribution Anaconda (open source des langages de programmation python et R), et qui permet aux utilisateurs de lancer des applications, mais aussi de gérer les librairies conda, les environnements et les canaux sans utiliser la moindre ligne de commande. Il est disponible pour Windows, MacOS et Linux. [32]

Les applications disponibles dans le navigateur : Jupyter lab, Jupyter notebook, Spyder, orange, glue, Rstudio, QtConsole, visuel studio code.

Dans notre projet nous avons utilisé les applications Jupyter notebook et Spyder.

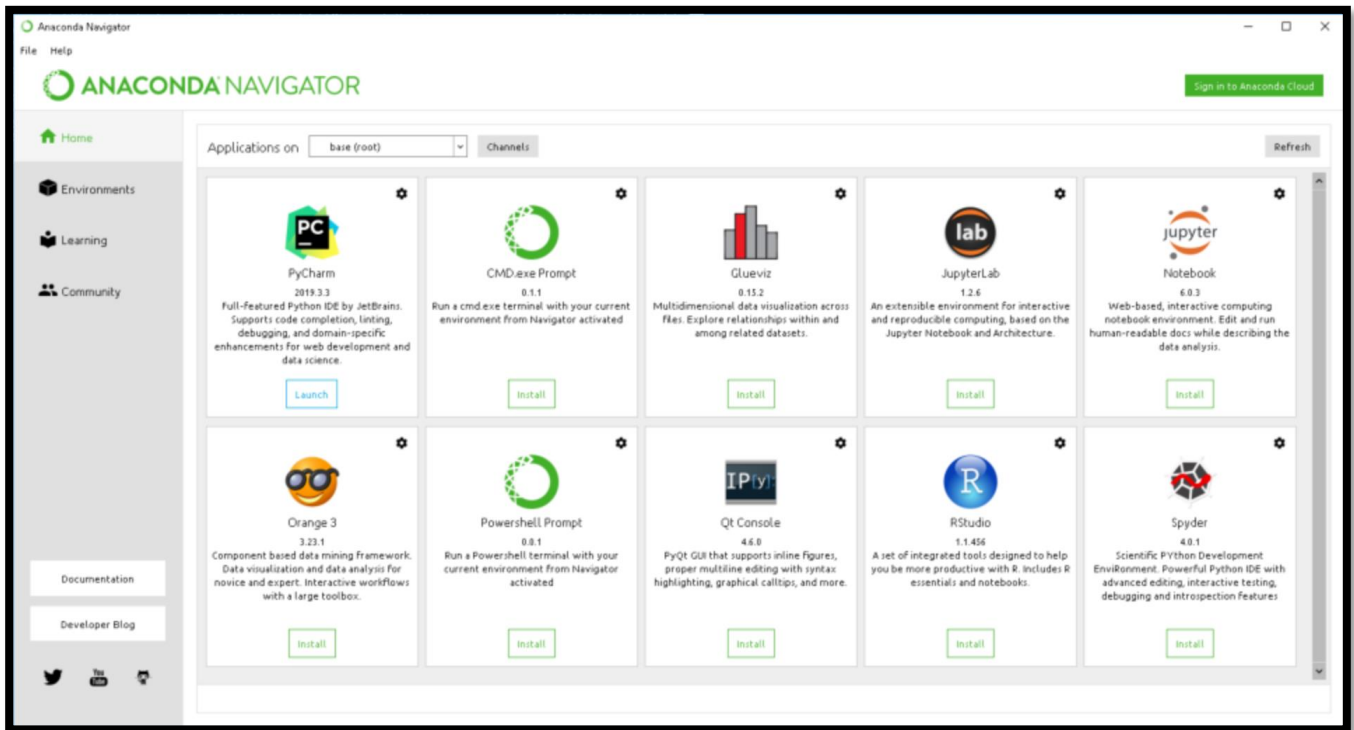


Figure 15: interface de logiciel ANACONDA NAVIGATOR

### 3.1.1. Jupyter notebook

Créé à partir de Python en 2014, Jupyter est un notebook de calcul (computational notebook) open source, gratuit et interactif. C'est une application web basée client permettant de créer et de partager du code, des équations, des visualisations ou du texte. Grâce à ses nombreux avantages, ce "bloc-notes" de calcul est devenu une référence incontournable pour les Data Scientists. [33]

### 3.1.2. Spyder

Spyder est un environnement de développement pour Python, Créé et développé par Pierre Raybaut en 2008. Spyder a un ensemble unique de fonctionnalités - multiplateforme, open-source, écrit en Python. Il intègre de nombreuses bibliothèques d'usage scientifique : Matplotlib, NumPy, SciPy et IPython. [34]

### 3.2. Langage python

Le langage Python est un langage de programmation open source multi-plateformes et orienté objet. Grâce à des bibliothèques spécialisées, Python s'utilise pour de nombreuses situations comme le développement logiciel, l'analyse de données, ou la gestion d'infrastructures. [35]

Les principales utilisations de Python par les développeurs sont : la programmation d'applications, la création de services web, la génération de code et la métaprogrammation. [36]



Figure 16: logo de python

#### Bibliothèque utilisée

Nom	Définition
<b>JSON</b>	JSON (JavaScript Object Notation) est un format d'échange de données léger. Facile d'analyser et de génère pour les machines qui suit une norme ECMA-262, aussi un format de texte qui utilise des conventions familières aux programmeurs de la famille de Python. [37]
<b>NumPy</b>	NumPy (diminutif de Python numérique) est une bibliothèque destinée à la manipulation des matrices et des tableaux multidimensionnels et les factions mathématiques. [38]
<b>Le module OS</b>	Le module OS est une fonction qui interagit avec le system d'exploitation [39]
<b>TensorFlow</b>	TensorFlow est un librairie open source d'apprentissage automatique. [40]
<b>Matplotlib</b>	Matplotlib une bibliothèque utilisée pour tracer et visualiser graphiquement les données, elle utilise des toolbox graphique telle que Tkinter, GTK, ect
<b>Math</b>	Bibliothèque intègre à Python qui contient toutes des fonction mathématique [41]

<b>Librosa</b>	Un package Python pour analyser les enchanions musicale et audio. Cette bibliothèque fournit les éléments de base au traitement du signal vocal ou musical [42]
<b>Sounddevice</b>	Un module Python qui fournit des liaisons avec le port audio pour lire et enregistrer des tableaux NumPy contenant des signaux audio [43]
<b>Scipy</b>	Une bibliothèque qui fournit des fonctions utilitaires pour l'optimisation et le traitement du signal. [44]

Tableau 1: définition des bibliothèques utilisés

### 3.3. App Inventor

App Inventor est une application développée par Google (Elle est actuellement entretenue par le « Massachusetts Institute of Technology (MIT) ») et est destiné aux utilisateurs d'Android. C'est un logiciel en ligne qui permet la création d'applications destinées à des systèmes Android (téléphones, tablettes, ...), à travers une interface purement visuelle et de configurer les actions de l'application par un système de blocs logiques. [45]



Figure 17: logo de app Inventor

### 3.4. Windows on Raspberry

Cette application est développée par 'Mario Balancia' en 2018. Il est principalement conçu pour les développeurs qui souhaitent réaliser des prototypes d'appareils IoT à l'aide du Pi. Il nous permet d'installer le Windows 10 IoT Core sur Raspberry pi. [46]

**Windows 10 IoT Core :** Ce n'est pas Windows comme la plupart des gens le comprennent. Au lieu de cela, Windows 10 IoT Core affiche aux utilisateurs une seule application Windows universelle en plein écran. Le système n'affiche que l'interface d'une seule application à la fois, mais d'autres logiciels peuvent être exécutés en arrière-plan. [46]

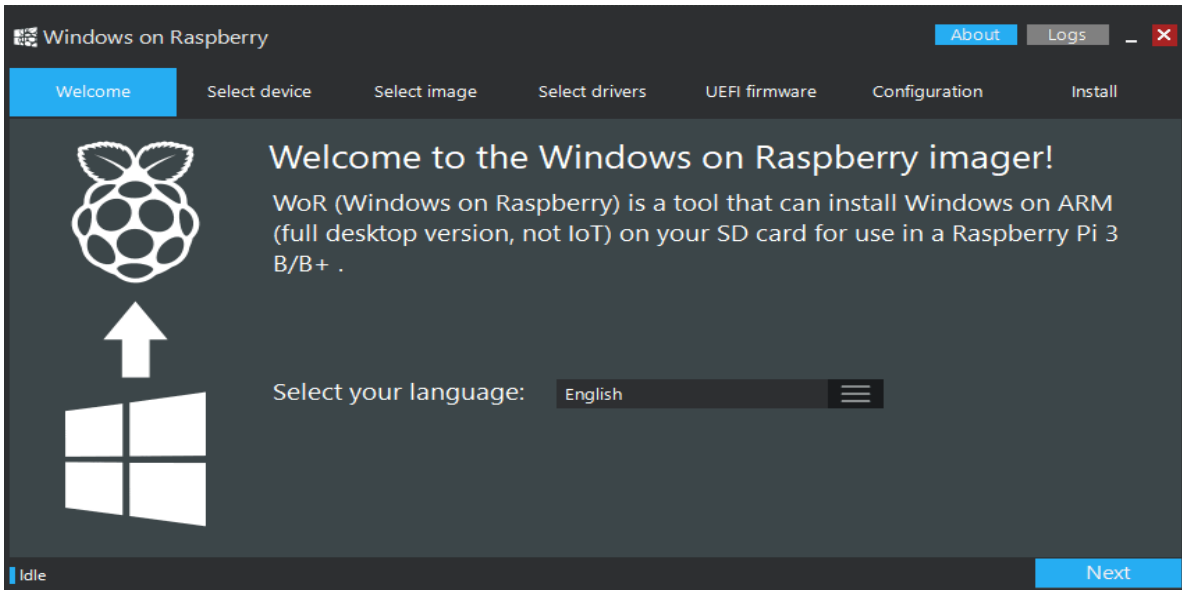


Figure 18: interface de logiciel "Windows on Raspberry"

#### 4. Mel-Frequency Cepstral Coefficients (MFCC)

Les MFCC est un moyen d'extraire des caractéristiques d'un fichier audio. Le MFCC utilise l'échelle MEL pour diviser la bande de fréquences en sous-bandes, puis extrait les coefficients cepstraux à l'aide de la transformée en cosinus discrète (DCT). L'échelle MEL est basée sur la façon dont les humains distinguent les fréquences, ce qui le rend très pratique pour traiter les sons. [47]

L'organigramme du processus MFCC est illustré dans la figure ci-dessous :

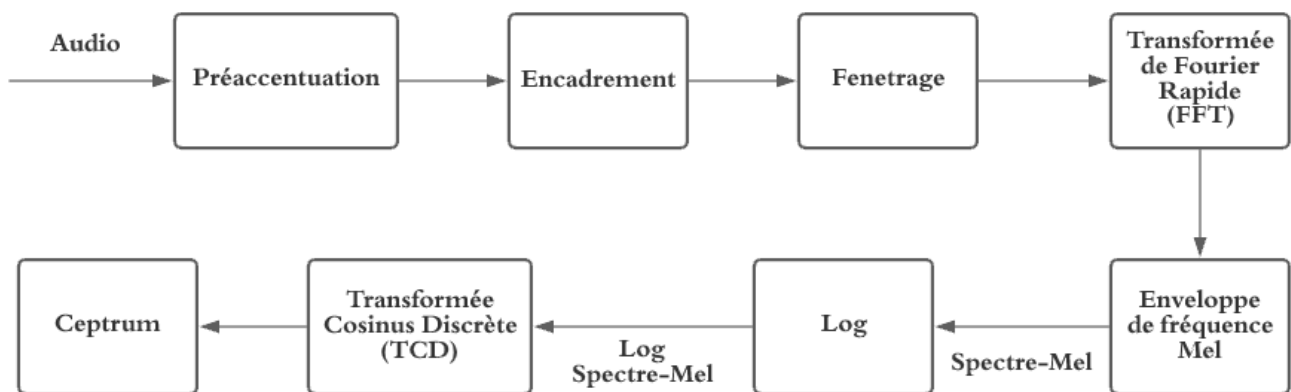


Figure 19: diagramme représente le processus MFCC

### *a. Préaccentuation*

Le processus de filtrage du signal sonore nécessitait un processus d'enregistrement ou d'échantillonnage. Le filtrage a pour but d'obtenir une forme spectrale plus douce de la fréquence sonore. La préaccentuation du filtre est basée sur la relation entrée/sortie dans le domaine temporel, exprimée dans l'équation : [48]

$$y[n] = s[n] - \alpha s[n - 1], \text{ avec } 0.9 \leq \alpha \leq 1$$

Où :

$y[n]$  = filtre de préaccentuation du signal de résultat

$s[n]$  = avant le filtre de préaccentuation du signal

### *b. Encadrement*

Les signaux sonores sont des signaux non stationnaires, ce qui signifie que leurs propriétés statistiques changent constamment avec le temps. Il n'est donc pas possible d'extraire les caractéristiques spectrales de la parole en une seule fois. Par conséquent, le caractère spectral des signaux sonores est extrait à travers une fenêtre en tant que signal sonore caractérisant une partie sonore particulière de sorte qu'il peut être supposé que les signaux sonores soient stationnaires. [48]

### *c. Fenêtrage*

Le processus de fenêtrage est le processus de filtrage de chaque image en multipliant chaque image avec certaines fonctions de fenêtres de la même taille que l'image. Le fenêtrage est également utilisé pour assurer un son continu de la trame initiale à la fin. Une bonne fonction de fenêtre doit être réduite au lobe principal et s'étendre à son lobe latéral. [48]

### *d. Transformée de Fourier rapide (FFT)*

Le but de la transformation de Fourier rapide (FFT) est de décomposer le signal en un signal sinusoïdal, qui se compose de deux unités, à savoir les unités réelles et les unités imaginaires. FFT Utilisé pour l'analyse fréquentielle, simplifiant ainsi le traitement de la voix car il correspond à l'audition

humaine. FFT Est un algorithme qui implémente la transformée de Fouries discrète (DFT). DFT est la transformation de chaque trame avec N échantillons Du domaine temporel au domaine fréquentiel. [48]

$$X_n = \sum_{k=0}^{N-1} X e(-2\pi j k \frac{n}{N})$$

avec : N : Le nombre de segments de trame, X : Valeur de données à k

#### *e. Enveloppe de fréquence Mel*

Enveloppe de fréquence Mel est généralement effectué en utilisant un banc de filtres « Filterbank » (c'est une forme de filtre conçu dans le but de déterminer la taille de l'énergie d'une bande de fréquence particulière en fréquence, mais à des fins MFCC), qui doit être appliqué dans le domaine fréquentiel. [48]

#### *f. Log*

Les valeurs logarithmiques sont obtenues en convertissant les valeurs DFT en une seule valeur. Réduisez la valeur de la banque de filtres mel en substituant chaque valeur de journal de base. En utilisant la commande Matlab "log" pour extraire la valeur logarithmique du filtre mel sur le segment sonore. L'oreille humaine est moins sensible pour entendre des sons de faible amplitude et de faible hauteur. [48]

#### *g. Transformé cosinus discrète (DCT)*

La DCT est la dernière étape du processus principal d'extraction de caractéristiques MFCC. Le concept de base de la DCT est de corrélér la valeur du spectre de mel afin de produire une bonne représentation de la propriété spectrale locale. [48]

#### *h. Cepstrum*

Le Cepstre est la valeur inverse du spectre. Cepstre Couramment utilisé pour obtenir des informations à partir d'un signal vocal prononcé par des humains. Dans cette dernière étape, le log mel du spectre converti en Cepstrums utilise la transformée en cosinus discrète (DCT) qui est la valeur de la fréquence mel résultante convertie en un domaine temporel. [48]



## 5. Apprentissage profond (Deep learning)

Depuis les années 2000, l'apprentissage profond est apparu comme une nouvelle zone de l'intelligence artificielle. Son architecture se base sur un réseau de neurones à plusieurs couches cachées dans le but d'imiter le fonctionnement du cerveau humain où un certain poids est assigné à différents neurones. Plus le nombre de couches cachées est élevé, plus le réseau est profond. [49]

L'apprentissage profond a connu un vaste succès en montrant une performance exceptionnelle dans différents domaines d'application en matière de vision par ordinateur, de reconnaissance de la parole, de traitement du langage naturel et de reconnaissance audio, etc. [50]

Les architectures les plus connues dans l'apprentissage profond sont Les réseaux de neurones convolutifs (détailler par la suite) grâce à leurs résultats impressionnants.

## 6. Les réseaux de neurones convolutifs

Les réseaux convolutifs sont une puissante application de l'intelligence artificielle (IA) au traitement des images, qui s'appuie sur l'apprentissage profond pour effectuer des tâches descriptives et génératives. Ils sont simplement des réseaux de neurones qui utilisent la convolution à la place de la multiplication de matrice générale dans au moins une de leurs couches. [51]

La structure d'un CNN est généralement constituée de quatre types de couche : la couche de convolution, la couche de pooling, la couche d'aplatissement et La couche entièrement connectée. Chaque couche de convolution est accompagnée par une fonction d'activation ReLU. [51]

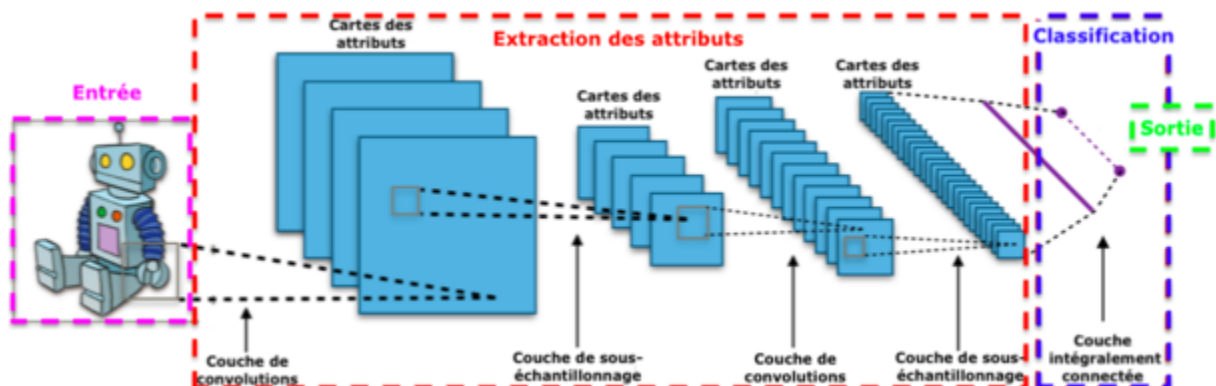


Figure 20: diagramme de réseaux de neurones convolutif

## 6.1. Les couches convolutives

La couche de convolution est la composante clé des réseaux de neurones convolutifs qui constitue les noyaux du réseau convolutif. Ces noyaux sont générés par le réseau compte tenu des caractéristiques (formes, couleurs, textures, etc.) qu'il souhaite extraire des images. [49]

Une convolution est une opération linéaire qui implique la multiplication de l'entrée avec l'ensemble de pondérations.

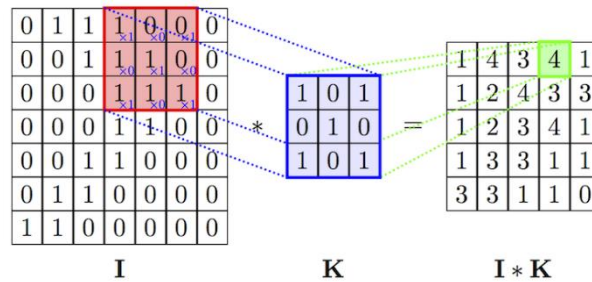


Figure 21 : représentation matricielle du produit de convolution [49]

La phase de convolution est ainsi formulée comme suit : [49]

$$\begin{aligned}
 (f * g)(x) &\equiv \int_{-\infty}^{+\infty} f(x-t)g(t)dt \\
 &= \int_{+\infty}^{-\infty} f(T)g(x-T)d(-T) \\
 &= \int_{-\infty}^{+\infty} f(T)g(x-T)dT \\
 &\equiv (g * f)(x)
 \end{aligned}$$

avec  $f$  : la fonction des entrées, et  $g$  représente la fonction du noyau.

La convolution est décrite comme une moyenne pondérée de  $f$ , avec  $g$  la pondération associée à  $f$ .

## 6.2. La fonction d'activation ReLU

Cette fonction permet d'effectuer un filtre sur nos données, est appliquée juste après la phase de convolution afin de réduire les calculs. Elle laisse passer les valeurs positives et les valeurs négatives sont ramenées à 0. Il y a aucun échange sur la taille des caractéristiques. [52]

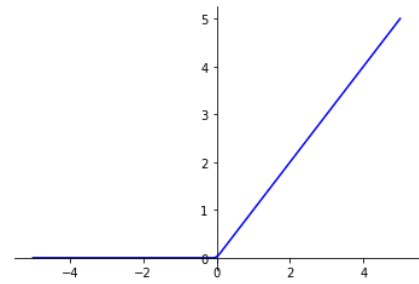


Figure 22 : graphe de filtre d'activation ReLU [52]

## 6.3. La couche pooling

La couche de pooling obtient une variance spatiale en réduisant la résolution des cartes de caractéristiques en préservant les informations les plus importantes qu'elle contient. Également, cette opération a un rôle de réduction du risque de sur-apprentissage. [50]

Les types de pooling les plus populaires sont le max et l'average pooling. [50]

Le pooling max : Chaque opération de pooling sélectionne la valeur maximale de la surface, et au même temps garder les caractéristiques détectées.

Le pooling average : Chaque opération de pooling Fait la moyenne de l'ensemble des valeurs sélectionnées.

## 6.4. La couche d'aplatissement

L'aplatissement consiste à convertir les données en un tableau unidimensionnel pour les introduire dans la couche suivante. Nous aplatissons la sortie des couches de convolution pour créer un seul long vecteur de caractéristiques. Et il est connecté au modèle de classification final, qui est appelé une couche entièrement connectée. [53]

## 6.5. La couche entièrement connectée

La couche entièrement connectée constitue toujours la dernière couche d'un réseau de neurones, convolutif ou non. Celui-ci est composé de deux parties, La particularité de la première partie est que tous ses neurones sont connectés avec tous les neurones de la couche précédente, mais aussi avec tous les neurones de la couche suivante. La deuxième partie est utilisée pour produire une distribution de

probabilité entre les différentes classes (chaque classe aura une valeur réelle comprise dans l'intervalle [0, 1]), et c'est fait par connecter la dernière couche à un softmax. [50]

✚ **Softmax** : La fonction Softmax permet de transformer un vecteur réel en vecteur de probabilité. On l'utilise souvent dans la couche finale d'un modèle de classification, notamment pour les problèmes multiclasse. Dans la fonction Softmax, chaque vecteur est traité indépendamment. L'argument axis définit l'axe d'entrée sur lequel la fonction est appliquée. [52]

## 7. La précision

La précision ou « accuracy » représente le ratio entre le nombre d'échantillon bien prédit et le nombre total d'échantillon. Elle est donnée par la formule suivante : [50]

$$ACC = \max \frac{\sum_{i=1}^N 1(r_i = m(c_i))}{N}$$

avec  $r_i$  : le vrai label de l'observation  $i$ .  $C_i$  : la classe contenant  $i$ , et  $m(.)$  : l'ensemble des possibilités d'affectation des observations dans les classes.

## 8. L'optimisation Adam

Adam est un algorithme d'optimisation qui peut être utilisé à la place de la procédure classique (l'optimisation par gradient de descente) pour mettre à jour les poids du réseau de manière itérative en fonction des données d'apprentissage afin de minimiser l'erreur (augmenter la précision). [54]

Le tableau (2) représente un exemple de différence de précision avant et après l'utilisation d'algorithme d'optimisation Adam.

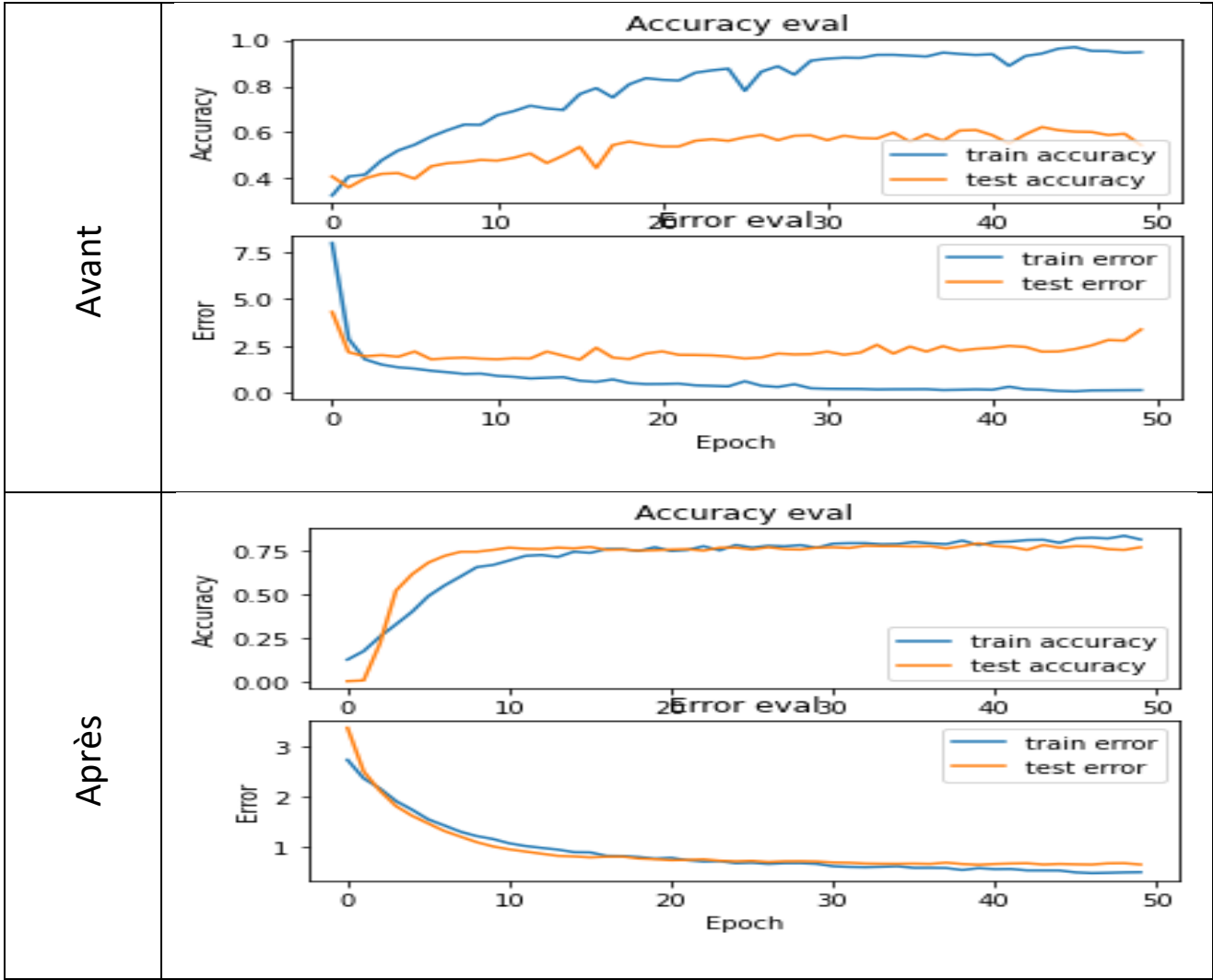


Tableau 2: la précision avant et après l'optimisation

## Chapitre IV : réalisation et implémentation

### 1. Introduction

Le chapitre présent devant nous expose la réalisation et l'implémentation de la méthode utilisée pour réaliser notre application.

### 2. Architecture du back-end de logiciel

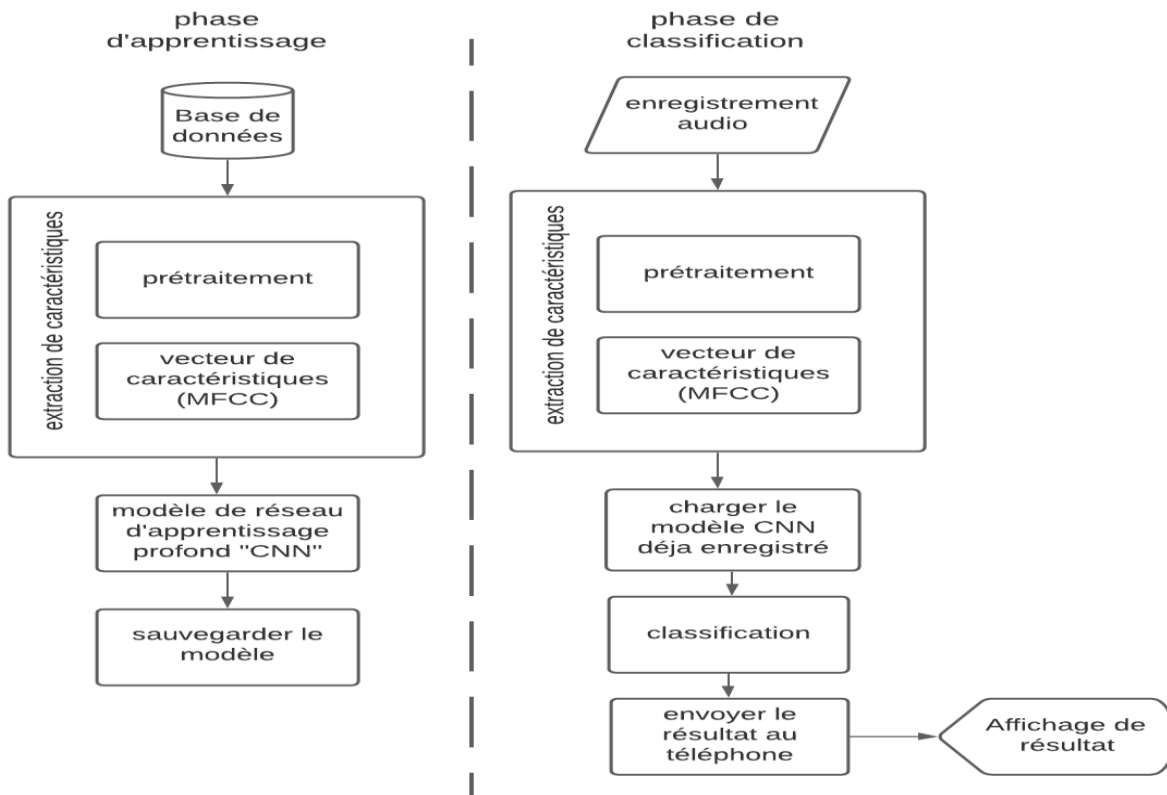


Figure 23: organigramme de back-end de "Baby Crying Classifier"

### 3. Première partie de projet

#### 3.1. Etape 01 : Préparation de base de données

Nous avons téléchargé notre base de données sur le site web cité en référence [55] .

Donateacry-corpus : c'est un corpus audio de pleurs de nourrissons qui a été construit grâce à la campagne Donate-a-cry (n'est plus active).

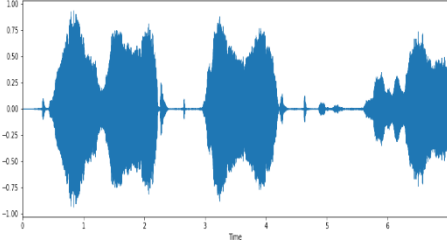
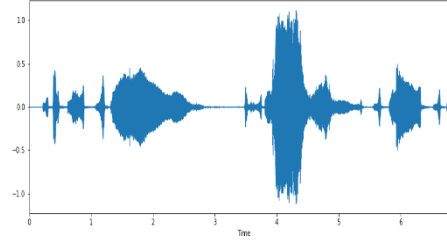
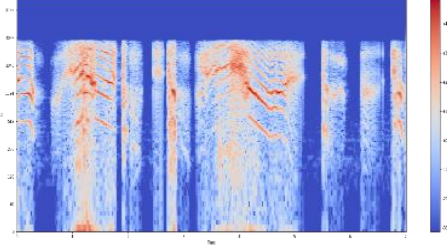
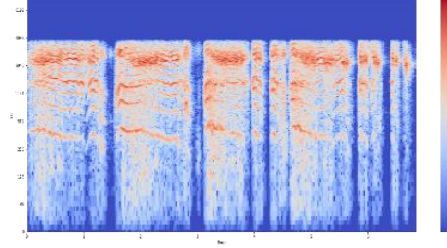
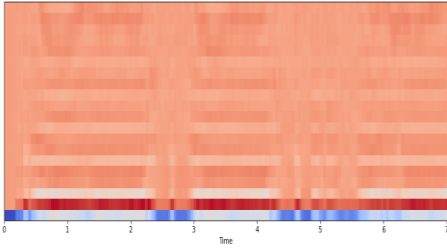
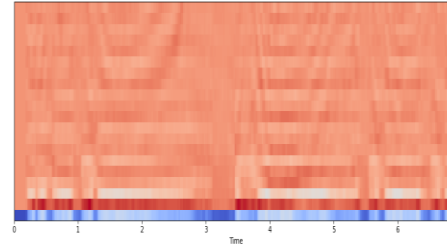
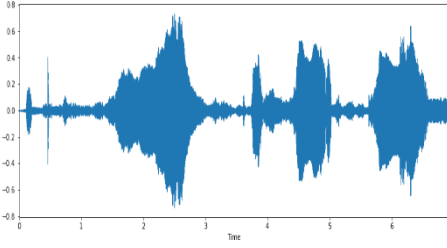
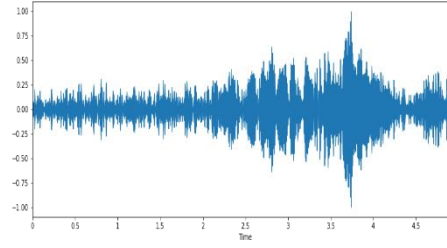
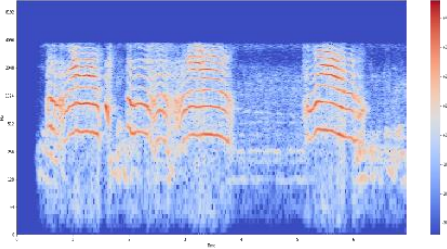
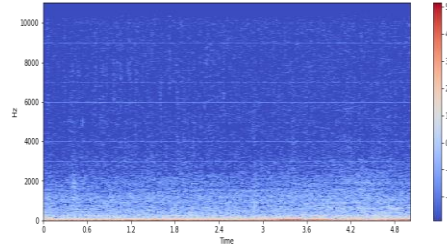
Ce dépôt contient des échantillons audios téléchargés par les utilisateurs dans leur forme originale, non modifiée et non vérifiée. Les échantillons audios ont été téléchargés à l'aide des applications mobiles Donate-a-cry pour Android et iOS.

Après filtrage de ces échantillons audio, nous avons obtenu quatre (04) catégories de données (le silence, la faim, l'inconfort et la douleur).

Nous descripteurs de chaque audio dans la base de données sont extrait par la méthode MFCC utilisant la fenêtre de Hamming avec nfft de 2048 et un chevauchement de 512, puis sauvegarder en fichier json (.json).



Figure 24: fichier json

Classe	Inconfort	Douleur
<b>Signal</b>		
<b>Spectrogramme</b>		
<b>MFCC</b>		
Classe	Faim	Silence
<b>Signal</b>		
<b>Spectrogramme</b>		



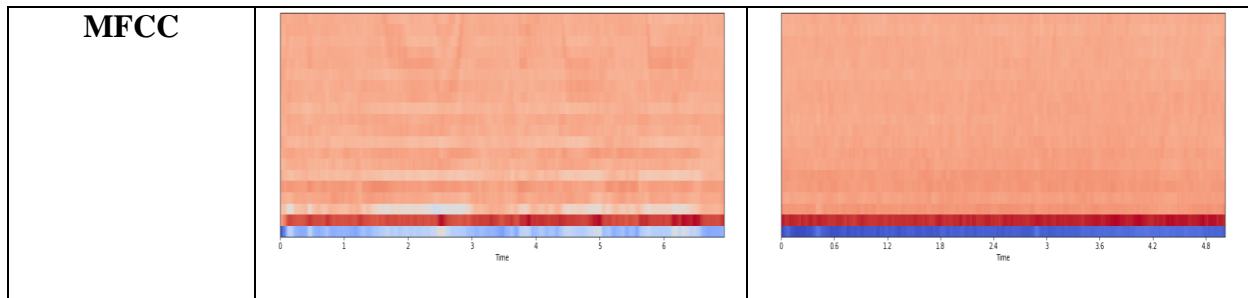


Tableau 3 : présentation de spectrogramme et MFCC pour chaque classe

### 3.2. Etape 02 : Création d'un modèle CNN

Le tableau suivant expose les paramètres choisis pour construire notre modèle de réseaux CNN utilisé dans notre application :

```

Model: "sequential_18"
Layer (type)                Output Shape                Param #
-----
conv2d_15 (Conv2D)          (None, 20, 11, 32)         320
max_pooling2d_14 (MaxPooling (None, 10, 6, 32)         0
batch_normalization_14 (Batc (None, 10, 6, 32)         128
conv2d_16 (Conv2D)          (None, 8, 4, 32)           9248
max_pooling2d_15 (MaxPooling (None, 4, 2, 32)           0
batch_normalization_15 (Batc (None, 4, 2, 32)           128
conv2d_17 (Conv2D)          (None, 3, 1, 32)           4128
max_pooling2d_16 (MaxPooling (None, 2, 1, 32)           0
batch_normalization_16 (Batc (None, 2, 1, 32)           128
flatten_17 (Flatten)        (None, 64)                  0
dense_60 (Dense)            (None, 64)                  4160
dropout_10 (Dropout)        (None, 64)                  0
dense_61 (Dense)            (None, 3)                   195
-----
Total params: 18,435
Trainable params: 18,243
Non-trainable params: 192

```

Figure 25: paramétrage du modèle de CNN

Conv2D	Crée une convolution de kernel qui est convolué avec la couche d'entrée pour produire un "tensor" de sorties [56]
Maxpooling	Un processus de discrétisation basé sur des échantillons, son objectif est de sous-échantillonner une représentation d'entrée (image, matrice de sortie de couche cachée, etc.) en réduisant sa dimension. [57]
Batch normalisation	Batch normalization est une technique d'apprentissage superviser qui converti l'inter couche de sortie du réseau neurone a une forme standard
Flatten	Les couches d'aplatissement modifient la forme des données d'un vecteur de matrices 2d (ou de matrices nd) en un format correct interprétable par la couche dense. Cette transformation permet aux données d'être exploitables par ce type de couche différent.
Dense	La couche dense est la couche de réseau de neurones profondément connectée. Il s'agit de la couche la plus courante et la plus fréquemment utilisée. La couche dense effectue les opérations sur l'entrée et renvoie les résultats en sortie. [58]

Tableau 4: définition des termes de paramétrage

### 3.3. Etape 03 : l'apprentissage pour le model

Avant d'entamer l'étape d'apprentissage on doit passer par une conversion de cette dernière en json file déjà citer en haut ; output (data. Json) et ensuite diviser cette base de données en 2/3 1/3 (2/3 des données pour l'apprentissage, et 1/3 pour le test)

Après avoir traité les données et lancer l'étape d'apprentissage viens l'étape de test ou on utilise les 1/3 des donnée restant, les résultats sont comme suit :

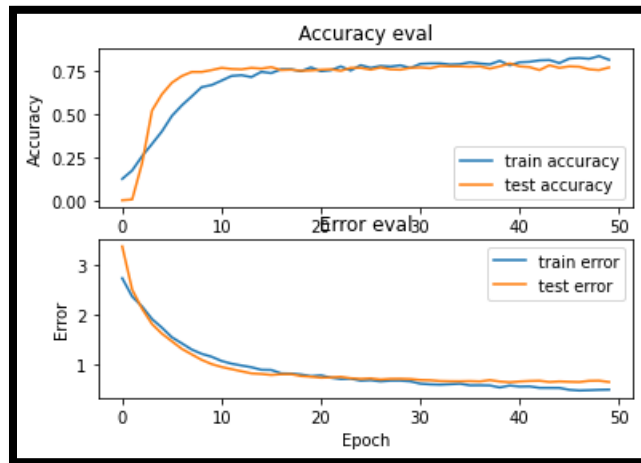


Figure 26: graphe d'erreur et de précision

Test accuracy: 0.7432432174682617

Notre modèle est de précision de 74%, ce qui est acceptable.

### 3.4. Etape 04 : Enregistrement de son

Après avoir sauvegarder le modèle, viens l'étape d'enregistrement pour traiter le crie du bebe, notre application utilise la bibliothèque Python « sounddevice » disponible en open source sur web.

Le son est enregistré sur une longueur de 5 secondes avec une fréquence d'échantillonnage de 22050Hz utilisant un microphone USB à haute sensibilité. Le son enregistré est exporté dans un fichier wave de type stéréo pour être préparé à l'étape suivante.

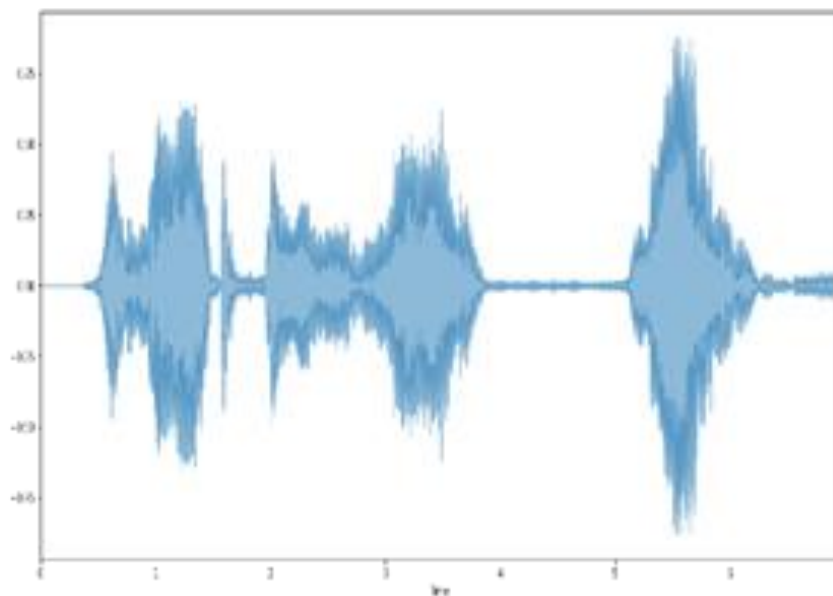


Figure 27: représentation de signal enregistré

### 3.5. Etape 05 : Extraction des paramètres MFCC et conversion en jsonfile

Après acquisition du signal vocal du cri du bébé à l'aide d'un microphone, celui-ci est enregistré en format wave. Par la suite vient l'étape d'extraction des paramètres MFCC (fréquentielle) une étape cruciale et importante car elle transforme le signal en un spectrogramme (figure 26) pour être utilisé sous forme de matrice dans le jsonfile et aussi puisque notre modèle (CNN) est mieux applicable sur des

images. Le spectrogramme extrait a partir du signal vocal est une matrice qu'on sauvegarde sous la forme d'un jsonfile.

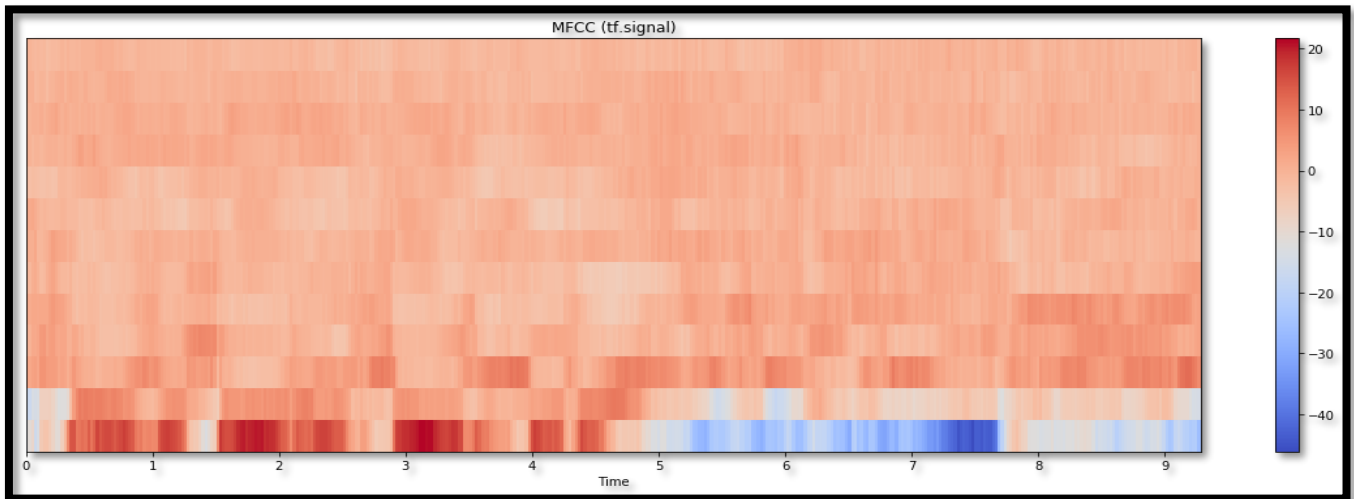


Figure 28: spectrogramme de signal enregistré

### 3.6. Etape 06 : Prédiction et envoi du résultat de la classification au smartphone

Dans cette partie nous rappelons le modèle (CNN) qui a été créé dans une partie précédente. Ensuite nous appliquons le jsonfile de notre enregistrement audio au modèle réalisé pour faire une prédiction de la classe (faim, malaise, douleur, silence).

Après avoir obtenu le résultat est affiché et envoyer au téléphone par SMS en utilisant un compte Twilio.

#### C'est quoi le Twilio ?

Twilio est une entreprise californienne créée en 2007. Présente dans plus de 15 pays avec 24 bureaux, elle est le leader du marché de l'envoi de sms dans le monde.

Twilio propose un logiciel pour envoyer des SMS automatiquement grâce à une API. L'outil s'adresse à des entreprises technologiques et particulièrement aux développeurs. [59]



## 4. Deuxième partie du projet

### 4.1. Initialisation et configuration du Raspberry pi

Par utilisation du logiciel « Windows on Raspberry » nous avons installé un système d'exploitation « Windows 10 pro » sur une carte mémoire MicroSD de 32 Gb.

Le premier démarrage du Raspberry pi va nécessiter un écran avec entrée HDMI, un clavier ainsi qu'une souris. Tout d'abord, nous branchons le câble HDMI, la souris et le clavier au Raspberry Pi en utilisant les ports HDMI et USB. Après ça on peut démarrer le Raspberry Pi en le branchant à une alimentation 5V, 2A.

Maintenant, nous installons l'interface logicielle « Anaconda Navigator » et les bibliothèques de python nécessaires pour lancer notre programme sur la carte Raspberry pi.

La figure 29 présente notre système hardware.

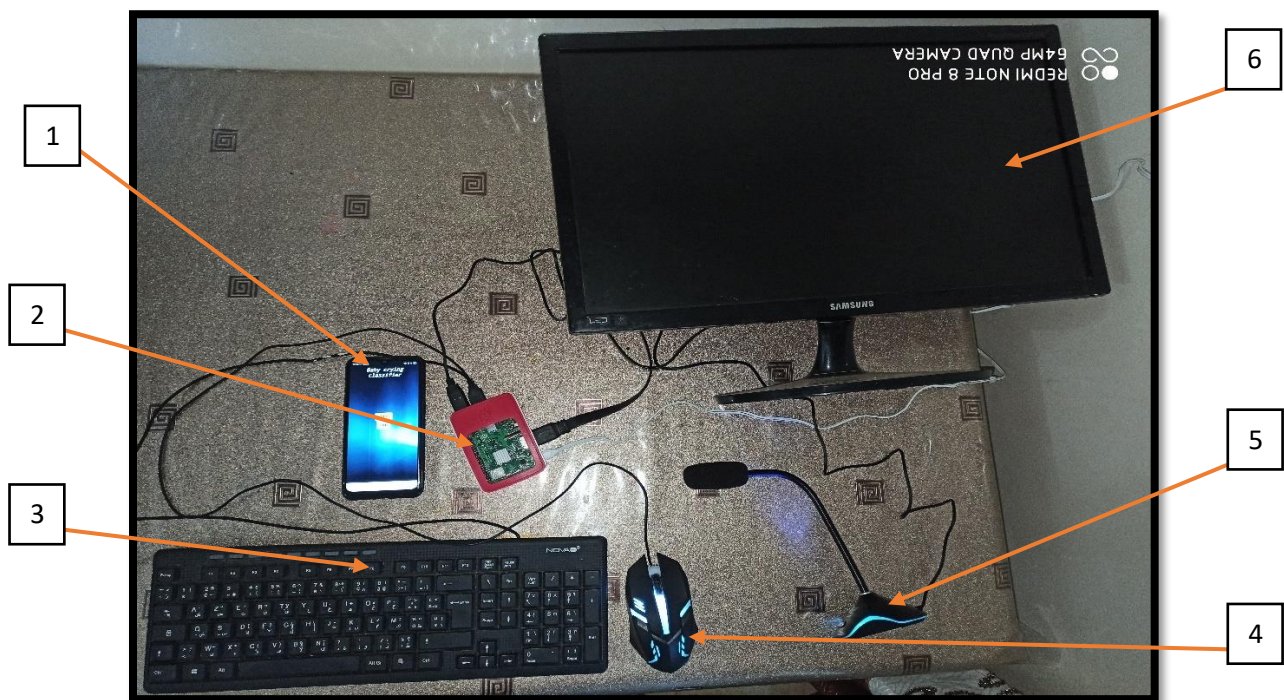


Figure 29: le système hardware du projet

1	Téléphone portable pour afficher les résultats
2	Raspberry pi 3 (modèle B)
3	Un clavier
4	Une souris
5	Microphone USB à haut sensibilité
6	Un écran

## 4.2. Application « Baby Crying Classifier »

L'interface d'application :



Figure 30: interface principale de l'application mobile



Figure 31: deuxième interface de l'application mobile

1	Bouton du lancement (pour démarrer l'application)
2	Bouton pour sortir de l'application
3	Bouton pour revenir au menu principal (page d'accueil)
4	Espace d'affichage de l'historique des précédente prédiction
5	Représentation textuelle de la prédiction courante
6	Représentation graphique de la prédiction courante



## 5. Conclusion

Dans ce chapitre nous avons présenté les résultats de notre projet de fin d'étude. Ce dernier est divisé en trois parties majeures. La première partie traite la phase de de l'extraction des attributs par la méthode MFCC et l'utilisation des réseaux de neurones Convolutionnels CNN pour faire une classification du signal vocal. Dans la deuxième partie, nous avons expliqué la méthode de la configuration de Raspberry pi. La dernière partie explique le fonctionnement de l'application « baby crying classifier » crée sous App Inventor.

## Conclusion générale

Avoir un bébé change le mode de vie. Le premier-né rend les parents inquiéter à trouver comment prendre soin du lui. Le problème s'empire si les parents ont des déficits auditifs.

Le cadre de ce projet découle de l'intérêt porté sur la conception d'un prototype destiné à détecter les pleurs de bébé et prédire la cause de ces pleurs.

Dans ce mémoire nous nous sommes intéressés aux propriétés et caractéristiques du son en particulier les pleurs de bébé.

La première étape de notre projet a été entamée par une acquisition du signal vocal en utilisant un microphone. Le principal volet de notre travail a été fait en combinant la méthode du traitement et extraction des caractéristiques du signal vocal MFCC associée aux réseaux de neurones convolutifs CNN. Les résultats de classification obtenus par cette approche sont très satisfaisants. Ce modèle permet de classifier le signal vocal du bébé en quatre classes : la douleur, l'inconfort, la faim et le silence.

Le résultat de la classification (état du bébé) est transmis directement par un SMS au téléphone portable. En plus du message une photo s'affiche montrant l'état du bébé.

Notons enfin que ce travail s'inscrit dans le contexte de l'aide les mamans à s'occuper leur enfant sans souci.

Après avoir achevé les étapes citées, nous pouvons dire que nous avons enrichi nos connaissances en apprentissage profond (Deep Learning) et que ce projet réalisé nous a été grandement profitable.

En raison de la pandémie du Covid 19, nous nous sommes confrontés à plusieurs problèmes. Nous n'avons pas pu créer une base de données dans notre région afin de déterminer la meilleure perception chez les algériens. Même sur internet, l'obtenir d'une base de données plus précise et bien filtré afin d'avoir un meilleur résultat était très difficile.

Nous avons également eu un problème en termes d'acquisition. L'acquisition avec le Raspberry Pi model B ne fonctionne pas car notre système a besoin d'une RAM avec des caractéristiques et capacité élevée (4 Gb en minimum).

Nous avons cherché une autre solution, qui est l'intégration de Raspberry avec un Jeston Nano, mais cette carte n'est pas disponible en Algérie.

Enfin, nous espérons sincèrement que ce projet pourra servir de base à d'autres recherches approfondies utilisant d'autres méthodes de caractérisation et d'autres approches de classification plus avancées, Avec des moyens d'acquisition plus puissants.

## Références

- [1] G.V.I.S. Silva, D.S. Wickramasinghe, «Infant Cry Detection System with Automatic Soothing and Video Monitoring Functions,» vol. 5, 2017.
- [2] équipe naitre et grandir, "les pleurs de bébé," juin 2016. [Online]. Available: [https://naitreetgrandir.com/fr/etape/0\\_12\\_mois/soins/fiche.aspx?doc=naitre-grandir-bebe-pleurs-comprendre](https://naitreetgrandir.com/fr/etape/0_12_mois/soins/fiche.aspx?doc=naitre-grandir-bebe-pleurs-comprendre).
- [3] Yizhar L., Coheny R., Ruinskiy D, "Baby Cry Detection in Domestic Environment using Deep Learning," 2016.
- [4] R. Haridy, "UCLA smartphone app translates meaning of your baby's cries," mai 2018. [Online]. Available: <https://newatlas.com/baby-cry-translation-research-autism-study/54722?fbclid=IwAR0I13NwMR3OGQThHDEUgp2Sh2M3ks7p7IcgiXw4InHdPuXY143HW2e6ffc>. [Accessed 2018].
- [5] Y, Pan C. Zhao W. Deng S. Wei W. Zhang Y. Xu, "The Methods of Realizing Baby Crying Recognition and Intelligent Monitoring Based on DNN-GMM-HMM," 2018.
- [6] Deepli B., Gokul V., Kavyashree S., Pawan K, "Smart Cradle For Baby Using FN-M16P Module," vol. 2, no. 2566-932X, 2019.
- [7] MEZOUAR B, BOUREGA M, "Smart infant cry detection system with raspberry pi," Mémoire de projet de fin d'étude, spécialité : Instrumentation Biomédicale, faculté de Technologie, université Abou Bakr Belkaid, Tlemcen, 2020.
- [8] [Online]. Available: <https://www.who.int/fr/news-room/fact-sheets/detail/deafness-and-hearing-loss> .
- [9] N. CHIALI, «le son,» unité des exploration ORL, Oran.
- [10] InternetArchiveBot, «Son,» 17 février 2020. [En ligne]. Available: [https://fr.wikipedia.org/wiki/Son\\_\(physique\)](https://fr.wikipedia.org/wiki/Son_(physique)).

- [11] "Mesurer une période à partir d'une courbe," [Online]. Available: <https://www.kartable.fr/ressources/physique-chimie/methode/mesurer-une-periode-a-partir-dune-courbe/20318>.
- [12] the GreenFacts Scientific Board, [En ligne]. Available: <https://www.greenfacts.org/fr/glossaire/def/frequence.htm>.
- [13] «Longueur d'onde,» 2020. [En ligne]. Available: <https://www.futura-sciences.com/sciences/definitions/physique-longueur-onde-4575/>.
- [14] S. GOYAL, «What are the characteristics of Sound Waves?,» 7 mai 2018. [En ligne]. Available: <https://www.jagranjosh.com/general-knowledge/what-are-the-characteristics-of-sound-waves-1525678871-1>.
- [15] Michel, «Phase à l'origine d'une fonction sinusoidale – déphasage définition,» 4 septembre 2016. [En ligne]. Available: <http://www.math15minutes.fr/phase-a-l-origine-fonction-sinusoidale-dephasage-definition/>.
- [16] "Les différents types de son," weebly, [Online]. Available: <https://tpe-briser-le-verre-avec-la-voix.weebly.com/les-types-de-son.html>.
- [17] C. Lesueur, Rayonnement acoustique des structures, Paris: Eyrolles, 1988.
- [18] R. E. Berg, «Sound,» The Physics of Sound, 21 février 2020. [En ligne]. Available: <https://www.britannica.com/science/sound-physics>.
- [19] S. Asthana, N. Varma, V. K. Mittal, "Preliminary analysis of causes of infant cry," 2014.
- [20] I. Bgnicg, H. Cucu, A. Buzo, D. Burileanu and C. Burileanu, "Baby cry recognition in real-world conditions," Vienna, Austria, 2016.
- [21] Antonio Verduzco-Mendoza Emilio Arch-Tirado Carlos A. Reyes García Jaime Leybón Ibarra Juan Licona Bonilla, Qualitative and Quantitative Crying Analysis of New Born Babies Delivered Under High Risk Gestation, México: COST School, 2008.
- [22] Chunyan Ji, Thosini Bamunu Mudiyanseelage, Yutong Gao et Yi Pan, «A review of infant cry analysis and classification,» vol. 8, 2021.
- [23] «Anatomie et fonctionnement de l'oreille,» 2021. [En ligne]. Available: <http://www.fo-rothschild.fr/soins/orl/informations-medicales/anatomie-fonctionnement-oreille.html>.

- [24] N. Quentin, «Oreille,» juillet 2016. [En ligne]. Available: <https://www.passeportsante.net/fr/parties-corps/Fiche.aspx?doc=oreille>. [Accès le 2021].
- [25] D. ZERGOUN, "Etude et réalisation d'un audiomètre médical," Mémoire de projet de fin d'étude, spécialité : Instrumentation Biomédicale, faculté de Technologie, université Abou Bakr Belkaid, Tlemcen, 2018.
- [26] A. Masliah, «comprendre les différents types de surdité,» 5 décembre 2019. [En ligne]. Available: <https://www.lescentresmasliah.com/fr/blogue/comprendre-les-differents-types-de-surdite>.
- [27] «Surdité,» mars 2021. [En ligne]. Available: <https://fr.wikipedia.org/wiki/Surdit%C3%A9>.
- [28] "10 faits sur la surdité," 2 mars 2021. [Online]. Available: <https://www.who.int/fr/news-room/facts-in-pictures/detail/deafness>.
- [29] J. Berke, «Deaf People Parenting Hearing Children,» 9 janvier 2020. [En ligne]. Available: <https://www.verywellhealth.com/deaf-parents-with-hearing-children-1046779>.
- [30] "Surdité et déficience auditive," 2 mars 2021. [Online]. Available: <https://www.who.int/fr/news-room/fact-sheets/detail/deafness-and-hearing-loss>.
- [31] "Qu'est ce qu'un Raspberry Pi ?," 23 octobre 2015. [Online]. Available: <https://grafikart.fr/blog/raspberry-pi-utilisation>.
- [32] "Anaconda (distribution Python)," 9 avril 2021. [Online]. Available: [https://fr.wikipedia.org/wiki/Anaconda\\_\(distribution\\_Python\)](https://fr.wikipedia.org/wiki/Anaconda_(distribution_Python)).
- [33] L. Bastien, «Jupyter Notebook : tout savoir sur le notebook préféré des Data Scientists,» 16 mars 2021. [En ligne]. Available: <https://www.lebigdata.fr/jupyter-notebook>.
- [34] "Spyder (logiciel)," 20 mai 2021. [Online]. Available: [https://fr.wikipedia.org/wiki/Spyder\\_\(logiciel\)](https://fr.wikipedia.org/wiki/Spyder_(logiciel)).
- [35] "Python : qu'est-ce que c'est ?," 2021. [Online]. Available: <https://www.futura-sciences.com/tech/definitions/informatique-python-19349/>.
- [36] JDN, "Python : définition et utilisation de ce langage informatique," 31 mars 2020. [Online]. Available: <https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1445304-python-definition-et-utilisation-de-ce-langage-informatique/>.
- [37] "Introducing JSON," [Online]. Available: <https://www.json.org/json-en.html>.

- [38] A. Neishabouri, «Découvrez les bibliothèques Python pour la Data Science,» 15 décembre 2020. [En ligne]. Available: <https://openclassrooms.com/fr/courses/4452741-decouvrez-les-bibliothèques-python-pour-la-data-science/4740941-plongez-en-detail-dans-la-bibliothèque-numpy>.
- [39] A. Rai, "OS Module in Python with Examples," 26 mai 2021. [Online]. Available: <https://www.geeksforgeeks.org/os-module-python-examples/>.
- [40] S. Yegulalp, "What is TensorFlow? The machine learning library explained," 18 juin 2019. [Online]. Available: <https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html>.
- [41] N. Samuel, "The Python Math Library," 2 mai 2019. [Online]. Available: <https://stackabuse.com/the-python-math-library>.
- [42] "Librosa," 2021. [Online]. Available: <https://librosa.org/doc/latest/index.html>.
- [43] M. Geier, "sounddevice," 28 octobre 2015. [Online]. Available: <https://pypi.org/project/sounddevice/0.3.0/>.
- [44] "SciPy Introduction," [Online]. Available: [https://www.w3schools.com/python/scipy/scipy\\_intro.php](https://www.w3schools.com/python/scipy/scipy_intro.php).
- [45] 2021. [Online]. Available: <https://appinventor.mit.edu/about-us>.
- [46] N. Heath, «Windows 10 on the Raspberry Pi: What you need to know,» 11 août 2015. [En ligne]. Available: <https://www.techrepublic.com/article/windows-10-on-the-raspberry-pi-what-you-need-to-know/>.
- [47] I. Mish, "MFCC implementation and tutorial," 2018. [Online]. Available: <https://www.kaggle.com/ilyamich/mfcc-implementation-and-tutorial>.
- [48] Yesy Afrillia, et al, "Performance Measurement Of Mel Frequency Cepstral Coefficient (MFCC) Method In Learning System Of Al- Qur'an Based In Nagham Pattern Recognition," no. 930 012036, 2017.
- [49] Y. Ghanney, "L'apprentissage profond pour la détection des signaux d'interférence radio fréquentiels," Université du Québec à Montréal, Maîtrise en informatique., Mémoire. Montréal (Québec, Canada), 2019.
- [50] D. Marie-Chantal, "COMME EXIGENCE PARTIELLE DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES," Université du Québec à Trois-Rivières, Canada, 2008.

- [51] Techtarget, "Réseau neuronal convolutif," juin 2019. [Online]. Available: <https://www.lemagit.fr/definition/Reseau-neuronal-convolutif>.
- [52] T. Keldenich, "fonction d'activation, comment ca marche?," 8 février 2021. [Online]. Available: <https://inside-machinelearning.com/fonction-dactivation-comment-ca-marche-une-explication-simple/>.
- [53] J. Jeong, "The Most Intuitive and Easiest Guide for Convolutional Neural Network," 24 janvier 2019. [Online]. Available: <https://towardsdatascience.com/the-most-intuitive-and-easiest-guide-for-convolutional-neural-network-3607be47480>.
- [54] J. Brownlee, "Gentle Introduction to the Adam Optimization Algorithm for Deep Learning," Deep Learning Performance, 3 juillet 2017. [Online]. Available: <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>.
- [55] G. Veres, "donateacry-corpus," 1 septembre 2020. [Online]. Available: <https://github.com/gveres/donateacry-corpus>.
- [56] "Conv2d," [Online]. Available: [https://keras.io/api/layers/convolution\\_layers/convolution2d/](https://keras.io/api/layers/convolution_layers/convolution2d/).
- [57] "Max Pooling," [Online]. Available: <https://deeptai.org/machine-learning-glossary-and-terms/max-pooling>.
- [58] "Keras - Dense Layer," 2020. [Online]. Available: [https://www.tutorialspoint.com/keras/keras\\_dense\\_layer.htm](https://www.tutorialspoint.com/keras/keras_dense_layer.htm).
- [59] "Présentation de Twilio," 6 avril 2021. [Online]. Available: <https://tool-advisor.fr/sms-pro/comparatif/twilio/>.