

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À
L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

COMME EXIGENCE PARTIELLE
À L'OBTENTION DE LA MAÎTRISE EN GÉNIE
CONCENTRATION RÉSEAUX DE TÉLÉCOMMUNICATIONS
M. Ing.

PAR
Khedidja NEKAB

ÉTUDE COMPARATIVE DES DIFFÉRENTES MÉTHODES D'ESTIMATION DE LA
FRÉQUENCE FONDAMENTALE DES CRIS DES NOUVEAU-NÉS

MONTREAL, LE 12 DÉCEMBRE 2013

©Tous droits réservés, Khedidja Nekab, 2013

PRÉSENTATION DU JURY

CE MÉMOIRE A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE :

M. Chakib Tadj, directeur de mémoire
Département de génie électrique à l'École de technologie supérieure

M. Jérémie Voix, président du jury
Département de génie mécanique à l'École de technologie supérieure

M. Ghyslain Gagnon, membre du jury
Département de génie électrique à l'École de technologie supérieure

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT UN JURY PUBLIC

LE 02 DÉCEMBRE 2013

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Je souhaite en premier lieu remercier Monsieur Tadj Chakib, directeur de mon projet, pour m'avoir accueillie dans son équipe de recherche et guidée pendant la réalisation de mon projet. Je souhaite remercier également les membres de jury d'avoir pris le temps d'évaluer mon travail.

J'adresse aussi mes remerciements aux membres de l'équipe du laboratoire 2446 et notamment à Dahmani Jamal.

J'aimerais dédier ce travail à mes parents, mon frère et mes sœurs ainsi qu'à tous mes amis(es) et mes proches avec une attention particulière pour Fares Ryma et R.Herroug.

ÉTUDE COMPARATIVE DES DIFFÉRENTES MÉTHODES D'ESTIMATION DE LA FRÉQUENCE FONDAMENTALE DES CRIS DES NOUVEAU-NÉS

Khedidja NEKAB

RÉSUMÉ

Au cours des trente dernières années, diverses méthodes ont été développées afin de mieux détecter la fréquence fondamentale des cris des nouveau-nés. En effet, l'analyse de la variation de la fréquence fondamentale dans le temps d'un signal de cri chez les nouveau-nés fournit des informations fiables très utiles au diagnostic médical concernant l'état de santé des bébés.

Nous présentons dans ce rapport une étude comparative des résultats de plusieurs méthodes d'extraction de la fréquence fondamentale des signaux de cris chez les nouveau-nés.

Les quatre algorithmes choisis à l'implémentation dans Matlab sont deux algorithmes temporels, AMDF «*Magnitude Difference Function*» et ACF «*Auto-Correlation Function*», un algorithme fréquentiel HPS «*Harmonic Product Spectrum*» ainsi que l'algorithme SIFT modifié «*modified Simple Inverse Filtering Tracking*». Ce dernier et le logiciel Praat sont utilisés à titre de comparaison. Nous étudierons plusieurs segments de cri : phonique, de haute-fréquence, hyperphonique et non-voisé.

Après évaluation de la performance des différentes méthodes analysées, nous avons remarqué que l'ACF a échoué à estimer la fréquence hyperphonique. D'autre part, nous avons observé que l'algorithme HPS a montré une bonne précision quant à l'évaluation de la fréquence fondamentale phonique.

Nous avons aussi constaté que les algorithmes AMDF et SIFT modifié sont moins précis dans l'identification de la fréquence fondamentale phonique. Cela nous a conduit à développer un nouvel algorithme CAMDA «*Combined Average Magnitude Difference and Autocorrelation*» permettant une combinaison entre le SIFT modifié et l'AMDF pour corriger les erreurs produites lors de la détection des segments phoniques. Ce dernier a montré qu'il est capable de remédier aux problèmes posés par l'AMDF et le SIFT modifié lors de la détection des segments phoniques.

Mots-clés : traitement du signal, cri des nouveau-nés, fréquence fondamentale.

COMPARATIVE STUDY OF DIFFERENT METHODS FOR ESTIMATING THE FUNDAMENTAL FREQUENCY OF CRY SOUNDS NEW BORN

Khedidja NEKAB

ABSTRACT

Over the last thirty years, various methods have been developed to improve the estimation process of the fundamental frequency. Indeed, detection and time domain analysis of changes in the fundamental frequencies of the newborn infants' cries are necessary to obtain reliable information that helps pediatrician to make an accurate diagnosis.

We present in this report a comparative study of the three fundamental frequency extraction methods and their performances on the baby's cry signals. The aforementioned algorithms are AMDF «*Average Magnitude Difference Function*», ACF «*Auto-Correlation Function*» and frequency algorithm HPS «*Harmonic Product Spectrum*» which are implemented in the MATLAB environment. In addition, the modified SIFT algorithm «*modified Simple Inverse Filtering Tracking*» and the Praat software are used for comparison study only. We have studied several cry segments: phonic, high-pitch, hyperphonic, unvoiced and undefined. In order to evaluate the performance of different algorithms, we have considered five types of error; sound error, hyperphonic error dysphonic error, unvoiced error and undefined error. We also present the classification procedure of different cry segments depending on the error type.

We have observed that the ACF is not capable to estimate the hyper-phonic frequency. Also, we have observed that the HPS estimate the fundamental phonic frequency with a higher precision than other method.

In addition, the AMDF and the modified SIFT are not so precise for estimation of the fundamental phonic frequency.

Because of this low precision, we have developed a new algorithm named CAMDA «*Combined Average Magnitude Difference and Autocorrelation*», which is a combination of the modified SIFT and the AMDF. The goal of this new algorithm is to correct the error proceeded during the detection of the phonic segments. The CAMDA has showed its capability to correct the problem given by the AMDF and the SIFT modified during the detection of the phonic segments.

Keywords : signal processing, cry sounds new born, algorithms, fundamental frequency .

TABLE DES MATIÈRES

	Page
INTRODUCTION	1
CHAPITRE 1 ÉTAT DE L'ART	3
1.1 Le signal de parole chez les adultes et le signal de cri chez les nouveau-nés.....	3
1.1.1 Processus de production d'un signal de parole	3
1.1.2 Processus de production d'un signal de cri chez les nouveau-nés.....	5
1.2 Les points de différences entre le signal de parole et le cri	6
1.3 La complexité et la variété du cri d'un nouveau-né.....	6
1.3.1 Description d'un signal de cri chez les nouveau-nés	7
1.3.1.1 Les catégories de cris.....	9
1.4 Les problèmes de l'estimation de la fréquence fondamentale	10
1.5 Les algorithmes de détection de la fréquence fondamentale	13
1.5.1 Les algorithmes temporels d'extraction de la fréquence fondamentale....	14
1.5.1.1 L'algorithme temporel AMDF.....	14
1.5.1.2 L'algorithme temporel ACF.....	15
1.5.1.3 L'algorithme temporel SIFT modifié.....	17
1.5.2 Les algorithmes spectraux d'extraction de la fréquence fondamentale ...	18
1.5.2.1 L'algorithme HPS	18
1.5.3 Les algorithmes hybrides d'extraction de la fréquence fondamentale.....	21
1.6 Les études comparatives des algorithmes d'estimation de la fréquence fondamentale	22
1.6.1 Les paramètres utilisés pour l'évaluation de la performance des algorithmes.....	22
1.6.2 Le cas d'un signal de parole.....	24
1.6.3 Le cas d'un signal de cri chez les nouveau-nés	28
1.7 Conclusion	28
CHAPITRE 2 IMPLÉMENTATION DES ALGORITHMES DE DÉTECTION DE LA FRÉQUENCE FONDAMENTALE.....	31
2.1 Introduction.....	31
2.2 Schéma général d'un algorithme	31
2.3 Les quatre méthodes de détection de la fréquence fondamentale.....	32
2.3.1 Architecture de l'algorithme AMDF	33
2.3.2 Architecture de l'algorithme ACF	35
2.3.3 Architecture de l'algorithme HPS.....	36
2.3.4 Architecture de l'algorithme SIFT modifié	38
2.4 Conclusion	40
CHAPITRE 3 RÉALISATION DE L'ÉTUDE COMPARATIVE DES MÉTHODES D'EXTRACTION DE LA FRÉQUENCE FONDAMENTALE.....	41

3.1	Introduction.....	41
3.2	Description de la méthode de comparaison	43
3.3	Analyse des résultats expérimentaux.....	44
3.3.1	Analyse des résultats de l'AMDF	47
3.3.2	Analyse des résultats de l'ACF	51
3.3.3	Analyse des résultats de l'HPS	54
3.3.4	Analyse des résultats de SIFT modifié	56
3.3.5	Analyse des résultats du logiciel Praat.....	58
3.4	Conclusion	60
CHAPITRE 4	DÉVELOPPEMENT ET MISE EN ŒUVRE D'UN ALGORITHME HYBRIDE CAMDA POUR L'ESTIMATION DE LA FRÉQUENCE FONDAMENTALE.....	63
4.1	Introduction.....	63
4.1.1	Description de l'algorithme hybride CAMDA	64
4.1.2	Principe de fonctionnement de l'algorithme CAMDA.....	64
4.1.3	Discussion des résultats	66
4.2	Conclusion	69
CONCLUSION.....		71
LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES.....		73

LISTE DES FIGURES

	Page
Figure 1-1 Section horizontale du larynx présentant l'emplacement de la glotte.....	3
Figure 1-2 L'ensemble des organes responsables de la production du signal vocal	4
Figure 1-3 La représentation de voyelle «a» et la consonne fricative «s» respectivement dans le temps (a) (c) et la fréquence (b) (d)	5
Figure 1-4 Le conduit vocal d'un nourrisson comparé à celui d'un adulte	6
Figure 1-5 Représentation de 7 secondes d'enregistrement d'un cri de bébé.....	8
Figure 1-6 Spectrogramme des différents modes des cris (a) phoniques (b) hyperphonique (c) dysphonique Tirée de Kheddache et Tadj (2013a, p.798).....	9
Figure 1-7 Deux différentes mesures utilisées pour définir les marqueurs du pic.....	12
Figure 1-8 Schéma bloc de l'algorithme SIFT modifié.....	17
Figure 1-9 Décimation d'un signal échantillonné par $D=3$	20
Figure 1-10 La décimation utilisée par HPS avec les valeurs D (1, 2, 3, 4 et 5)	20
Figure 1-11 Exemple des valeurs de $F0s(n)$ et $F0e(n)$ du même segment	23
Figure 2-1 Les blocs essentiels d'un algorithme de détection de la fréquence fondamentale	31
Figure 2-2 Schéma bloc de l'algorithme AMDF.....	33
Figure 2-3 Schéma bloc de l'algorithme ACF	35
Figure 3-1 L'application de la validation croisée à cinq blocs sur la comparaison des techniques d'estimation de la fréquence fondamentale	41
Figure 3-2 Les modes des segments qui correspondent à l'intervalle de variation de la fréquence fondamentale	44
Figure 3-3 L'ordre de l'erreur hyperphonique Er_{Hy_pc} selon le mode du segment traité par le modèle de test.....	44
Figure 3-4 L'ordre de l'erreur haute-fréquence (Er_{Hi_pc}) selon le mode du segment traité par le modèle de test.....	45

Figure 3-5 L'ordre de l'erreur phonique (Er_Ph_pc) selon le mode du segment traité par le modèle de test	46
Figure 3-6 L'ordre de l'erreur non voisée (Er_Nv_pc) selon le mode du segment traité par le modèle de test	46
Figure 3-7 L'ordre de l'erreur non défini (Er_Nan_pc) selon le mode du segment traité par le modèle de test	47
Figure 3-8 Pourcentages de la moyenne des différentes erreurs produites par rapport à l'AMDF.....	48
Figure 3-9 L'erreur moyenne des quatre méthodes par type d'erreur	51
Figure 3-10 Pourcentages de la moyenne des différentes erreurs produites par rapport à l'ACF	52
Figure 3-11 L'erreur moyenne des quatre méthodes par type d'erreur	53
Figure 3-12 Pourcentages de la moyenne des différentes erreurs produites par rapport à l'HPS.....	54
Figure 3-13 L'erreur moyenne des quatre méthodes par type d'erreur	55
Figure 3-14 Pourcentages de la moyenne des différentes erreurs produites par rapport au SIFT modifié.....	56
Figure 3-15 L'erreur moyenne des quatre méthodes par type d'erreur	57
Figure 3-16 Pourcentages de la moyenne des différentes erreurs produites par rapport au Praat	59
Figure 3-17 L'erreur moyenne des quatre méthodes par type d'erreur	60
Figure 4-1 Principe de fonctionnement de l'algorithme CAMDA	65
Figure 4-2 Pourcentages de la moyenne des différentes erreurs.....	67
Figure 4-3 L'erreur moyenne des quatre méthodes par type d'erreurs.....	68

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

AMDF	Average Magnitude Difference Function
ACF	Auto-Correlation Function
AICD	Automatic Infant Cry Detection
C-AMDF	Clipping -Average Magnitude Diffrence Function
CAMDA	Combined Average Magnitude Difference and Autocorrelation
CEP	Cepstrum
DARD	DARD
Erreur(V/NV)	Erreur Voisée/Non-Voisée
Erreur(NV/V)	Erreur Non-Voisée/Voisée
Er_Hy_pci	Percentage Error Hyper-phonation order i
Er_Hy_pc1	Percentage Error Hyper-phonation 1
Er_Hy_pc2	Percentage Error Hyper-phonation 2
Er_Hy_pc3	Percentage Error Hyper-phonation 3
Er_Hy_pc4	Percentage Error Hyper-phonation 4
Er_Ph_pci	Percentage Error phonation order i
Er_Ph_pc1	Percentage Error phonation 1
Er_Ph_pc2	Percentage Error phonation 2
Er_Ph_pc3	Percentage Error phonation 3
Er_Ph_pc4	Percentage Error phonation 4
Er_Nv_pc	Pourcentage d'Erreur Non-voisé
Er_Nv_pc1	Pourcentage d'Erreur Non-voisé 1
Er_Nv_pc2	Pourcentage d'Erreur Non-voisé 2
Er_Nv_pc3	Pourcentage d'Erreur Non-voisé 3
Er_Nv_pc4	Pourcentage d'Erreur Non-voisé 4
Er_Nan_pci	Pourcentage d'Erreur Non-définis ordre i
Er_Nan_pc1	Pourcentage d'Erreur Non-définis 1
Er_Nan_pc2	Pourcentage d'Erreur Non-définis 2
Er_Nan_pc3	Pourcentage d'Erreur Non-définis 3

Er_Nan_pc4	Pourcentage d'Erreur Non-définis 4
Er_Hi_pci	Percentage Error High pitched order i
Er_Hi_pc1	Percentage Error High pitched 1
Er_Hi_pc2	Percentage Error High pitched 2
Er_Hi_pc3	Percentage Error High pitched 3
Er_Hi_pc4	Percentage Error High pitched 4
F _{0s}	Fréquence fondamentale Standard
F _{0e}	Fréquence fondamentale Estimée
F ₀	Fréquence fondamentale
Fe	Fréquence d'échantillonnage
HPS	Harmonic Product Spectrum
LPC	Linear Predictive Coding
MACC	Modified Autocorrelation Center-Clippin
M_err_Hi	Moyennes d'erreurs High pitched
M_err_Ph	Moyenne d'erreur Phonique
M_err_Nv	Moyenne d'erreur Non-voisé
M_err_Hy	Moyenne d'erreur Hyperphonique
M_err_Nan	Moyenne d'erreur Non-définis
MM _{Nv}	Moyennes de M_err_Nv
MM _{Hy}	Moyennes de M_err_Hy
MM _{Hi}	Moyennes de M_err_Hi
MM _{Ph}	Moyennes de M_err_Ph
MM _{Nan}	Moyennes de M_err_Nan
PDyi	Pourcentage haute-fréquence ordre i
PNai	Pourcentage Non-définis ordre i
PNvi	Pourcentage Non-voisé ordre i
PPhi	Pourcentage Phonique ordre i
PHyi	Pourcentage Hyperphonique ordre i
PPROC	Parallel Processing
PESEC	Paramètre d'Erreur Sans Erreur Commise

SIFT modified	modified Simple Inverse Filtering Tracking
STE	Short-Time Energy
TFD	Transformée de Fourier Discrète
VAD	Détection d'Activité Vocale

INTRODUCTION

Dans la littérature de traitement du signal sonore, la fréquence fondamentale est un paramètre de base caractérisant tout phénomène. Elle est en corrélation directe avec les maladies liées au fonctionnement du cerveau. Cette caractéristique se manifeste dès les premiers stades de développement de l'enfant. Ainsi, la précision de mesure de la fréquence fondamentale et de sa variation dans le temps est un élément essentiel pour l'obtention d'informations fiables sur l'état de santé des nouveau-nés.

Dans ce projet, la fréquence fondamentale est la fréquence la plus basse de la vibration des cordes vocales. Elle précise les segments vocaux dans un signal de parole ou de cri. La valeur de cette fréquence varie d'une personne à une autre, en fonction du sexe et de l'âge. En effet, elle varie de 80 à 200 Hz chez les hommes, de 150 à 450 Hz chez les femmes et de 200 à 600 Hz chez les enfants.

L'importance de la fréquence fondamentale a poussé les chercheurs à trouver des méthodes et des techniques pour détecter cette valeur avec une haute précision. De nombreux algorithmes ont été expérimentés. Cependant, les résultats obtenus par ces algorithmes ne se sont pas révélés entièrement satisfaisants. C'est la raison pour laquelle les chercheurs tentent continuellement d'améliorer le principe de fonctionnement de ces algorithmes.

Notre projet se situe dans cet axe de recherche d'amélioration des algorithmes de détection de la fréquence fondamentale. En effet, le but de ce mémoire est de faire une étude comparative sur les algorithmes de détection de la fréquence fondamentale déjà existants dans la littérature. Cette étude consiste à comprendre le principe de fonctionnement de chaque algorithme choisi et à identifier ses principales caractéristiques pour sélectionner l'algorithme le plus performant.

Notre travail consiste à utiliser une base de données de signaux de cris des nouveau-nés afin d'extraire la fréquence fondamentale de chaque signal selon les différents algorithmes AMDF «*Average Magnitude Difference Function*», ACF «*Auto-Correlation Function*», HPS «*Harmonic Product Spectrum*», SIFT modifiée «*modified Simple Inverse Filtering Tracking*» et le logiciel Praat. Ensuite, des paramètres de comparaison seront appliqués pour mesurer la performance de ces techniques.

Par la suite, un nouvel algorithme CAMDA est proposé pour mieux détecter la fréquence fondamentale et réduire les faiblesses de l'algorithme AMDF dans l'identification des segments phoniques.

Structure du document

Ce mémoire est organisé de la manière suivante.

Le premier chapitre est composé d'un bilan de la littérature. Il présente les différentes recherches sur le sujet.

Le second chapitre traite de l'implémentation des différents algorithmes AMDF, ACF, HPS et SIFT modifié en décrivant le processus d'estimation de la fréquence fondamentale pour chaque algorithme sous forme de schémas blocs.

Le troisième chapitre sera entièrement consacré à l'analyse et la discussion des résultats de simulation des différentes méthodes utilisées.

Le quatrième chapitre présente notre contribution originale dans ce projet par le développement et la mise en œuvre d'un algorithme hybride intitulé CAMDA.

CHAPITRE 1

ÉTAT DE L'ART

1.1 Le signal de parole chez les adultes et le signal de cri chez les nouveau-nés

Dans ce chapitre, nous entamons une description générale de la génération du signal de cri de parole chez les adultes et les nouveau-nés, tout en citant les caractéristiques essentielles pour chacun d'eux.

1.1.1 Processus de production d'un signal de parole

La production de la parole est basée généralement sur la quantité d'air poussée par l'appareil respiratoire à travers la trachée-artère. Une pression d'air est modulée au niveau du larynx avant d'être appliquée au conduit vocal (Boite et al., 2000).

Le larynx est l'organe qui assure le passage de l'air vers les poumons lors de l'inspiration. Il contient les cordes vocales et les muscles responsables de la modulation des sons. En effet, les cordes vocales sont deux lèvres symétriques qui peuvent fermer complètement le larynx. Le repoussement progressif de ces deux lèvres crée une ouverture triangulaire appelée la glotte (Figure 1.1).

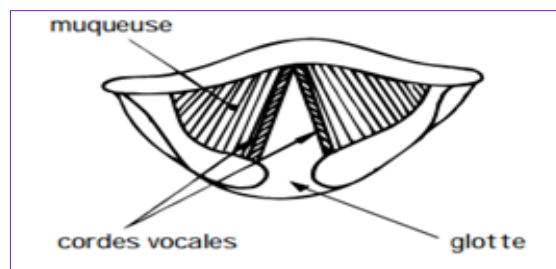


Figure 1-1 Section horizontale du larynx présentant
l'emplacement de la glotte
Tirée de Boite et al (2000, p.14)

Le passage de l'air crée, au niveau des cordes vocales, des vibrations qui peuvent divulguer la fréquence fondamentale du signal parole. Les différents organes responsables de la production des phonèmes sont présentés dans la Figure 1.2.

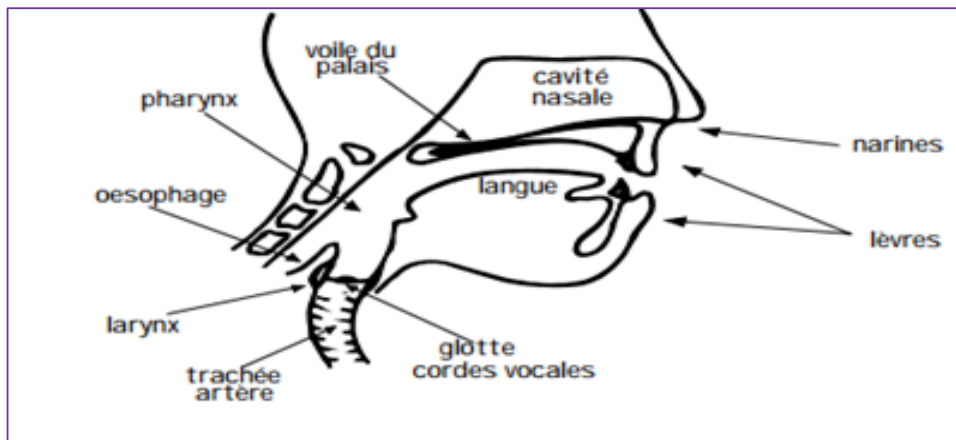


Figure 1-2 L'ensemble des organes responsables de la production du signal vocal
Tirée de Boite et al (2000, p.15)

Il existe deux catégories de phonèmes dans la parole : les phonèmes voisés et non voisés, représentés respectivement par les voyelles et les consonnes fricatives. En effet, la production des sons est contrôlée par le débit de l'air expiré, la tension sur les cordes vocales et la forme du conduit vocal. Pour illustrer les deux catégories, une représentation fréquentielle et temporelle est indiquée dans la Figure 1.3 (Valin et Lefebvre, 2000) pour une voyelle « a » puis une consonne fricative « s ».

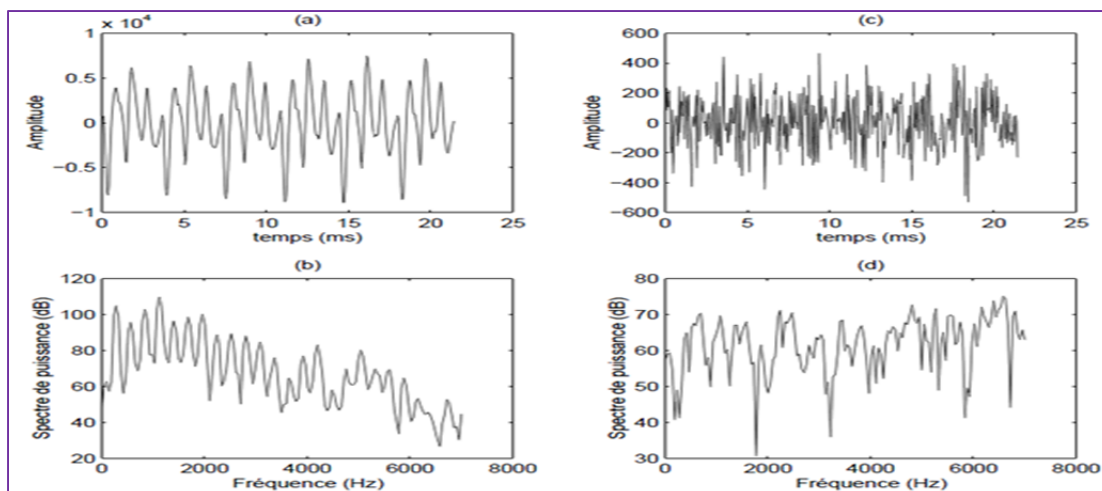


Figure 1-3 La représentation de voyelle «a» et la consonne fricative «s» respectivement dans le temps (a) (c) et la fréquence (b) (d)
Tirée de Valin (2001, p.6)

Dans le cas de la voyelle «a», l'ouverture et la fermeture de la glotte sont régulières dans le temps et par conséquent, le signal produit est périodique. En revanche, la glotte reste ouverte durant la production de la consonne fricative «s», produisant un signal non périodique (Valin, 2001).

1.1.2 Processus de production d'un signal de cri chez les nouveau-nés

Le cri est le seul moyen de communication entre le nouveau-né et son environnement. Il a été prouvé par plusieurs recherches que les caractéristiques acoustiques du cri sont influencées directement par l'état physique ou psychologique du nouveau-né ou par divers facteurs externes (Lederman, 2002).

La production du signal de cri chez les nouveau-nés ressemble au processus de production du signal de parole chez les adultes. La première étape de la production du cri est initiée dans le cerveau du nouveau-né par des facteurs externes ou internes comme la douleur ou la faim par exemple, sous forme d'une commande. La deuxième étape permet de traduire cette commande en une série d'impulsions à travers le système nerveux de la parole et les organes respiratoires qui sont responsables de la création des sons au niveau physiologique.

Le processus de la production de cri consiste à expirer l'air des poumons vers le conduit vocal et les cordes vocales, soit un processus identique à celui de la production de la parole (Valin, 2001).

1.2 Les points de différences entre le signal de parole et le cri

La forme du conduit vocal chez les nourrissons diffère de celui des adultes au niveau de l'organe qui s'appelle «pharynx». Ce dernier forme une pente douce descendante vers la glotte chez les nourrissons et un angle droit chez les adultes. Ceci conduit à une diminution de l'espace pharyngé chez les nouveau-nés (Figure 1.4). Donc, cet espace qui constitue un résonateur très important pour la production des sons, est nettement différent chez un nourrisson et un adulte (MARC, 2009).

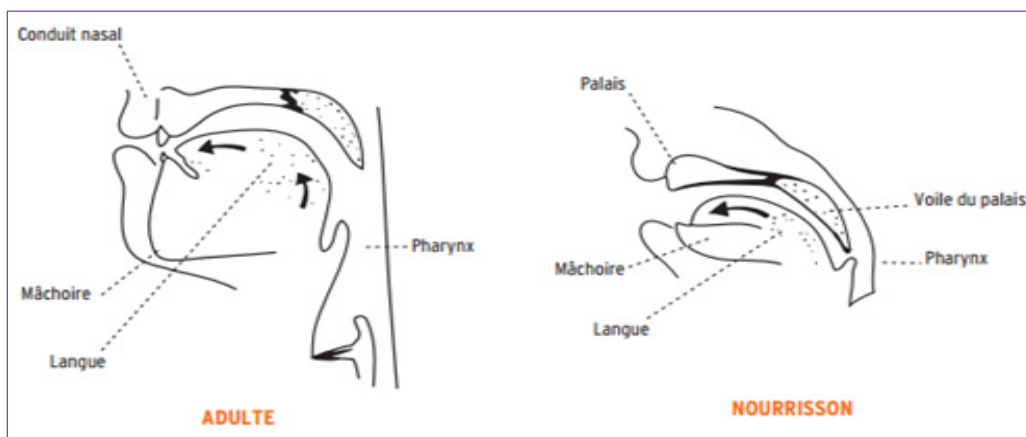


Figure 1-4 Le conduit vocal d'un nourrisson comparé à celui d'un adulte
Tirée de MARC (2009, p.21)

1.3 La complexité et la variété du cri d'un nouveau-né

La complexité et la variabilité du signal de cri d'un nouveau-né ont rendu l'analyse de ce dernier difficile. D'une manière générale, le cri d'un nouveau-né est considéré comme le seul langage de communication avec le monde extérieur et reste encore incompréhensible. En effet, la plupart du temps, les parents ne peuvent pas savoir pour quelle raison leur bébé pleure.

Les propriétés intrinsèques du signal de cri varient d'un nourrisson à un autre en fonction de son état de santé, de son état émotif et de son attitude (Manfredi, Tocchioni et Bocchi, 2006).

L'analyse acoustique des signaux des nouveau-nés est de grande importance (Manfredi, Tocchioni et Bocchi, 2006). Elle pourrait servir d'aide au diagnostic clinique. Cependant, les résultats de mesure de la fréquence fondamentale ne sont pas toujours fiables. En fait, le cri du nouveau-né se caractérise par une très grande plage de fréquence fondamentale F_0 , des changements brusques et des segments voisés / non voisés de très courtes durées (Manfredi, Tocchioni et Bocchi, 2006).

De plus, les fréquences de résonance du conduit vocal sont très différentes dans la plupart des cas. Elles ont besoin d'un suivi précis. En fait, l'analyse de leur évolution dans le temps et au cours des premiers mois de la vie du nouveau-né peut donner des informations utiles sur l'évolution de ses capacités phonatoires et les éventuels dysfonctionnements du système nerveux central (Manfredi, Tocchioni et Bocchi, 2006).

1.3.1 Description d'un signal de cri chez les nouveau-nés

Un signal de cri est composé d'un mélange de sons tels que la toux, le hoquet, etc. Pour identifier la partie importante du signal, la procédure de segmentation est appliquée. La segmentation permet de déterminer les segments de cris trouvés dans l'ensemble d'un enregistrement donné. Cette étape est souvent appelée détection d'activité vocale (VAD).

Dans l'étude réalisée par (Várallyay Jr, Illényi et Benyó, 2008), les auteurs ont introduit deux nouvelles méthodes avec modification de la fonction de la génération du spectre harmonique (HPS). Cette méthode peut être utilisée pour la segmentation des cris. Le résultat souhaité est le déploiement d'un algorithme capable de détecter automatiquement les segments de cris.

Les cris sont produits lors de l'expiration par la vibration des cordes vocales et peuvent affecter le son harmonique. De plus, il pourrait y avoir des sons et/ou événements particuliers lors de l'émission d'un son tels que la toux, les silences, etc. Ceci pourrait toutefois affecter le processus de détection et d'analyse des sons réels des cris.



Figure 1-5 Représentation de 7 secondes d'enregistrement d'un cri de bébé

Le signal de la Figure 1.5 montre un cri chez un nouveau-né composé par des segments de cris (a) et des segments (b) (c) non utilisés, comme le silence, le bip d'une machine et les paroles des infirmières.

Pour extraire et identifier uniquement les segments de cris (a), il faut éliminer les autres segments, soit les segments (b) et (c).

1.3.1.1 Les catégories de cris

Le terme «cri» est traité le plus souvent en référence à la performance sonore. Les sons de cri, en particulier chez les nourrissons, sont utilisés par les médecins comme des critères supplémentaires pour l'évaluation de la fonction neurologique en général et spécifiquement de la fonction respiratoire. Le son de cri est évidemment très accessible à l'observation (Messaoud et Tadj, 2011).

En tant qu'outil de clinique, certaines manifestations des cris ont souvent été utilisées pour tenter d'identifier diverses conditions pathologiques chez les nourrissons (Messaoud et Tadj, 2011).

Il existe trois modes principaux du cri qui sont : phonique, dysphonique et hyperphonique. Le moins complexe et le plus élémentaire d'entre eux est le cri phonique. Comme le montre la Figure 1.6.

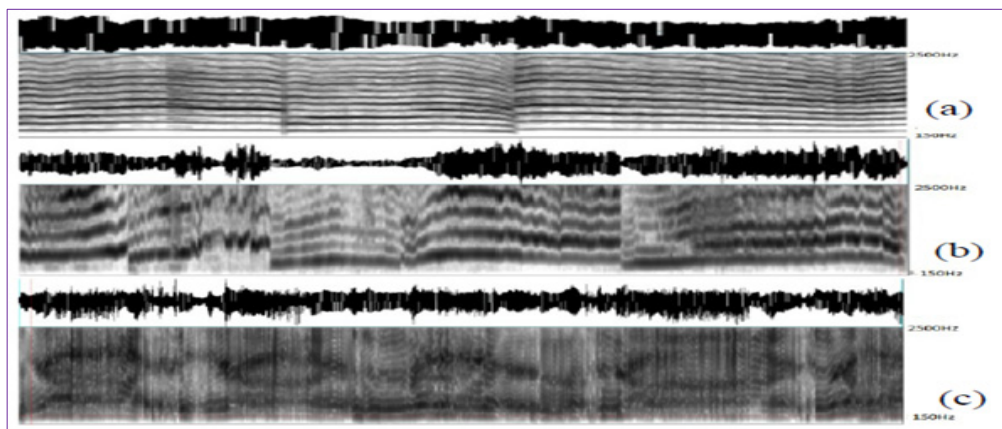


Figure 1-6 Spectrogramme des différents modes des cris (a) phoniques
(b) hyperphonique (c) dysphonique
Tirée de Kheddache et Tadj (2013a, p.798)

La catégorie (a) représente un cri phonique à une fréquence fondamentale entre 350-750Hz. L'estimation de la fréquence fondamentale de cette catégorie de cri est semblable à celle de la parole des adultes.

La « dysphonation » est un type de cri de catégorie (c) facile à distinguer à partir du schéma d'histogramme en se basant sur les turbulences qui y apparaissent (Truby et Lind, 1965). C'est le résultat d'un effort extrême. L'intensité générale associée entraîne une surcharge au niveau du larynx le long du conduit vocal. Il est toujours typique de la réponse la plus bruyante à l'inconfort.

En raison de la nature acoustique et de l'apparence de l'aberration proprement dite, les auteurs désignent habituellement la «dysphonation» comme une turbulence (Truby et Lind, 1965).

La dernière catégorie (b) représente un cri hyperphonique de fréquence fondamentale qui varie entre 1000-2000 Hz (Yasmina Kheddache, 2013). Dans ce type de cri, l'estimation de la fréquence fondamentale n'est pas semblable à celle de la parole des adultes à cause de l'élargissement de la structure harmonique et de la séparation de la contribution de l'excitation spectrale (Abdullah-Al-Mamun, Sarker et Muhammad, 2009).

1.4 Les problèmes de l'estimation de la fréquence fondamentale

L'estimation fiable et précise de la fréquence fondamentale pour un signal de cri ou un signal de la parole s'est souvent révélée une tâche difficile.

Premièrement, la forme d'onde de l'excitation glottale d'un signal de cri n'est pas un enchaînement des impulsions parfaitement périodique. Elle produit un signal non stationnaire dont la détection de la période et de la fréquence fondamentale sont complexes (Hui et al., 2006).

Le deuxième problème, le plus difficile, est la difficulté de séparer les segments voisés des segments non voisés. Dans plusieurs cas, des segments voisés avec une énergie réduite ont été considérés comme des segments non voisés et, des segments non voisés avec une grande énergie ajoutée par le bruit comme des segments voisés (Rabiner et al., 1976) .

Le troisième problème se manifeste lors de l'interaction entre le conduit vocal responsable de la production des formants et l'excitation glottale qui est à son tour responsable de la production de la fréquence fondamentale. En effet, les formants peuvent modifier la structure de la forme d'onde glottale de sorte que cette influence rend l'extraction de la fréquence fondamentale difficile (Hui et al., 2006).

L'utilisation de l'algorithme de prédiction linéaire résiduel LPC «*Linear Prediction Residual*» est une solution qui permet d'éliminer l'information du conduit vocal et le bruit à haute-fréquence. Ceci améliore la précision de détection de la fréquence fondamentale dans une certaine mesure. Mais, le modèle du conduit vocal est semblable à un filtre passe-bas qui coupe les fréquences harmoniques les plus élevées. Par conséquent, la fréquence fondamentale souhaitée sera atténuée et difficile à détecter. Une solution a été proposée par (Hui et al., 2006) pour corriger et améliorer ce type de problème d'estimation de la fréquence atténuée sous forme d'une combinaison entre l'algorithme LPC «*based Cepstrum*» et l'algorithme HPS.

Le quatrième problème est la difficulté à définir exactement le début et la fin d'une période durant un segment voisé du signal de cri. Cependant, il est souvent pratiqué un choix arbitraire pour fixer le début et la fin d'une période (Rabiner et al., 1976). À titre d'exemple, on peut citer le cas d'une onde acoustique, dont le choix du début et de la fin d'une période est basé sur les valeurs maximales successives ou sur le passage par zéro avant chaque maximum comme indiqué sur la Figure 1.7. La seule condition à une telle mesure est qu'elle soit conforme, période par période, afin pouvoir définir l'emplacement exact du début et de la fin de chaque période du pic.

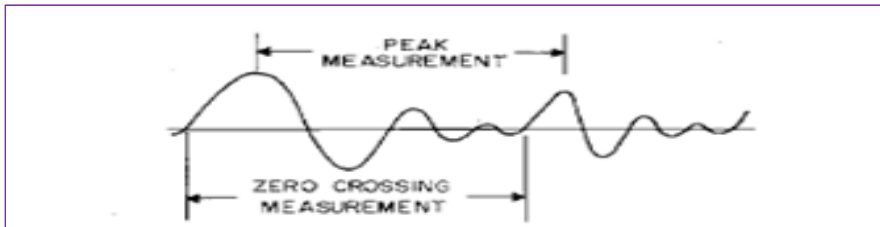


Figure 1-7 Deux différentes mesures utilisées pour définir les marqueurs du pic
Tirée de Rabiner et al (1976, p.400)

Les autres difficultés conduisant à des erreurs dans l'estimation de la fréquence fondamentale pour le cas des segments hyperphoniques, sont la séparation de la contribution de l'excitation spectrale du conduit vocal et l'identification des sous-harmoniques de la fréquence fondamentale.

Ce problème n'existe pas dans le cas des segments phoniques, car l'estimation de la fréquence fondamentale d'un cri de bébé est presque semblable à celle de la parole d'un adulte (Lederman, 2010).

Un autre problème identifié dans l'étude citée ci-dessus correspond à l'estimation incorrecte des sous-harmoniques de la fréquence fondamentale. En effet, le doublement de la fréquence fondamentale produit une mauvaise classification du segment. Un segment phonique sera par erreur classifié comme un segment hyperphonique et vice-versa, un segment hyperphonique sera classé comme un segment phonique par l'erreur de la sous-estimation de la fréquence fondamentale. La solution proposée consiste à lisser le contour de la fréquence fondamentale obtenue (Lederman, 2010).

La détection du cri est une procédure qui sélectionne le son du cri dans un enregistrement donné. Cependant, le problème le plus difficile dans cette sélection est de reconnaître le son inspiratoire et de le séparer du son voisé (Várallyay Jr, Illényi et Benyó, 2009).

Un système appelé Automatic Infant Cry Detection «AICD» a été créé par (Várallyay Jr, Illényi et Benyó, 2009) pour résoudre le problème de la séparation des sons inspiratoires et des sons voisés.

Ce système analyse d'abord le contenu d'énergie de chaque segment enregistré en appliquant la fonction STE «*Short-Time Energy Function*» qui fournit une représentation de l'amplitude de l'énergie pour chaque segment enregistré. Les plus grandes amplitudes détectées représentent donc les segments voisés et les plus petites amplitudes les segments inspiratoires.

Les chercheurs G. Várallyay Jr, A. Illényi et Z. Benyó ont constaté que dans la plupart des segments inspiratoires, une quantité de bruit est ajoutée par l'environnement d'enregistrement. Ainsi, l'amplitude d'énergie des segments augmente par l'ajout de l'énergie du bruit (Várallyay Jr, Illényi et Benyó, 2009). Ceci produit une séparation erronée des segments voisés et inspiratoires. La fonction HPS est proposée par les auteurs pour résoudre ce type de problème.

Cette fonction consiste à faire une analyse spectrale du contenu des segments enregistrés dans le but de détecter les bruits et de les supprimer. En effet, l'algorithme HPS détecte la fréquence fondamentale et ses harmoniques directement à partir du contenu spectral du signal analysé pour chaque segment enregistré. Un segment est identifié comme son voisé si le contenu spectral de la fréquence fondamentale et ses harmoniques présentent une variation régulière. Dans le cas contraire, le segment est sélectionné comme son inspiratoire.

1.5 Les algorithmes de détection de la fréquence fondamentale

L'extraction de la fréquence fondamentale dans un signal de cri a été abondamment abordée durant les trois dernières décennies. Ainsi, un grand nombre d'algorithmes d'extraction de la fréquence fondamentale a été développé.

Dans la littérature, les algorithmes d'extraction de la fréquence fondamentale sont classés par (Ykhlef et al., 2007) en trois catégories essentielles : temporels, spectraux et hybrides.

1.5.1 Les algorithmes temporels d'extraction de la fréquence fondamentale

1.5.1.1 L'algorithme temporel AMDF

L'un des algorithmes temporels les plus utilisés dans le traitement du signal de cri est la fonction AMDF «*Average Magnitude Difference Function*» (Yu-Min et al., 2003). Celle-ci est d'ailleurs à l'origine de l'amélioration de plusieurs algorithmes.

La fonction AMDF correspond à une analyse d'autocorrélation. Un signal de différence est formé entre le signal de la voix, le signal du cri ou le signal de musique retardé et celui d'origine. A chaque valeur de retard la fonction de magnitude absolue est calculée.

Pour un bloc de N échantillons, la fonction AMDF est définie par la relation suivante (Yu-Min et al., 2003):

$$R(m) = \frac{1}{N} \sum_{n=0}^{N-1-m} (|x(n) - x(n+m)|), 0 \leq m \leq M_0 \quad (1.1)$$

$x(n)$ est le bloc du signal original, $x(n-m)$ est le signal décalé de la valeur m dans le temps et M_0 est le nombre de points de la fonction AMDF. $R(m)$ est la somme de la différence entre les deux signaux. Le calcul du minimum de la fonction nous renseigne sur la valeur m qui est égale ou très proche de la période fondamentale.

L'algorithme AMDF a été souvent utilisé pour la détection de la fréquence fondamentale grâce à sa simplicité de calcul. Comme tous les algorithmes de détection de la fréquence fondamentale, cette méthode produit deux erreurs dans l'estimation de la fréquence fondamentale qui sont souvent en relation avec la complexité des formes d'ondes du signal de cri ou parole.

La première erreur est l'estimation de la fréquence fondamentale avec une valeur double de la valeur réelle. La deuxième erreur se manifeste par une valeur réelle qui est multiple de la valeur estimée. Les raisons principales de ces erreurs sont la tendance à la baisse des pics d'énergie élevés de la fonction AMDF (Abdullah-Al-Mamun, Sarker et Muhammad, 2009). L'inconvénient de cette tendance est d'augmenter les risques d'erreur d'octave.

Dans l'échelle d'octave, un son S1 de fréquence f_1 est une octave au-dessus d'un son S2 de fréquence f_2 , si la fréquence de S1 est deux fois la fréquence de S2. Dans le domaine fréquentiel, les erreurs d'octaves sont essentiellement présentes en raison de la teneur faible en énergie dans les ordres inférieurs des harmoniques.

Une erreur d'octave désigne donc la détection non correcte d'une fréquence : par exemple la détection d'une fréquence de f_1 au lieu d'une fréquence de $f_2=2*f_1$.

Parmi les méthodes qui corrigent la tendance de l'AMDF existe notamment celle de la combinaison par multiplication de la fonction ACF avec le miroir de la fonction AMDF (Abdullah-Al-Mamun, Sarker et Muhammad, 2009). Au début, le problème de la tendance produit par l'AMDF est corrigé par la modification de l'AMDF original en ajoutant un seuil de deux tiers pour le dernier échantillon de chaque trame. L'effet de ce seuil est d'aligner au maximum les pics de la fonction AMDF. De plus, la multiplication entre les fonctions ACF et AMDF en miroir est appliquée pour diminuer l'effet du bruit et éliminer les fréquences non souhaitées. En effet, la multiplication entre les points des deux fonctions ACF et AMDF aboutit en général à des valeurs plus grandes facilitant l'extraction précise de la fréquence fondamentale (Abdullah-Al-Mamun, Sarker et Muhammad, 2009).

1.5.1.2 L'algorithme temporel ACF

La fonction ACF «*Auto-Correlation Function*» est une autre méthode utilisée dans le domaine temporel pour la détection de la fréquence fondamentale des signaux de la parole, de la musique et des cris chez les nouveau-nés. Elle se base sur la détection de la valeur maximale de la fonction d'autocorrélation dans un intervalle de temps donné .

Pour un signal discret, la fonction d'autocorrélation est généralement définie comme suit (Upadhyaya et Wankhede, 2012):

$$F(m) = \lim_{N \rightarrow \infty} \left(\frac{1}{(2N+1)} \sum_{n=-N}^N (x(n) \times x(n+m)) \right), \quad 0 \leq m \leq M_0 \quad (1.2)$$

N est la longueur de la fenêtre analysée et M_0 le nombre de points de la fonction d'autocorrélation. Si on suppose que $x(n)$ est la séquence périodique de période p , $x(n) = x(n+p)$ pour tout n . Il est démontré que la fonction d'autocorrélation est aussi périodique à la même période $F(m) = F(m+P)$.

Dans le cas d'un signal non stationnaire, comme la parole et le cri, le calcul par l'équation (1.2) n'est pas fiable à longue durée (Upadhyaya et Wankhede, 2012).

Dans ce travail, les segments utilisés sont nécessairement courts avec un nombre d'échantillons fini. Donc, l'utilisation de la fonction d'autocorrélation de courte durée est primordiale et définie comme suit :

$$F(m) = \frac{1}{N} * \sum_{n=0}^{N-1-m} (x(n) \times x(n+m)), \quad 0 \leq m \leq M_0 \quad (1.3)$$

m est le décalage ou le retard. La fréquence fondamentale est définie par la valeur de m au point maximum de la fonction $F(m)$.

L'idée derrière le déploiement de cette fonction est de déterminer à quel point deux échantillons successifs d'un signal se ressemblent. L'algorithme ACF est utilisé fréquemment pour l'estimation de la fréquence fondamentale. Les résultats obtenus sont satisfaisants, pour un signal qui ne varie pas trop d'une période à une autre.

L'algorithme ACF présente l'avantage de permettre une estimation de la fréquence fondamentale avec un calcul de multiplication simple. Donc, le temps de calcul dans les processeurs sera réduit.

1.5.1.3 L'algorithme temporel SIFT modifié

La méthode SIFT modifié «*modified Simple Inverse Filtering Tracking*» a été publiée pour la première fois par Dror Lederman en 2010. Cette méthode d'analyse temporelle a été souvent utilisée dans l'extraction de la fréquence fondamentale avec beaucoup de précision. Le processus de fonctionnement de cet algorithme est présenté selon le schéma bloc suivant :

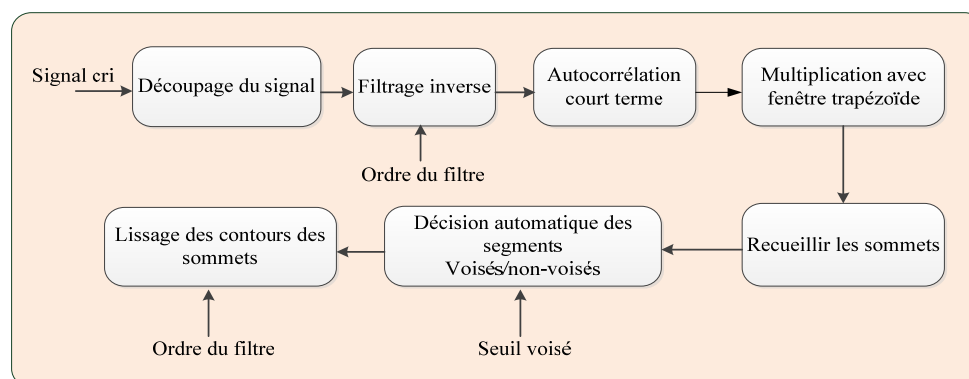


Figure 1-8 Schéma bloc de l'algorithme SIFT modifié
Tirée de Lederman (2002, p.3)

D'après la figure ci-dessus, le processus se déroule en sept étapes : l'analyse du signal de cri se fait sur des segments courts liés à sa nature non stationnaire. Ainsi, le signal traité est segmenté sous forme de trames entrelacées. Par la suite, chaque trame est multipliée par la fenêtre de Hamming appropriée.

Dans une seconde étape, l'application d'un filtrage inverse pour atténuer l'influence du conduit vocal est effectuée. Cette opération est suivie par une estimation de la fonction d'autocorrélation, que représente dans la figure ci-dessus le troisième bloc.

Une multiplication par une fenêtre trapézoïdale est alors effectuée en quatrième étape.

Ensuite, la somme des produits précédents est comparée avec le seuil voisé de 0.4 afin de trouver la fréquence fondamentale (Lederman, 2010).

L'apport de cet algorithme est l'élimination du problème de la sous-estimation et de la surestimation de la fréquence fondamentale, grâce à une étape supplémentaire réalisée après l'estimation de la fréquence fondamentale. Elle permet de lisser le contour de la fréquence fondamentale afin de diminuer ses changements brusques (Lederman, 2010).

1.5.2 Les algorithmes spectraux d'extraction de la fréquence fondamentale

Ce type d'algorithme permet la mesure de la fréquence fondamentale dans le domaine fréquentiel en temps réel. En effet, la fréquence fondamentale est estimée à partir des informations contenues dans les harmoniques du signal. L'algorithme HPS ou la fonction de produit harmonique spectrale est un exemple d'algorithme spectral.

1.5.2.1 L'algorithme HPS

La fonction HPS a été publiée pour la première fois par R.Noll en 1970. Cette méthode mesure le maximum des coïncidences des produits des harmoniques de chaque spectre d'un segment de signal suivant les équations (1.4, 1.5) (Hui et al., 2006):

$$P(e^{jrw}) = \prod_{r=1}^R |X(e^{jrw})| \quad (1.4)$$

$$\hat{P} = \max_{wi} \{P(e^{jw})\} \quad (1.5)$$

R représente le nombre d'harmoniques choisi dans le segment pour l'analyse spectrale. Le meilleur choix est la valeur de cinq harmoniques (le nombre de cinq est optimal, car le contenu en énergie des harmoniques d'ordre supérieur est moins important).

Dans chaque trame spectrale $X(e^{jrw})$, w_i représente la valeur de la fréquence choisie d'un intervalle qui contient les fréquences fondamentales possibles. $P(e^{jrw})$ représente le résultat de la matrice de corrélation périodique et \hat{P} est la valeur maximale du produit $P(e^{jrw})$ (De La Cuadra, Master et Sapp, 2001).

L'algorithme HPS se base sur les cinq étapes essentielles suivantes.

La première étape est la segmentation du signal original en trame à court terme qui varie entre 20ms et 30ms.

La deuxième étape applique la transformée de Fourier discrète (TFD) qui permet de décrire un signal discret dans le domaine fréquentiel. La TFD d'un signal $x(n)$, composé d'une suite de N termes ($x(0), x(1), \dots, x(N-1)$) est la suite de N termes ($X(0), X(1), \dots, X(N-1)$). Le calcul de ces coefficients est défini par la relation suivante (Gargour et Ramachandran, 2006) :

$$TFD\{x(n)\} = X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi \frac{nk}{N}} \quad (1.6)$$

N représente le nombre d'échantillons d'un signal $x(n)$, la variable k varie de 0 à $N-1$ et $X(k)$ est l'approximation de la TFD de ce signal aux N échantillons (Gargour et Ramachandran, 2006).

L'étape suivante applique les décimations fréquentielles à des valeurs $1/1, 1/2, 1/3, 1/4$ et $1/5$. La décimation ou « Downsampling » d'un signal discret $x(n)$ par un entier D consiste à conserver uniquement un échantillon sur D de ce signal. La relation entre le signal original $x(n)$ et sa version décimée $x_{\downarrow D}(n)$ par une valeur D est notée selon l'équation (1.7).

Un exemple de la décimation dans le cas où $D=3$ est présenté dans la Figure 1.9 (Gargour et Ramachandran, 2006).

$$x_{\downarrow D}(n) = x(Dn) \quad (1.7)$$

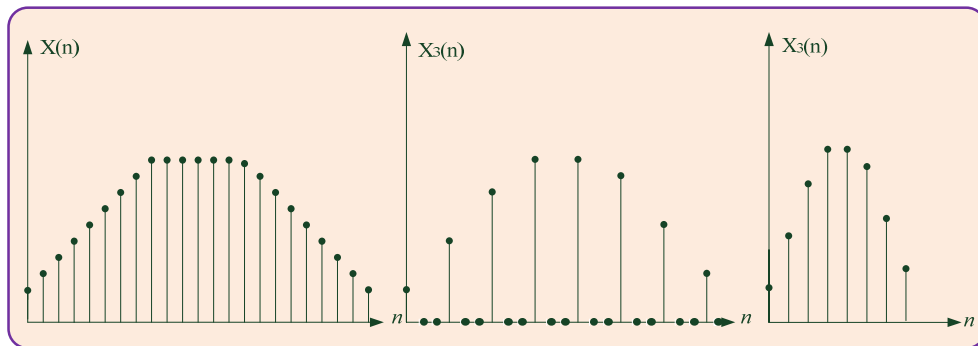


Figure 1-9 Décimation d'un signal échantillonné par $D=3$
Tirée de Gargour et Ramachandran (2006, p.347)

Un exemple d'un spectre TFD d'un signal donné, représenté par sa fréquence fondamentale F_0 et ses harmoniques $2F_0$, $3F_0$, $4F_0$ et $5F_0$ est présenté dans la Figure 1.10. Les valeurs de décimation D utilisées sont 2, 3, 4, et 5.

Le résultat maximal du produit des coefficients TFD décimés de TFD1/1, TFD1/2, TFD1/3, TFD1/4 et TFD1/5 donne l'information nécessaire sur la valeur exacte de la fréquence fondamentale F_0 .

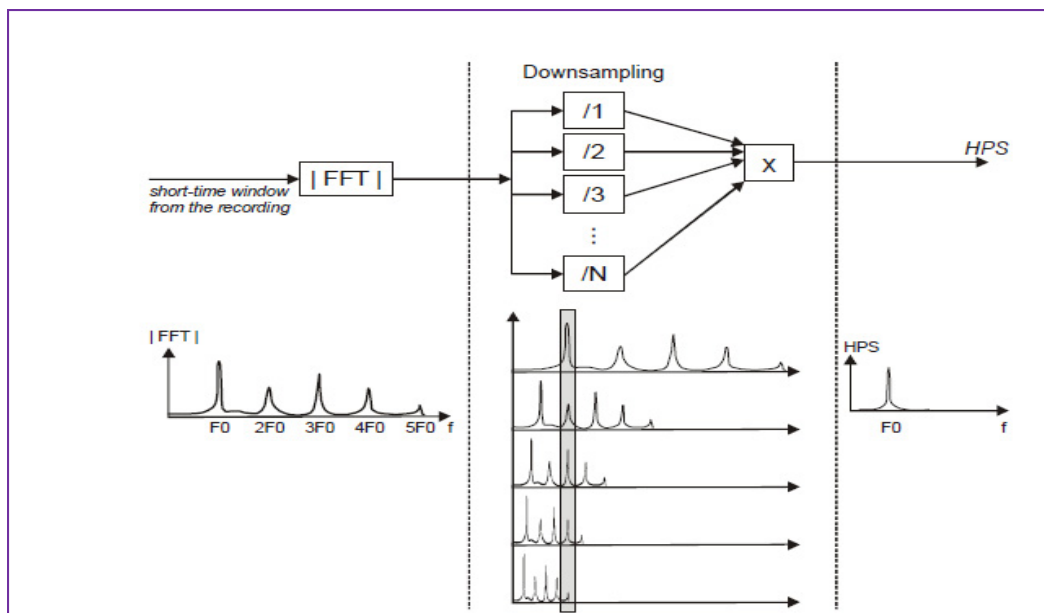


Figure 1-10 La décimation utilisée par HPS avec les valeurs D (1, 2, 3, 4 et 5)
Tirée de De La Cuadra, Master et Sapp (2001, p.2)

L'avantage de l'algorithme HPS se manifeste dans sa capacité à éliminer le bruit et les segments non voisés au cours de l'estimation de la fréquence fondamentale pour chaque segment. Mais l'inconvénient primordial de l'HPS est la nécessité d'améliorer la détection des basses fréquences avec la méthode de remplissage par des zéros (zéro padding).

L'introduction des valeurs nulles sur un signal à basse fréquence peut augmenter la résolution ΔF du spectre de ce signal avant sa transformation de sorte que le spectre peut être interpolé au demi-ton le plus proche.

De cette façon, une perte se produit dans beaucoup de calculs non nécessaires, puisque l'interpolation peut affecter aussi les fréquences les plus élevées.

L'erreur d'octave est un problème courant dans l'estimation de la fréquence fondamentale avec l'algorithme HPS. Dans la majorité des cas de ces erreurs, le pic de F_0 est détecté à une valeur d'octave plus haute. Un traitement postérieur, sous forme de règle dans le choix du pic F_0 avec l'erreur d'octave le plus faible possible, peut corriger ce genre de problème. De la même manière, le problème courant d'erreur du déroulement de pic F_0 est corrigé avec un traitement postérieur (De La Cuadra, Master et Sapp, 2001).

1.5.3 Les algorithmes hybrides d'extraction de la fréquence fondamentale

La troisième catégorie correspond aux méthodes hybrides d'extraction de la fréquence fondamentale. Celles-ci se basent sur la combinaison de deux ou plusieurs algorithmes qui existent déjà pour augmenter la précision d'extraction des données fréquentielles.

La combinaison de deux algorithmes différents peut améliorer ou éliminer une faiblesse dans l'une des deux approches. Par exemple, la combinaison de l'algorithme ACF avec AMDF peut améliorer la précision des calculs des résultats de la fréquence fondamentale (Abdullah-Al-Mamun, Sarker et Muhammad, 2009).

L'application de cet algorithme n'a pas encore été effectué sur le signal de cri chez les nouveau-nés. Mais il a donné de bons résultats d'estimation de la fréquence fondamentale avec le signal de la parole (Abdullah-Al-Mamun, Sarker et Muhammad, 2009).

1.6 Les études comparatives des algorithmes d'estimation de la fréquence fondamentale

Cette partie présente la description de différentes études comparatives de performance d'estimation de la fréquence fondamentale. Les paramètres de comparaison seront introduits et définis pour l'utilisation ultérieure dans l'analyse des résultats obtenus.

1.6.1 Les paramètres utilisés pour l'évaluation de la performance des algorithmes

Les chercheurs ont appliqué dans leurs travaux les paramètres d'erreurs cités ci-dessous (Rabiner et al., 1976) pour l'évaluation des performances des algorithmes d'estimation de la fréquence fondamentale.

Soient $F_{0s}(n)$ et $F_{0e}(n)$ les valeurs de la fréquence fondamentale mesurée par l'analyse standard et estimée par les algorithmes, respectivement.

- l'erreur voisée/non voisée : Ce paramètre consiste à calculer l'erreur produite par la classification des segments. En effet, cette erreur est définie par l'équation suivante :

$$Erreur(V/NV) = \frac{\text{nombre de segments}(F_{0s}(n) \neq 0 \text{ et } F_{0e}(n) = 0)}{\text{nombre de segments}(F_{0s}(n) \neq 0)} \quad (1.8)$$

Où $F_{0s}(n) \neq 0$ et $F_{0e}(n) = 0$ représentent respectivement le $n^{\text{ième}}$ segment qui est classifié comme segment voisé par l'analyse standard (test effectué par un algorithme de référence choisi). Mais il est classifié comme segment non voisé par l'un des évaluateurs de test déployé (test effectué par un algorithme non identifié comme la référence principale de test).

- l'erreur non voisée/voisée : Dans ce cas, l'erreur est produite par la classification des segments non voisés effectuée par l'analyse standard. Les mêmes segments sont classés voisés par l'estimateur du test. Cette erreur est définie par l'équation suivante :

$$Erreur(NV/V) = \frac{\text{nombre de segments}(F_{0s}(n)=0 \text{ et } F_{0e}(n) \neq 0)}{\text{nombre de segments}(F_{0s}(n)=0)} \quad (1.9)$$

$F_{0s}(n) = 0$ et $F_{0e}(n) \neq 0$ représentent le nombre de fois où le même segment est détecté non voisé par l'analyse standard et voisé par l'estimateur du test, respectivement.

$F_{0s}(n) = 0$ représente le nombre de segments non voisé détecté par l'analyse standard.

- dans le cas où la fréquence fondamentale estimée par l'analyse standard et l'un des algorithmes sont différents du zéro $F_{0s}(n) \neq 0$ et $F_{0e}(n) \neq 0$, un autre type d'erreur sera produit en fonction des valeurs de $F_{0s}(n)$ et $F_{0e}(n)$.

Supposons $F_{0s}(n) = F_1 \neq 0$ et $F_{0e}(n) = F_2 \neq 0$, le calcul d'erreur dans ce cas est le suivant :

$$E(n) = |F_1 - F_2| \quad (1.10)$$

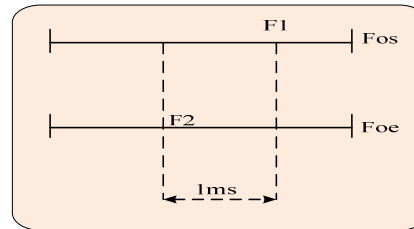


Figure 1-11 Exemple des valeurs de $F_{0s}(n)$ et $F_{0e}(n)$ du même segment

La Figure 1.11 montre deux valeurs différentes de F_0 pour le même segment analysé par deux méthodes différentes. La valeur de 1ms représente le seuil en temps qui délimite et classe les types d'erreurs affectant le segment voisé.

- si $E(n) \geq X$ échantillons (plus de 1ms d'erreur d'estimation de la période T_0 pour une fréquence d'échantillonnage Y) : cette erreur d'analyse est connue sous le nom d'erreur grossière. Elle se manifeste par l'estimation erronée double ou triple de la fréquence fondamentale. Par conséquent, nous concéderons que l'algorithme du test a échoué dans la détection de F_0 ;
- si $E(n) < X$ échantillons (moins de 1ms d'erreur d'estimation de la période T_0 pour une fréquence d'échantillonnage Y) : dans ce cas, l'erreur d'analyse est connue sous le nom d'erreur fine et par conséquent, la valeur de F_0 est estimée d'une manière plus précise et plus proche de la valeur exacte;
- si $E(n)=0$: c'est le cas idéal où il n'y a pas d'erreur d'analyse commise. En effet, c'est un cas particulier de l'erreur fine, connu comme le Paramètre d'Erreur Sans Erreur Commise (PESEC).

1.6.2 Le cas d'un signal de parole

Dans le domaine du traitement de la parole, plusieurs études comparatives des algorithmes de détection de la fréquence fondamentale ont été réalisées. En effet, le but de cette comparaison est de sélectionner l'algorithme le plus performant pour l'estimation de la fréquence fondamentale et de réduire le taux d'erreur de la mesure.

Une étude comparative a été réalisée par (Ykhlef et al., 2007) dans le but de sélectionner l'algorithme le plus performant pour estimer la fréquence fondamentale. Ce travail consiste à traiter quatre algorithmes, dont deux sont temporels MACC « *modified autocorrelation center-clippin* » et C-AMDF « *Clipping -Average Magnitude Diffrence Function* » et deux autres fréquentiels HPS et CEP « *cepstrum* ».

- le MACC est une méthode qui se base sur le même principe que l'algorithme d'autocorrélation. La seule différence est le centre de coupure qui correspond à la transformation non linéaire du signal. En général, le centre de coupure représente un seuil de valeur d'environ 50 % de l'amplitude de l'échantillon maximal dans une trame du signal analysé. Cette quantité est intégrée et ajoutée par l'algorithme ACF à la valeur du signal dans chaque trame au niveau de l'étape de prétraitement;
- le C-AMDF est une version modifiée de l'AMDF basé sur le clippage. Ce dernier consiste à appliquer une transformation non linéaire sur les segments à court terme du signal traité (Ykhlef et al., 2007);
- le CEPSTRE est une technique très reconnue pour la détection de la fréquence fondamentale en séparant le signal source de la réponse impulsionnelle du conduit vocal. En effet, en passant par plusieurs blocs d'analyse du signal, le cepstre est obtenu par le calcul de la transformée de Fourier inverse de ce signal (Manfredi et al., 2000) .

L'évaluation de ces algorithmes est basée sur l'analyse des erreurs d'estimation de la fréquence fondamentale qui sont : l'erreur grossière, l'erreur fine et le paramètre d'extraction sans erreur commise.

La base de données utilisée dans cette étude (Ykhlef et al., 2007) est composée par des enregistrements de voix féminines, masculines et d'enfants. Le résultat de l'évaluation finale de cette analyse a permis de classer la méthode du cepstre parmi les techniques les plus performantes, générant ainsi les meilleurs résultats d'estimation de la fréquence fondamentale.

Les algorithmes HPS et C-AMDF donnent de bons résultats pour détecter les régions de transition voisées et non voisées par rapport à l'algorithme MACC. Ainsi, l'estimation de l'erreur fine a été exacte avec l'algorithme HPS, bien qu'il soit parfois à l'origine d'erreurs grossières majeures (Ykhlef et al., 2007).

Une autre étude comparative entre sept algorithmes d'estimation de la fréquence fondamentale a été abordée dans l'article (Rabiner et al., 1976). La base de données utilisée pour réaliser cette étude est composée de trois hommes, trois femmes et un enfant. Ainsi, une estimation de la fréquence fondamentale a été réalisée par un programme interactif sophistiqué à haute précision dans le but de comparer cette mesure avec les résultats obtenus par la simulation des algorithmes étudiés.

Les algorithmes testés dans cette étude (Rabiner et al., 1976) sont: MACC, CEP, SIFT et l'AMDF présentés précédemment auxquels s'ajoutent les algorithmes PPROC «*Parallel processing method*», DARD «*Data reduction method*», LPC «*Linear predictive coding*».

Voici une brève description de ces derniers algorithmes:

LPC : cette méthode se base tout d'abord sur le classement des segments vocaux en segments voisés/ non voisés. Dans un second temps les segments voisés seront filtrés par un filtre passe-bas. Ensuite, une décimation d'ordre cinq est appliquée sur ce signal suivie par une TFD appliquée pour une représentation bien précise des harmoniques spectraux du signal. Finalement, une transformation de Newton est réalisée (Rabiner et al., 1976).

PPROC : cette méthode propose de filtrer le signal par un filtre passe-bas. Ensuite, une série de mesures est effectuée sur les sommes et les valeurs du signal filtré pour donner naissance à six fonctions séparées. Chaque fonction est traitée par un estimateur de période élémentaire. Ces six fonctions sont ensuite combinées par un algorithme de décision sophistiqué qui détermine la fréquence fondamentale (Rabiner et al., 1976) .

DARD : elle est basée sur le principe de détection du pic de la réduction des données de Miller. Ce détecteur de pic met en place des marqueurs de pics directement sur le signal filtré par un filtre passe-bas de 0 à 900 Hz. Il s'appelle donc un détecteur de pic avec pas synchronisé.

Dans le but d'obtenir des marqueurs de pics appropriés, la méthode de réduction de données détecte, premièrement, les cycles d'excursion dans la base de la forme d'onde sur des intervalles entre les principaux passages à zéro.

Ensuite, l'algorithme essaie d'isoler et de déterminer les cycles d'excursions principales qui correspondent à de véritables périodes de pics. Ceci est accompli à travers une série d'étapes utilisant la mesure d'énergie, la logique fondée sur les périodes des pics autorisées et la variation anticipée du taux syllabique des pics. Enfin, un correcteur d'erreurs est intégré dans l'algorithme pour fournir une mesure raisonnable dans la continuité des marqueurs de pics (Rabiner et al., 1976).

Différents paramètres d'évaluation de performance sont appliqués dans ce test : l'erreur grossière, l'écart-type de l'erreur fine, la valeur moyenne d'erreur fine, l'erreur de détection des segments voisés / non voisés et l'erreur de détection des segments non voisés / voisés.

On peut résumer les résultats d'évaluation des sept méthodes comme suit : l'algorithme CEP a une forte tendance à classer les segments voisés comme non voisés, mais avec un taux d'erreur très faible dans le classement de segments non voisés comme voisés.

Les quatre algorithmes MACC, PPROC, LPC et AMDF ont donné de bons résultats concernant l'erreur de confusion des segments voisés et non voisés. L'algorithme DARD a présenté des résultats moins satisfaisants, tandis que le CEP a donné la plus faible performance.

Dans le cas de l'évaluation de l'erreur de classement des segments non voisés et voisés, les algorithmes MACC, SIFT, DARD, PPROC et AMDF avaient tous une performance équivalente. La méthode utilisant LPC a eu la plus basse performance.

Les algorithmes MACC, CEP et LPC ont une performance relativement élevée dans le cas d'évaluation de l'écart-type d'erreur fine. En effet, ils ont donné des résultats presque similaires.

Les deux détecteurs DARD et PPROC ont obtenu une performance moins bonne. Enfin, la mauvaise performance dans cette catégorie est réalisée par l'algorithme AMDF (Rabiner et al., 1976).

La valeur moyenne de l'erreur fine a été nulle pour les sept méthodes utilisées. Par contre, il n'est pas possible de faire une évaluation comparative basée uniquement sur ce paramètre de la valeur moyenne.

1.6.3 Le cas d'un signal de cri chez les nouveau-nés

Contrairement au signal de la parole, nous avons remarqué un manque d'intérêt et une absence d'études comparatives pour déterminer la performance des algorithmes d'évaluation de la fréquence fondamentale d'un signal de cri chez les nouveau-nés.

Cette lacune nous a encouragé à entamer une étude sur un terrain vierge et à participer par la suite à l'évolution continue de ce domaine. Ce travail consiste à programmer en MATLAB et à faire l'analyse comparative de performance des quatre algorithmes AMDF, ACF, HPS et SIFT modifié.

1.7 Conclusion

Dans le présent chapitre, nous avons présenté les différentes caractéristiques d'un signal de parole et de cri. La caractéristique commune de ces signaux est le processus de production des sons.

Cette étude nous a permis de parcourir diverses méthodes appliquées sur le signal de parole d'une manière générale. Elle nous a donné une base de connaissances pour mener notre propre étude de comparaison sur le signal de cri des nourrissons.

Ce chapitre était dédié à la présentation de la fréquence fondamentale dans la parole comme dans le cri. Nous avons ensuite présenté les problèmes les plus connus d'estimation de la fréquence fondamentale. Enfin, nous avons exposé les différentes catégories d'algorithmes dans la littérature (temporelle, spectrale et hybride).

L'algorithme SIFT présente un problème pour l'identification correcte des sous-harmoniques de la fréquence fondamentale. L'algorithme SIFT modifié l'a corrigé en utilisant le lissage des contours de la fréquence fondamentale.

L'affaiblissement de la fonction AMDF est corrigé par un simple ajustement des blocs du signal traité. Un seuil de deux tiers d'amplitude du dernier échantillon de chaque trame est ajouté à ce dernier pour aligner au maximum la fonction AMDF. De plus, la multiplication de la fonction ACF avec l'AMDF en miroir améliore la résolution de l'extraction de la fréquence fondamentale.

Le problème de l'identification et de la séparation du son voisé du son inspiratoire dans un signal de cri est résolu par la détection de l'énergie d'amplitude pour chaque segment enregistré avec la fonction STE. L'utilisation de la fonction HPS supprime les bruits pour mieux sélectionner les sons voisés des sons inspiratoires.

Un autre avantage de l'HPS consiste à résoudre le problème des perturbations produites par l'interaction entre le conduit vocal et l'excitation glottale. En effet, l'HPS a été combiné avec LPC-based Cepstrum pour améliorer la performance de cette dernière.

Dans le second chapitre, nous présentons une description plus approfondie suivie d'une simulation Matlab des algorithmes AMDF, ACF et HPS. Dans la suite du chapitre, nous présentons une étude théorique sous forme d'un schéma bloc de l'algorithme SIFT modifié. La base de données utilisée est un ensemble d'enregistrements des cris des nouveau-nés qui sont déjà traités pour filtrer le bruit et le silence.

CHAPITRE 2

IMPLÉMENTATION DES ALGORITHMES DE DÉTECTION DE LA FRÉQUENCE FONDAMENTALE

2.1 Introduction

Ce chapitre aborde essentiellement les algorithmes AMDF, ACF, HPS et le SIFT modifié. L'objectif est de simuler sous Matlab les algorithmes capables de détecter la fréquence fondamentale des signaux de cris. Dans un premier temps, nous présenterons dans la section 2.2 la structure générale d'un algorithme de détection de la fréquence fondamentale sous forme de blocs de traitement spécifique. Dans un second temps, nous donnerons une description détaillée du fonctionnement des algorithmes choisis pour réaliser cette étude.

2.2 Schéma général d'un algorithme

Généralement, la chaîne de traitement des algorithmes d'estimation de la fréquence fondamentale est composée de trois phases essentielles qui sont la phase de prétraitement, la phase de traitement et la phase de post-traitement tel que montré à la Figure 2.1.

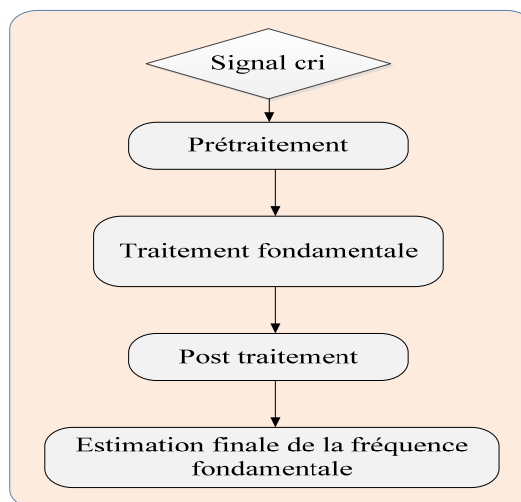


Figure 2-1 Les blocs essentiels d'un algorithme de détection de la fréquence fondamentale

La phase de prétraitement consiste à préparer le signal et le mettre sous une forme utilisable à l'analyse en utilisant la segmentation et d'autres outils. Parmi les outils utilisés, citons le fenêtrage de Hamming et le choix de la durée des trames analysées. Cette dernière varie généralement entre 20 et 30 ms. Le but de cette étape est d'élaborer la bonne représentation de la structure périodique d'un signal en supprimant l'information inutile comme (le bruit, le silence, etc.) et en extrayant l'information utile.

La phase de traitement est la partie la plus importante. Celle-ci dépend de l'algorithme choisi. Dans cette étape, l'estimation de la fréquence fondamentale est effectuée.

La phase post-traitement est nécessaire pour corriger et diminuer les erreurs grossières qui peuvent être produites lors du calcul des valeurs de fréquences par l'algorithme choisi et sert aussi à appliquer des algorithmes de lissage des valeurs estimées de la fréquence fondamentale (Ykhlef et al., 2007).

2.3 Les quatre méthodes de détection de la fréquence fondamentale

Les quatre méthodes AMDF, ACF, HPS et le SIFT modifié sont les algorithmes les plus utilisés dans la littérature pour la détection de la fréquence fondamentale des signaux musicaux et de parole. Leurs applications sur les signaux de cri des nouveau-nés restent encore un domaine d'exploration et de recherche. Cela nous a motivés pour les étudier et tester leurs performances sur les signaux de cris. Une description bien détaillée de leurs fonctionnements est présentée sous forme d'un schéma bloc.

Pour pouvoir comparer les quatre algorithmes, nous avons choisi des valeurs pour les paramètres communs comme la fréquence d'échantillonnage, la longueur du segment d'un signal à analyser et le seuil d'énergie nécessaire pour distinguer entre un segment voisé et non voisé. Ce choix est basé sur les travaux trouvés dans la littérature et les notions de base du traitement de signal. La fréquence d'échantillonnage d'un signal de cri a la valeur de 44.1Khz 44.1Khz (Kheddache et Tadj, 2013b).

La longueur d'un segment, cité dans la littérature pour ce type de traitement, varie entre 20 ms et 30 ms, nous avons choisi la valeur de 20 ms (Kheddache et Tadj, 2013b). Le seuil de décision des segments voisés et non-voisés varie entre 0.3 et 0.4, nous l'avons fixé sur 0.4 (Lederman, 2010).

2.3.1 Architecture de l'algorithme AMDF

L'implémentation de l'algorithme AMDF se décompose en cinq étapes essentielles de calcul qui sont décrites par la Figure 2.2.

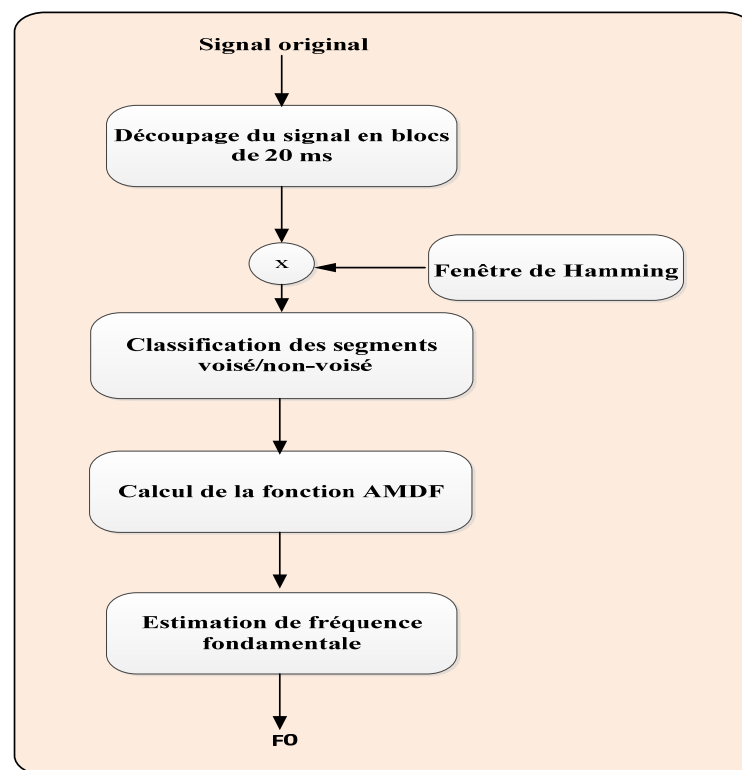


Figure 2-2 Schéma bloc de l'algorithme AMDF

1^{ère} étape : dans la phase de découpage, le signal original du cri d'un nouveau-né est segmenté en bloc de 20 ms. La fréquence d'échantillonnage utilisée est de 44 100 Hz qui correspond à un nombre de 882 échantillons par bloc. Ce nombre N est calculé de la façon suivante :

$F_e = 44.1 \text{ kHz}$ implique que : $T_e = 1/F_e = 1/44\,100 = 22.7 \mu\text{secs}$.

Le nombre d'échantillons N durant un temps de 20 ms est le résultat de la division de 20 ms sur la durée d'un échantillon de $22.7 \mu\text{sec}$. Donc $N = 20 \text{ ms} / 22.7 \mu\text{sec} = 882$.

2^{ème} étape : c'est une étape commune pour la plupart des algorithmes d'estimation de la fréquence fondamentale. En termes de spectre, la fenêtre de Hamming est utilisée pour augmenter le rapport signal/bruit des pics du signal échantillonné (Cormier, 2013).

3^{ème} étape : l'énergie de chaque segment de 20 ms est calculée. Ensuite, une condition de décision est appliquée sur cette valeur d'énergie qui sert à différencier entre les segments voisés et non voisés. La condition applique une comparaison avec un seuil d'énergie d'une valeur de 0.4. Si les segments ont une valeur d'énergie supérieure à 0.4, ils sont classés comme des segments voisés. Dans le cas contraire, les segments sont classés comme des segments non voisés.

4^{ème} étape : la fonction AMDF est calculée uniquement pour les segments voisés. Les segments non voisés se verront attribuer une fréquence nulle. L'AMDF d'un segment donné est calculé de la façon suivante : Le segment sera décalé de 1 à $N/2$ fois et après chaque décalage, la somme de la différence entre le signal original et celui décalé sera calculée. Ceci génère un vecteur de dimension $N/2$. Par la suite, le minimum de ce vecteur donne l'indice de position de la valeur minimale. La recherche du minimum de l'AMDF est appliquée sur un intervalle de temps de t_1 à t_2 . Le choix de t_1 et t_2 est justifié par la forme de la fonction résultante d'AMDF.

Si $t_1 = 0 \text{ ms}$, le pic d'AMDF est toujours minimal à ce point et par conséquence, le minimum de la fonction AMDF qui présente une forme d'onde descendante fournit toujours un résultat erroné de la fréquence fondamentale au point $t_1=0 \text{ ms}$. Nous avons opté pour la valeur de $t_1=2 \text{ ms}$ (Ykhlef et al., 2007) et $t_2 = (N/2) \text{ ms}$.

5^{ème} étape : elle représente la dernière étape qui nous permet d'obtenir le vecteur final de la fréquence fondamentale du signal traité. En effet, la fréquence fondamentale de chaque segment est calculée puis enregistrée dans un vecteur de fréquence. La relation suivante permet de calculer la fréquence fondamentale: $F_0 = F_e / (\text{indice} + t_1 - 1)$.

2.3.2 Architecture de l'algorithme ACF

L'implémentation de l'algorithme ACF se décompose en cinq étapes essentielles de calcul qui sont décrites par la Figure 2.3.

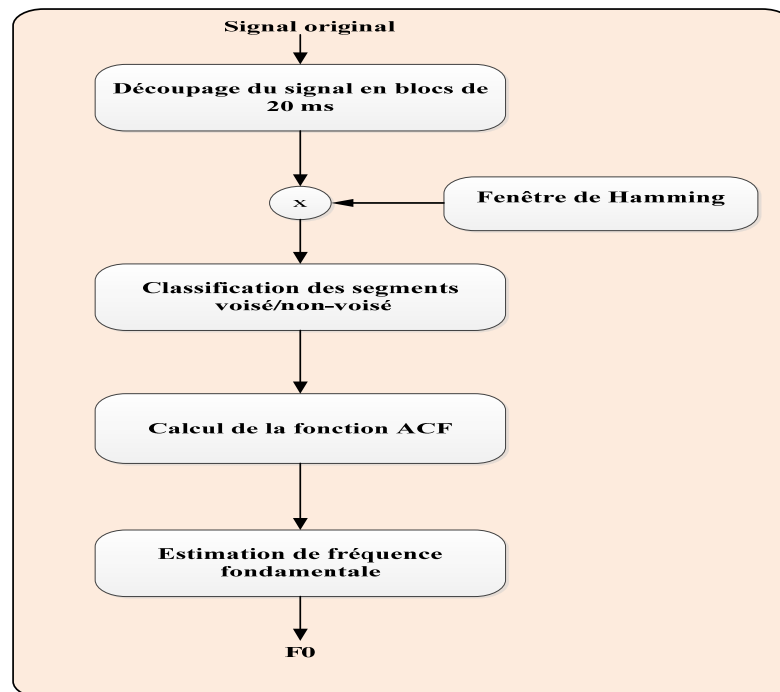


Figure 2-3 Schéma bloc de l'algorithme ACF

1^{ère} étape : le signal original du cri dans cette étape sera découpé en segment de 20 ms, avec une fréquence d'échantillonnage de 44100Hz qui correspond à un nombre de 882 échantillons par bloc.

2^{ème} étape : cette étape consiste à multiplier chaque segment de 20 ms par la fenêtre de Hamming.

3^{ème} étape : l'énergie de chaque segment de 20 ms est calculée. Ensuite, une condition de décision est appliquée sur cette valeur d'énergie qui sert à différencier les segments voisés et non voisés. La condition applique une comparaison avec un seuil d'énergie d'une valeur de 0.4. Si les segments ont une valeur d'énergie supérieure à 0.4, ils sont classés comme des segments voisés. Dans le cas contraire, les segments sont classés comme des segments non voisés.

4^{ème} étape : la fonction d'autocorrélation utilisée est la fonction $c=xcorr(x)$ implémentée dans Matlab et représente la séquence d'autocorrélation du vecteur x . Étant donné que la fonction d'autocorrélation est paire, seule la partie positive de la fonction sera traitée. La somme du produit entre le signal original et le signal décalé sera calculée et enregistrée dans un vecteur. Par la suite, le maximum de ce vecteur donne l'indice de position de la valeur maximale. La recherche du maximum de l'ACF est appliquée sur un intervalle limité entre t_1 et t_2 . Le choix de t_1 et t_2 est justifié par la forme de la fonction résultante d'ACF. Si $t_1 = 0$ ms, le pic d'ACF est toujours maximal à ce point et par conséquent, le maximum de la fonction ACF qui présente une forme d'onde descendante fournit un résultat erroné de la fréquence fondamentale toujours au point $t_1=0$ ms. Nous avons opté pour la valeur de $t_1=2$ ms et $t_2 = (N/2)$ ms.

5^{ème} étape : elle représente la dernière étape qui sert à calculer la fréquence fondamentale de chaque segment traité suivant la relation : $F_0 = F_e / (\text{indice} + t_1 - 1)$, puis à enregistrer le résultat dans un vecteur de fréquence.

2.3.3 Architecture de l'algorithme HPS

L'implémentation de l'algorithme HPS se décompose en cinq étapes essentielles de calcul qui sont décrites par la Figure 2.4.

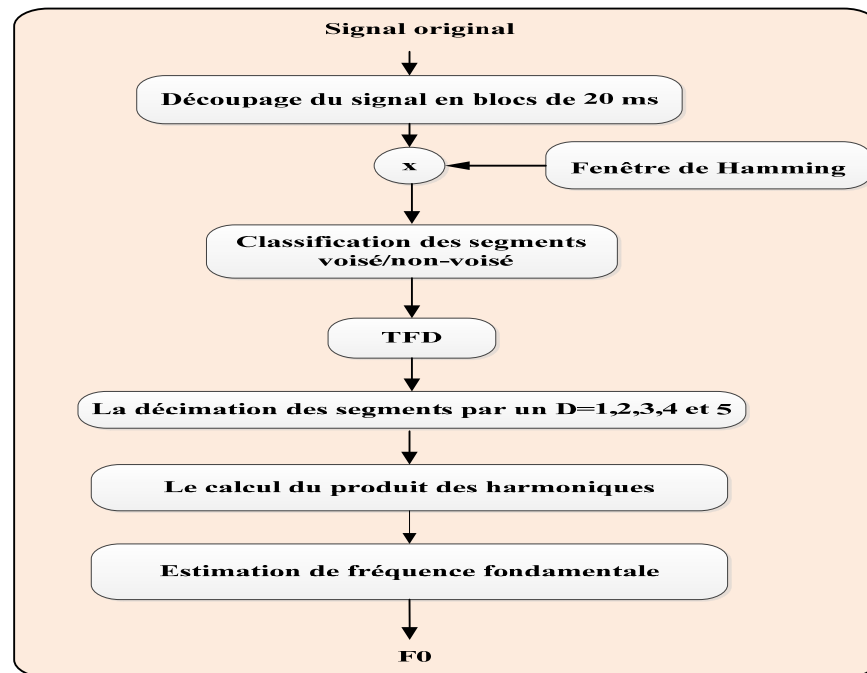


Figure 2.4 Schéma bloc de l'algorithme HPS

1^{ière} étape: à cette étape, le signal original du cri d'un nouveau-né sera découpé en segments de 20 ms en utilisant une fréquence d'échantillonnage de 44100Hz qui correspond à un nombre de 882 échantillons par bloc.

2^{ème} étape: chaque segment de 20 ms est multiplié par la fenêtre de Hamming.

3^{ème} étape : c'est la même étape de classification des segments voisés et non voisés du signal analysé, décrite dans le troisième bloc des algorithmes ACF et AMDF.

4^{ème} étape: dans ce bloc, la transformée de Fourier est appliquée sur les coefficients résultants de l'équation de la multiplication précédente et produit ainsi le spectre TFD du signal de 20 ms représenté par la fréquence fondamentale et ses harmoniques qui sont des multiples d'un entier ($2F_0$, $3F_0$, $4F_0$, $5F_0$, ...).

5^{ème} étape: l'application de la décimation fréquentielle avec les valeurs $1/2$, $1/3$, $1/4$ et $1/5$ est utilisée pour le signal qui sort du troisième bloc.

Ensuite, le calcul de la dimension d est effectué. La dimension d délimite le nombre commun de coefficients à utiliser pour chaque vecteur $\text{TFD}_{1/D}$. Elle représente la plus grande valeur de décimation $D = 5$ divisé par 2.

Le résultat du maximum du produit des coefficients TFD avec les coefficients décimés de $\text{TFD}_{1/2}$, $\text{TFD}_{1/3}$, $\text{TFD}_{1/4}$ et $\text{TFD}_{1/5}$ donne l'information nécessaire sur la valeur exacte de la fréquence fondamentale F_0 pour un bloc du signal traité.

2.3.4 Architecture de l'algorithme SIFT modifié

Conformément à ce que nous avons présenté dans le premier chapitre, l'implémentation de l'algorithme SIFT modifié, présentée sur la Figure 2.5, est basée sur six blocs que nous décrirons par la suite. Les résultats de simulation du SIFT modifié ont déjà été publiés par (Yasmina Kheddache, 2013), et seront utilisés pour la comparaison de la performance de cet algorithme avec les algorithmes AMDF, ACF, HPS et le Praat.

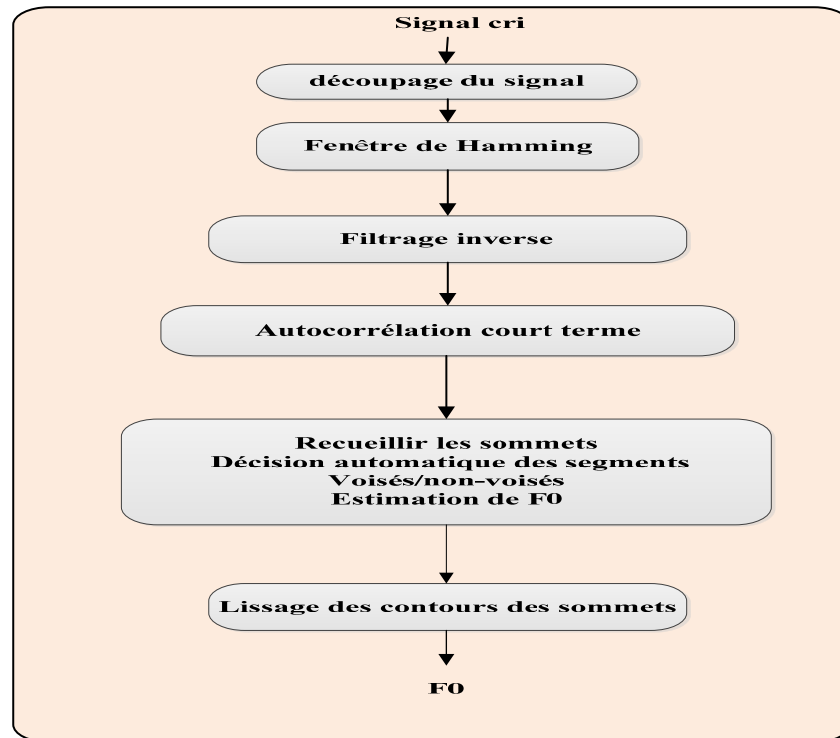


Figure 2.5 Diagramme bloc de l'algorithme SIFT modifié
Tirée de Kheddache et Tadj (2013, p.6)

1^{ère} étape : lors de l'opération de découpage du signal, chaque segment de 20 ms est mis en chevauchement de 10 ms avec les segments voisins. Étant donné que la fréquence d'échantillonnage est de 44.1 kHz, le nombre d'échantillons d'un segment de 20 ms est de 882.

2^{ème} étape : on applique le fenêtrage d'Hamming sur chaque segment.

3^{ème} étape : durant cette étape, on applique le filtrage inverse afin d'atténuer l'influence du conduit vocal. Le choix raisonnable de l'ordre du filtre qui a été retenu est $M = 4$ (Fort et Manfredi, 1998).

4^{ème} étape : l'estimation de la séquence d'autocorrélation du signal est effectuée à cette étape.

5^{ème} étape : lors de cette étape, les sommes de l'autocorrélation sont recueillies. Un algorithme de décision est appliqué par la suite. Durant cette opération, la valeur du pic de l'autocorrélation est comparée à un seuil voisé choisi de 0,4. Ce seuil sert à différencier entre les segments voisés et non voisés. Un segment d'énergie spectrale inférieur à 0.4 sera considéré comme segment non voisé. Un segment d'énergie spectrale supérieur à 0.4 sera considéré comme segment voisé. La fréquence estimée est définie par la formule suivante :

$$T_0 = \frac{1}{F_0} = \underset{n}{\operatorname{argmax}} \{ri(\eta)\} \quad (2.1)$$

Où $ri(\eta)$ est la séquence d'autocorrélation.

6^{ème} étape : la dernière étape consiste à lisser le contour de la fréquence fondamentale obtenue. Ceci est réalisé en utilisant un filtre médian d'ordre 4, afin de faire limiter les changements brusques dans le contour de la fréquence fondamentale et d'obtenir un contour lissé.

2.4 Conclusion

Dans ce chapitre nous avons présenté le principe de fonctionnement des algorithmes : AMDF, ACF et HPS implémentés dans Matlab sous forme de schémas blocs, ainsi qu'une explication détaillée pour chaque tâche de calcul réalisée dans le code Matlab.

De la même façon, nous avons exposé le principe de fonctionnement de l'algorithme SIFT modifié tel qu'il a été déjà implémenté dans Matlab par notre collègue.

CHAPITRE 3

RÉALISATION DE L'ÉTUDE COMPARATIVE DES MÉTHODES D'EXTRACTION DE LA FRÉQUENCE FONDAMENTALE

3.1 Introduction

Ce chapitre présente une étude comparative des différentes techniques d'extraction de la fréquence fondamentale. Cette étude comparative est validée par des simulations sous Matlab.

Le manque des valeurs réelles de la fréquence fondamentale de notre base de données nous a amené à utiliser le test statistique de la validation croisée afin de comparer les valeurs réelles avec celles estimées par les techniques étudiées (Preux, 2011).

La validation croisée consiste à diviser la base de données en deux parties : la partie de référence et la partie du test comme il est illustré sur la Figure 3.1.

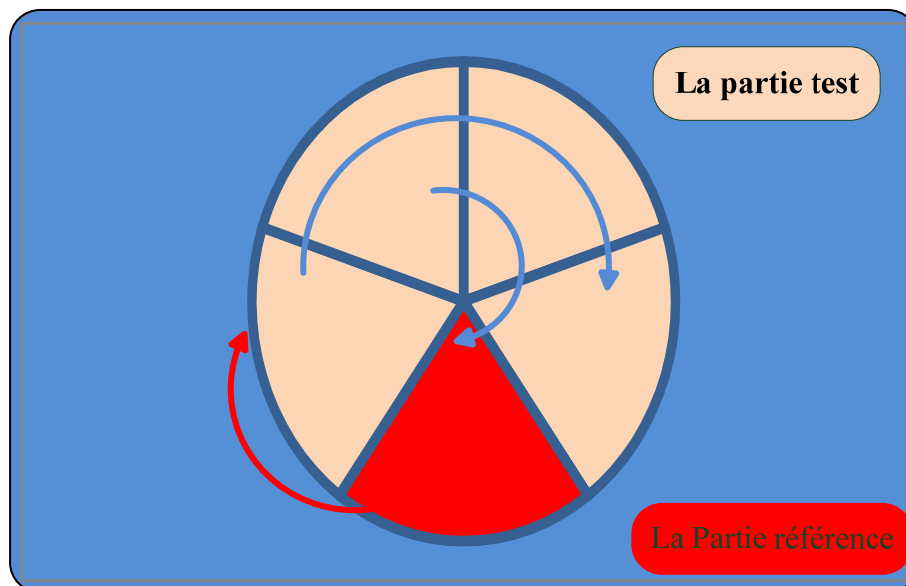


Figure 3-1 L'application de la validation croisée à cinq blocs sur la comparaison des techniques d'estimation de la fréquence fondamentale

La partie test est composée par les résultats de simulation de quatre techniques et la partie de référence est composée par les résultats d'une seule technique. En effet, à chaque fois nous fixons une méthode : L'AMDF, l'ACF, l'HPS, le SIFT et le Praat comme référence de comparaison. Les autres méthodes seront utilisées pour le test, puis nous répétons cette tâche cinq fois avec rotation des méthodes.

A titre d'exemple, nous supposons que la partie référence représente l'AMDF, ce qui implique qu'elle est la technique la plus performante pour la comparaison. Par la suite, les résultats d'estimation de la fréquence fondamentale de chaque méthode de la partie test seront comparés avec les valeurs estimées par la technique de référence afin de calculer les erreurs produites durant l'extraction de la fréquence fondamentale.

Prenons en considération le choix fixé au début de la partie référence et selon les techniques d'estimation de la fréquence fondamentale existantes. Nous obtiendrons les cinq différents cas d'analyse selon le tableau suivant :

Tableau 3.1 Les cinq différents cas d'analyse selon la référence choisie

Référence	Test
AMDF	ACF, HPS, SIFT modifié et PRAAT
ACF	AMDF, HPS, SIFT modifié et PRAAT
HPS	AMDF, ACF, SIFT modifié et PRAAT
SIFT modifié	AMDF, ACF, HPS et PRAAT
Praat	AMDF, ACF, HPS et SIFT modifié

Cette étude comparative nous permet de déterminer l'algorithme qui offre les meilleurs résultats d'extraction de la fréquence fondamentale.

3.2 Description de la méthode de comparaison

Une base de données composée de cinquante signaux des cris chez des nouveau-nés est utilisée pour réaliser ce travail de comparaison. La durée des signaux utilisés varie entre 7 et 8 secondes. Les résultats d'exécution de la simulation des signaux utilisés sous PRAAT donnent des fréquences fondamentales variables selon les caractéristiques des segments d'un signal donné.

Segment non défini : représente les valeurs des fréquences fondamentales non définies par une technique de détection donnée.

Segment non voisé : représente les valeurs des fréquences fondamentales égales à 0 Hz (Kheddache et Tadj, 2013a).

Segment phonique : ce sont les valeurs des fréquences fondamentales inférieures à 750 Hz (Kheddache et Tadj, 2013a).

Segment haute-fréquence «High-pitch»: représente les valeurs des fréquences fondamentales qui varient entre 750 et 1000 Hz (Kheddache et Tadj, 2013a).

Segment hyperphonique : représente les fréquences fondamentales qui varient entre 1000 et 2000 Hz (Kheddache et Tadj, 2013a).

Chaque intervalle de fréquence est représenté par un chiffre qui correspond à un mode de segment spécifique. Le but de cette représentation est de transformer la matrice des fréquences fondamentales en une matrice simplifiée par des chiffres de 1 à 5 selon le schéma suivant :

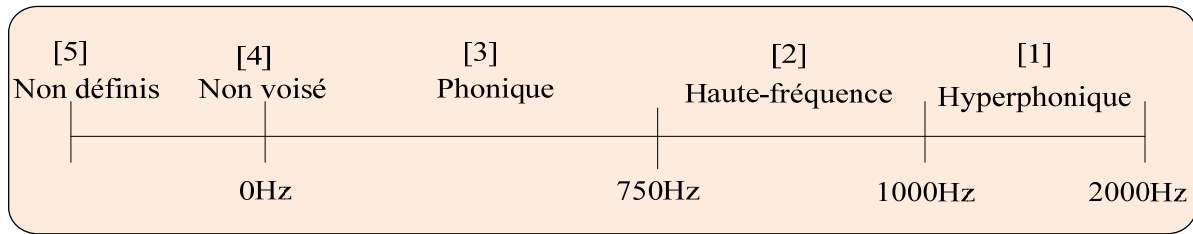


Figure 3-2 Les modes des segments qui correspondent à l'intervalle de variation de la fréquence fondamentale

Le calcul des différents types d'erreur sera effectué selon la nouvelle base de données qui est construite selon les modes des segments traités.

3.3 Analyse des résultats expérimentaux

Les paramètres d'erreur appliqués sont exprimés en pourcentage et sont définis comme suit.

L'erreur hyperphonique (Er_{Hy_pci}) se présente quand on analyse avec le modèle de référence qui classe le segment traité en segment hyperphonique, alors que le modèle de test le classe comme haute-fréquence, phonique, non voisé ou non défini. En fonction de la classification du modèle de test, l'erreur hyperphonique se décompose en plusieurs ordres $i=1, 2, 3$ et 4 cités dans le tableau suivant :

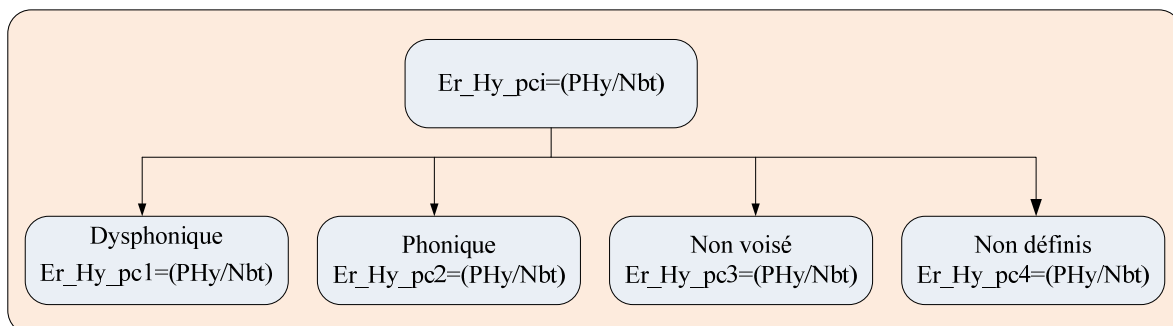


Figure 3-3 L'ordre de l'erreur hyperphonique Er_{Hy_pc} selon le mode du segment traité par le modèle de test

PHy est le nombre de segments dont le mode dans le modèle de test est différent de celui du mode hyperphonique du modèle référence. Le paramètre Nbt représente le nombre de segments du signal traité.

Exemple : si $i=1$; nous avons une erreur hyperphonique d'ordre 1 (Er_Hy_pc1) ; c'est le cas où le modèle de référence classifie le segment traité en segment hyperphonique, tandis que le modèle de test le classifie comme haute-fréquence.

L'erreur haute-fréquence (Er_Hi_pci) se présente dans le cas d'analyse du modèle de référence qui classifie le segment traité en segment haute-fréquence, alors que le modèle de test le classifie comme hyperphonique, phonique, non voisé ou non défini. En fonction de la classification du modèle de test, l'erreur haute-fréquence se décompose en plusieurs ordres cités dans le tableau suivant :

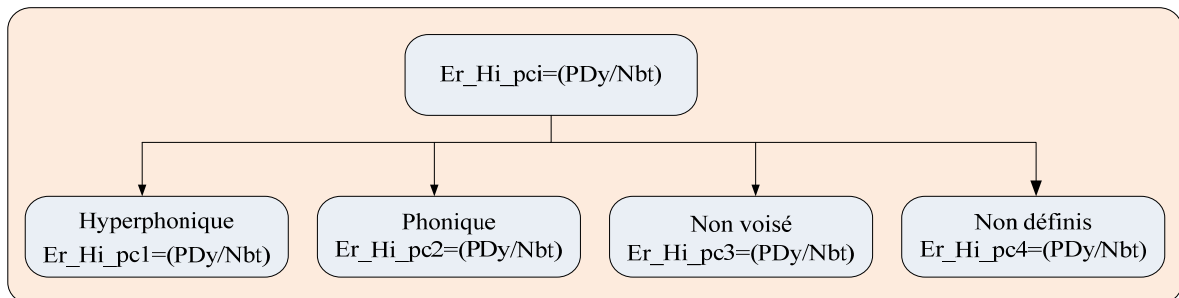


Figure 3-4 L'ordre de l'erreur haute-fréquence (Er_Hi_pc) selon le mode du segment traité par le modèle de test

PDy est le nombre de segments dont le mode dans le modèle de test est différent de celui du mode haute-fréquence du modèle de référence.

L'erreur phonique (Er_Ph_pci) se présente dans le cas d'analyse du modèle de référence qui classifie le segment traité en segment phonique, alors que le modèle de test le classifie comme hyperphonique, haute-fréquence, non voisé ou non défini. Selon la classification du modèle de test, l'erreur phonique se décompose en plusieurs ordres cités dans le tableau suivant.

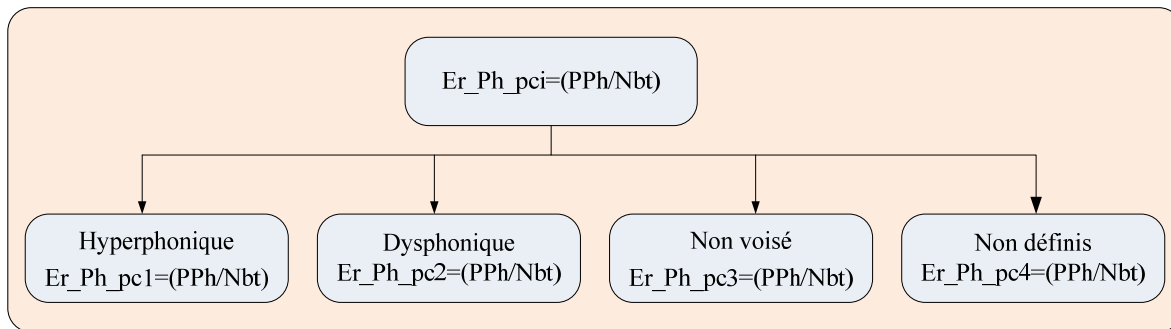


Figure 3-5 L'ordre de l'erreur phonique (Er_Ph_pc) selon le mode du segment traité par le modèle de test

PPh est le nombre de segments dont le mode dans le modèle de test est différent de celui du mode phonique en modèle référence.

L'erreur non voisée (Er_Nv_pci) se présente dans le cas d'analyse du modèle de référence qui classifie le segment traité en segment non voisé, alors que le modèle de test le classifie comme hyperphonique, haute-fréquence, phonique ou non défini. Selon la classification du modèle de test, l'erreur non voisée se décompose en plusieurs ordres cités dans le tableau suivant.

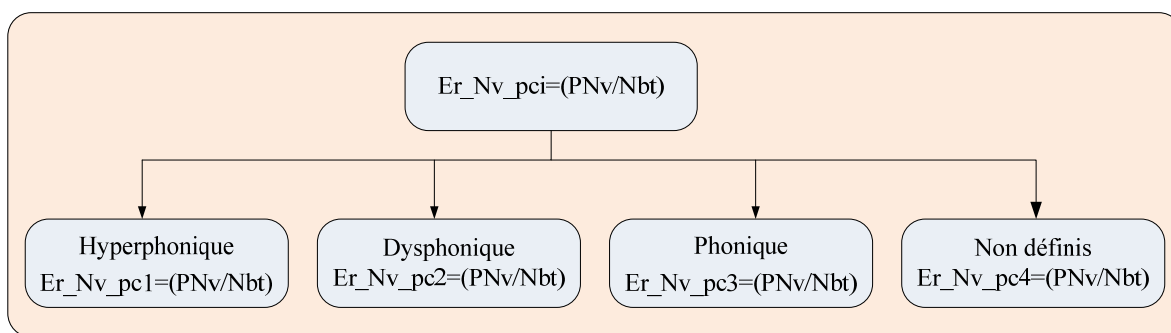


Figure 3-6 L'ordre de l'erreur non voisée (Er_Nv_pc) selon le mode du segment traité par le modèle de test

PNv est le nombre de segments dont le mode, dans le modèle de test, est différent de celui du mode non voisé en modèle référence.

L'erreur non définie (Er_Nan_pci) se présente dans le cas d'analyse du modèle référence qui classe le segment traité en segment non défini, par contre le modèle de test le classe soit hyperphonique, haute-fréquence, phonique ou non voisé. Selon la classification du modèle de test, l'erreur non définie se décompose en plusieurs ordres présentés ci-dessous (Figure 3.7).

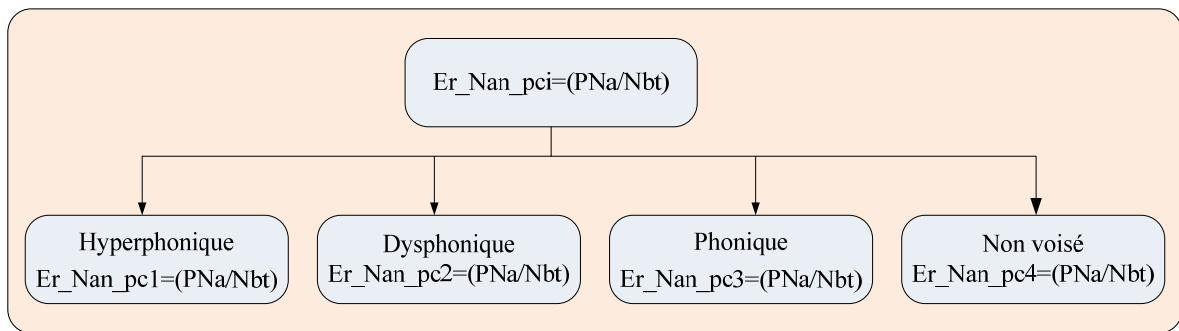


Figure 3-7 L'ordre de l'erreur non défini (Er_Nan_pc) selon le mode du segment traité par le modèle de test

PNa est le nombre de segments dont le mode dans le modèle de test est différent de celui du mode non défini en modèle référence.

3.3.1 Analyse des résultats de l'AMDF

Dans cette partie, est présentée une analyse quantitative et qualitative des différents algorithmes. Nous considérons l'algorithme AMDF comme modèle de référence et nous le comparons aux algorithmes ACF, HPS, SIFT modifié, ainsi qu'au logiciel Praat.

Définissons M_err_Hi , M_err_Ph , M_err_Nv , M_err_Hy et M_err_Nan comme étant respectivement les moyennes des erreurs : Er_Hi_pci (haute-fréquence), Er_Ph_pci (phonique), Er_Nv_pci (non voisée), Er_Hy_pci (hyperphonique) et Er_Nan_pci (non défini) pour chaque méthode par rapport l'AMDF. Ces paramètres sont calculés selon :

$$M_{err_Hi} = \frac{1}{4} * \sum_{i=1}^4 Er_Hi_pc_i \quad (3.1)$$

$$M_{err_Ph} = \frac{1}{4} * \sum_{i=1}^4 Er_Ph_pc_i \quad (3.2)$$

$$M_{err_Nv} = \frac{1}{4} * \sum_{i=1}^4 Er_Nv_pc_i \quad (3.3)$$

$$M_{err_Hy} = \frac{1}{4} * \sum_{i=1}^4 Er_Hy_pc_i \quad (3.4)$$

$$M_{err_Nan} = \frac{1}{4} * \sum_{i=1}^4 Er_Nan_pc_i \quad (3.5)$$

i représente l'ordre de chaque type d'erreur.

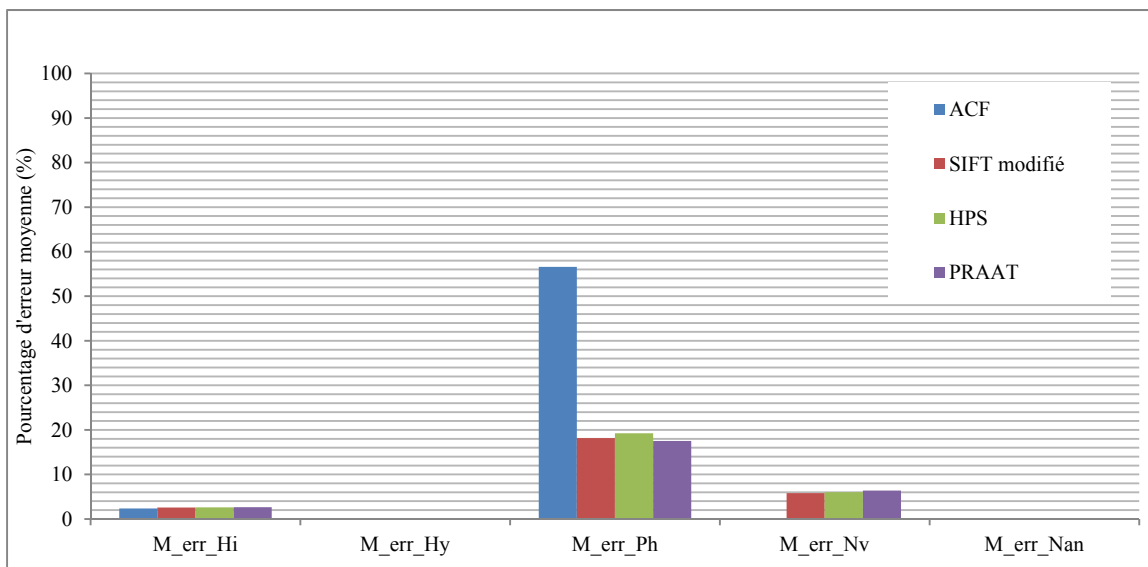


Figure 3-8 Pourcentages de la moyenne des différentes erreurs produites par rapport à l'AMDF

La Figure 3.8 représente les résultats de la moyenne des erreurs : $Er_{Hi_pc_i}$, $Er_{Ph_pc_i}$, $Er_{Hy_pc_i}$, $Er_{Nan_pc_i}$ et $Er_{Nv_pc_i}$ en pourcentage produite par les modèles de test par rapport au modèle de référence. D'après cette figure, nous remarquons que la valeur du paramètre M_err_Hy est nulle ce qui implique que l'AMDF a réussi à 100% à détecter les segments hyperphoniques.

Le résultat de l' M_err_Nan nul était prévisible parce qu'il n'y a aucun segment classifié par l'AMDF comme étant un segment non défini (Nan).

De plus, nous remarquons que la valeur de M_err_Hi garde une valeur similaire pour les quatre méthodes avec une valeur maximale de 2,64%. Ainsi, nous pouvons considérer que l'AMDF a un taux de réussite de 97% dans la classification des segments haute-fréquence.

L'AMDF présente une valeur de M_err_Nv négligeable par rapport à celle de l'ACF. Cependant il présente une valeur similaire à celles de SIFT modifié, HPS et PRAAT qui varient autour de 6%. Nous pouvons ainsi considérer les valeurs de M_err_Nv comme des résultats satisfaisants, ce qui implique que le pourcentage de fiabilité de classification des segments non voisés avec l'AMDF est de 93%.

D'autre part, nous observons que le taux de M_err_Ph varie autour 20% pour chacune des techniques SIFT modifié, HPS et PRAAT. Ce taux présente un dépassement de 50% pour l'ACF. Nous pouvons conclure que l'AMDF est moins fiable pour identifier les segments phoniques.

Dans le but de généraliser l'analyse des résultats de la méthode AMDF, nous définissons les paramètres suivants :

$$MM_{Nv} = \frac{1}{4} * \sum_{k=1}^4 M_{err_Nv_k} \quad (3.6)$$

$$MM_{Hy} = \frac{1}{4} * \sum_{k=1}^4 M_{err_Hy_k} \quad (3.7)$$

$$MM_{Hi} = \frac{1}{4} * \sum_{k=1}^4 M_{err_Hi_k} \quad (3.8)$$

$$MM_{Ph} = \frac{1}{4} * \sum_{k=1}^4 M_{err_Ph_k} \quad (3.9)$$

$$MM_{Nan} = \frac{1}{4} * \sum_{k=1}^4 M_{err_Nan_k} \quad (3.10)$$

MM_{Nv} , MM_{Hy} , MM_{Hi} , MM_{Ph} et MM_{Nan} représentent respectivement la moyenne de chacun des paramètres $M_{err_Nv_k}$, $M_{err_Hy_k}$, $M_{err_Hi_k}$, $M_{err_Ph_k}$ et $M_{err_Nan_k}$ relatifs aux différentes méthodes SIFT modifié, HPS, ACF et PRAAT par rapport à l'AMDF. L'indice $k = 1, 2, 3, 4$ représente chacune des méthodes testées SIFT modifié, HPS, ACF et PRAAT.

Les résultats du calcul précédant sont présentés dans la Figure 3.9. Ces résultats ont confirmé que l'AMDF est fiable lors de l'estimation des segments hyperphoniques, haute-fréquence et non voisés avec une moyenne d'erreur qui varie entre 0% et 4.6%. Cependant, il s'est révélé moins précis pour la détection des segments phoniques avec une moyenne d'erreur de 27.88%.

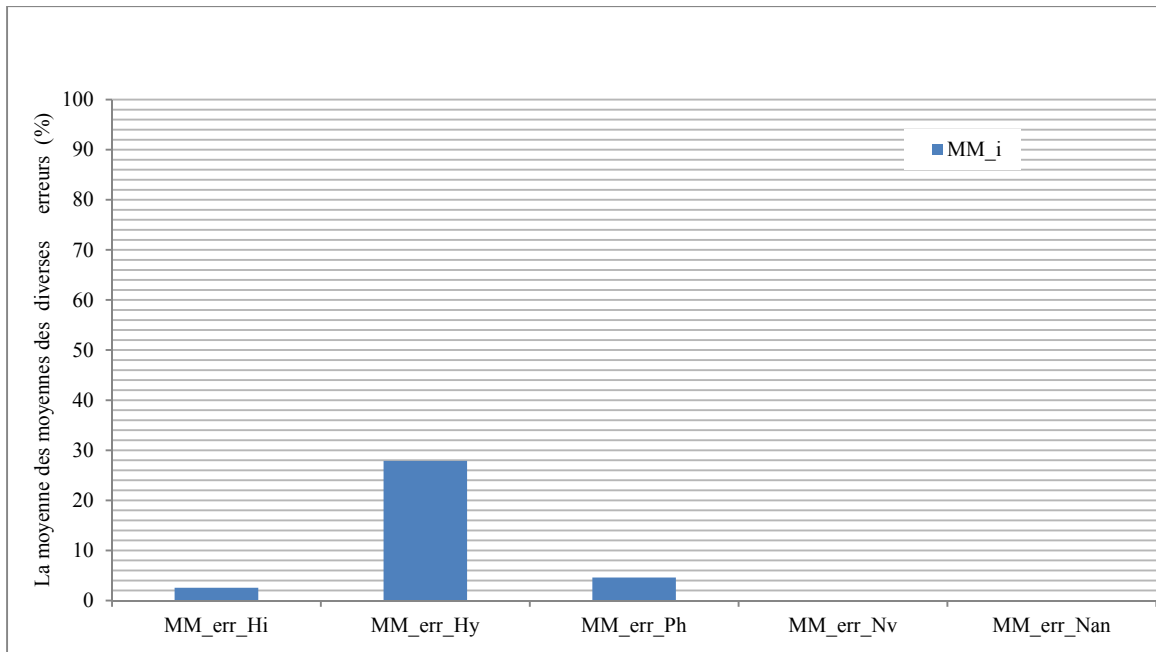


Figure 3-9 L'erreur moyenne des quatre méthodes par type d'erreur

3.3.2 Analyse des résultats de l'ACF

Nous considérons maintenant l'algorithme ACF comme modèle de référence et le comparons aux algorithmes AMDF, HPS, SIFT modifié, ainsi qu'au logiciel Praat. Nous présentons ensuite une analyse des résultats du calcul des erreurs produites lors de l'estimation de la fréquence fondamentale.

La procédure de calcul est similaire à celle présentée dans la partie d'analyse des résultats de l'AMDF. En utilisant les formules de calcul 3.1 à 3.5, nous obtenons les résultats présentés dans la Figure 3.10.

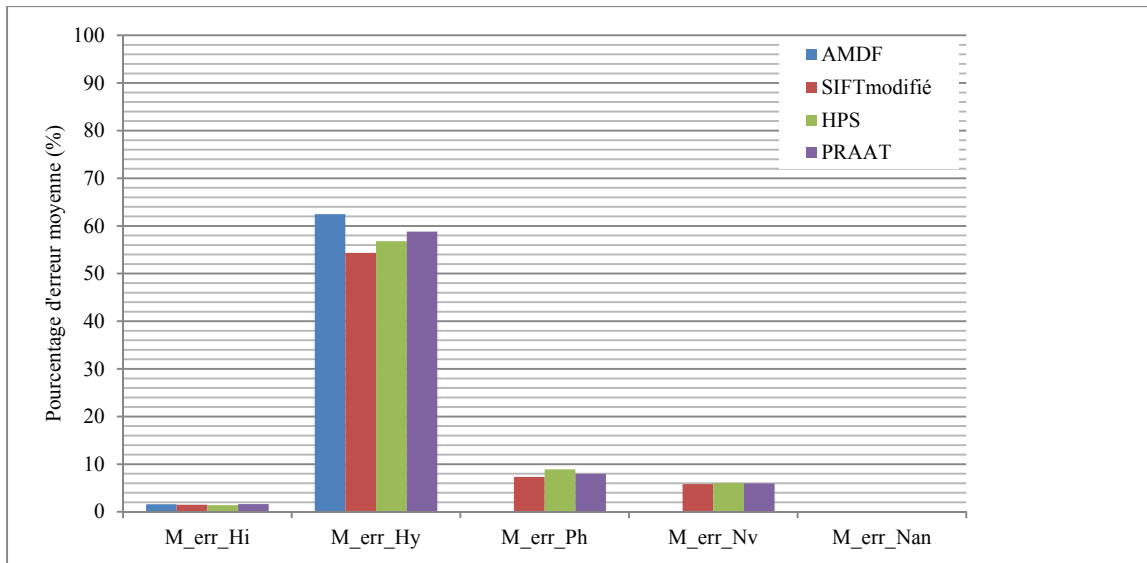


Figure 3-10 Pourcentages de la moyenne des différentes erreurs produites par rapport à l'ACF

Nous remarquons que l'ACF présente un taux d'erreur M_err_Hi minimum avec une valeur maximale de 1.63% pour les quatre méthodes d'évaluation. L'ACF a réussi donc l'extraction des segments haute-fréquence avec un pourcentage de 98%.

L'ACF présente un M_err_Nv nul par rapport à l'AMDF. Cependant, il présente une valeur similaire qui varie autour de 6% avec le SIFT modifié, l'HPS et le PRAAT, ce qui implique que l'ACF est fiable à 94% dans l'estimation des segments non voisés.

D'autre part, nous trouvons un M_err_{ph} nul pour l'AMDF. De plus il garde une valeur similaire qui converge vers 9% pour le SIFT modifié, HPS et le PRAAT. L'ACF donne ainsi une précision de 91% à classer les segments phoniques.

Nous observons que le taux de la moyenne d'erreur hyperphonique M_err_{Hy} est très élevé par rapport aux quatre méthodes : AMDF, SIFT modifié, HPS et le PRAAT.

La valeur du M_err_{ph} qui varie entre 50% et 60%, nous permet de dire que l'ACF n'est pas fiable pour détecter les segments hyperphoniques. En effet, cela confirme les résultats trouvés dans la littérature.

Nous avons refait le calcul secondaire conformément aux équations 3.6 à 3.10 pour confirmer l'analyse primaire des résultats de l'ACF par rapport au SIFT modifié, l'HPS, l'AMDF et le PRAAT. Les résultats obtenus sont présentés dans la Figure 3.11.

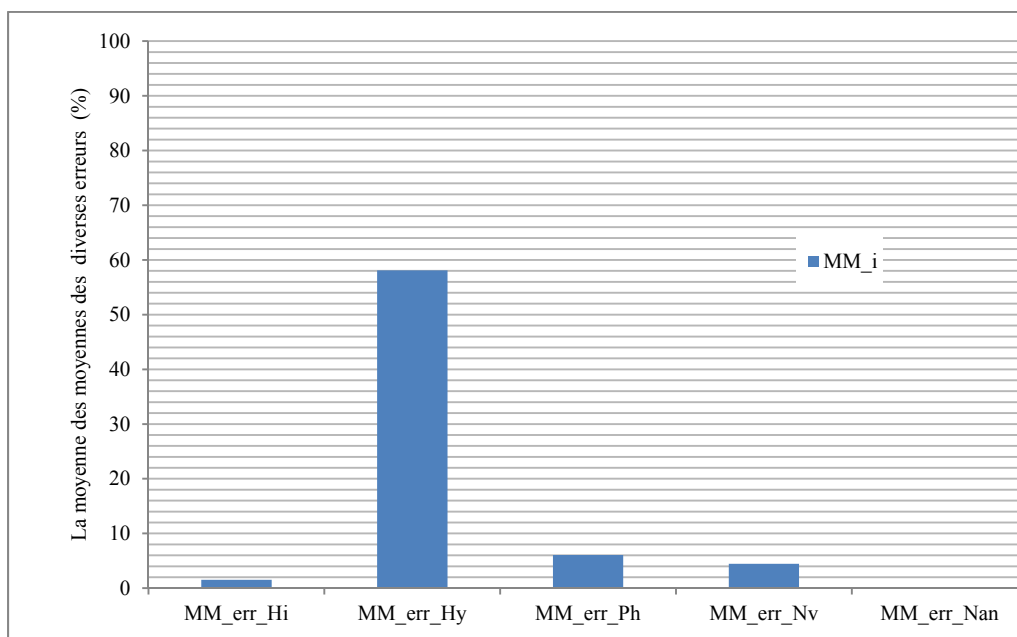


Figure 3-11 L'erreur moyenne des quatre méthodes par type d'erreur

Selon la Figure 3.11, nous constatons que l'ACF est un algorithme fiable pour reconnaître tous les segments d'un signal traité avec un M_err_Nan égal à 0%.

En plus, l'ACF a réussi de classifier les segments haute-fréquence à 98%. Il a une précision d'estimation des segments non voisés proches de 95%.

L'ACF possède un taux de précision à extraire les segments phoniques de 94%. Cependant, le taux de classification des segments hyperphoniques est proche de 40%.

3.3.3 Analyse des résultats de l'HPS

Dans cette section, est présentée une analyse quantitative et qualitative de la performance de l'HPS lors de l'estimation de la fréquence fondamentale. Nous avons refait la même procédure de calcul que citée précédemment dans 3.1 à 3.5.

Les résultats de la simulation du modèle de référence HPS avec les différents algorithmes d'apprentissage : l'AMDF, l'ACF, le SIFT modifié ainsi qu'avec le logiciel Praat, sont présentés dans la Figure 3.12.

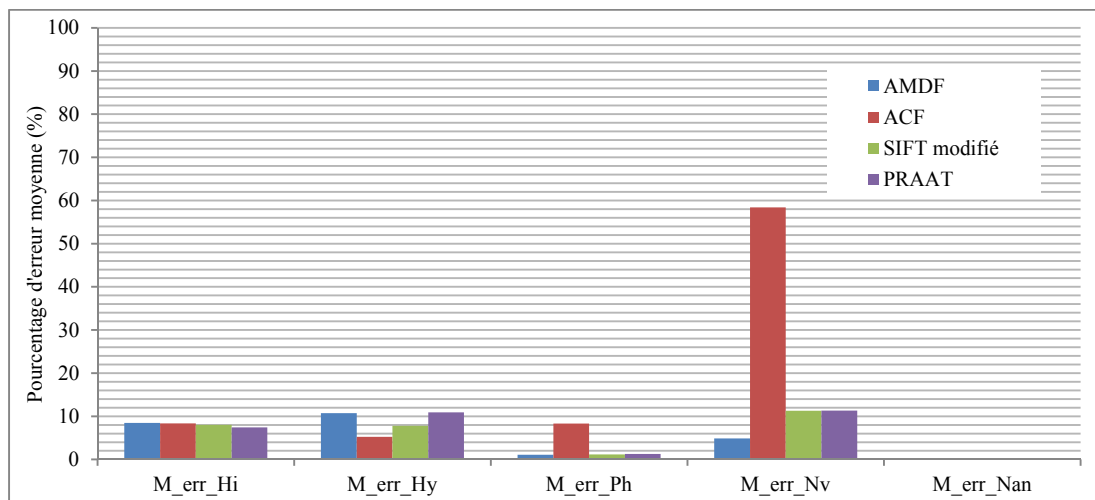


Figure 3-12 Pourcentages de la moyenne des différentes erreurs produites par rapport à l'HPS

Nous nous basons sur le M_err_Nan nul pour justifier la capacité de l'HPS à identifier ou quantifier tout segment d'un signal traité.

L'algorithme HPS est caractérisé par un taux d'M_err_Ph minimal qui ne dépasse pas 1.3% par rapport aux algorithmes SIFT modifié, l'AMDF et le logiciel Praat, alors que par rapport l'ACF le taux d'M_err_Ph égal 8.3%.

Tous les algorithmes : AMDF, le SIFT modifié, l'ACF et le logiciel Praat présentent une valeur similaire d' M_err_Hi par rapport au HPS qui varie autour 8%.

De plus, nous observons que le taux d' M_err_Hy varie entre 5% et 11% ce qui nous permet de fixer une marge de précision variant de 89% à 95% pour la détection des segments hyperphoniques par l'HPS.

Nous remarquons que l'HPS présente un taux d'erreur M_err_Nv important de 58% par rapport à l'ACF. Cependant il présente un taux assez bas variant de 4% à 11% par rapport aux AMDF, SIFT modifié et au Praat.

Afin de confirmer l'analyse des résultats précédents, nous avons réalisé le calcul de 3.6 à 3.10 où nous avons obtenu les résultats présentés dans la Figure 1.13.

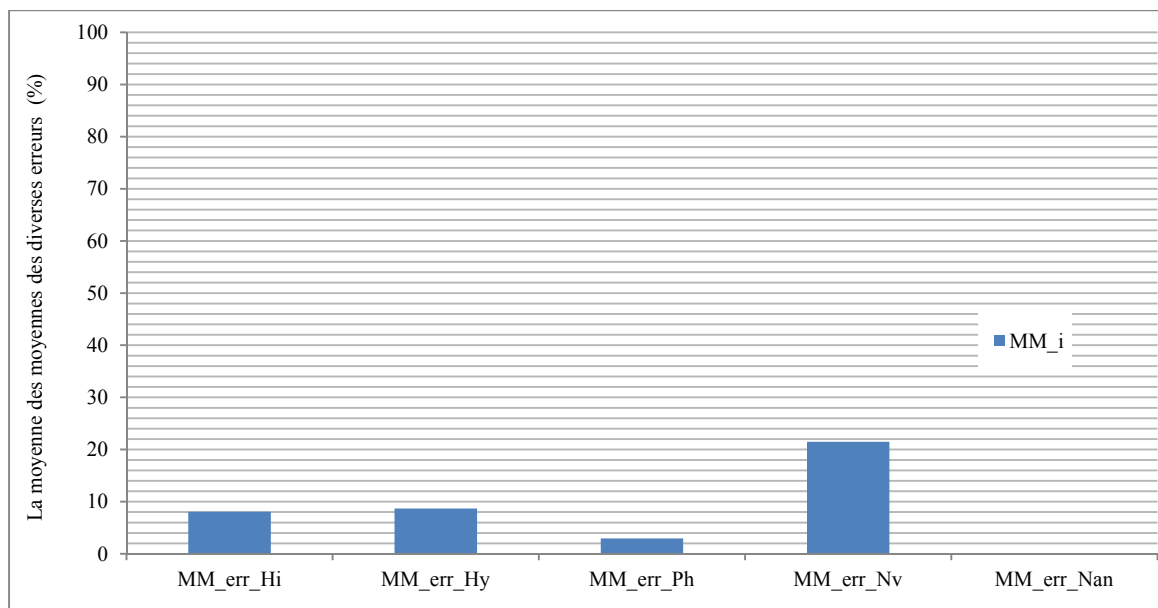


Figure 3-13 L'erreur moyenne des quatre méthodes par type d'erreur

D'après la figure ci-dessus, nous pouvons confirmer, d'après l' MM_err_Nan nul, que l'HPS n'a aucune difficulté à identifier toute partie d'un signal traité. De plus, il présente un taux d' MM_err_Ph maximal de 3%, ce qui leur donne une capacité de performance de 97% à détecter les segments phoniques.

D'autre part, le taux d'MM_err_Hy et d'MM_err_Hi ont une valeur presque similaire qui ne dépasse pas 9%, ce qui permet au HPS d'avoir une capacité maximale de 91% pour classer les segments hyperphonique et haute-fréquence. En prenant en considération le taux d'MM_err_Nv de 21.45%, nous constatons que l'HPS présente une faible précision à détecter les segments non voisés.

3.3.4 Analyse des résultats de SIFT modifié

Nous considérons dans cette partie que le modèle de référence est l'algorithme SIFT modifié et nous le comparons avec les algorithmes ACF, HPS, AMDF, ainsi qu'avec logiciel Praat.

Selon les mêmes étapes d'analyse que la section AMDF, nous avons obtenu les résultats de calcul des erreurs produites lors de la comparaison du SIFT modifié par l'ACF, l'AMDF, l'HPS et le Praat. Les résultats sont présentés dans la Figure 3.14.

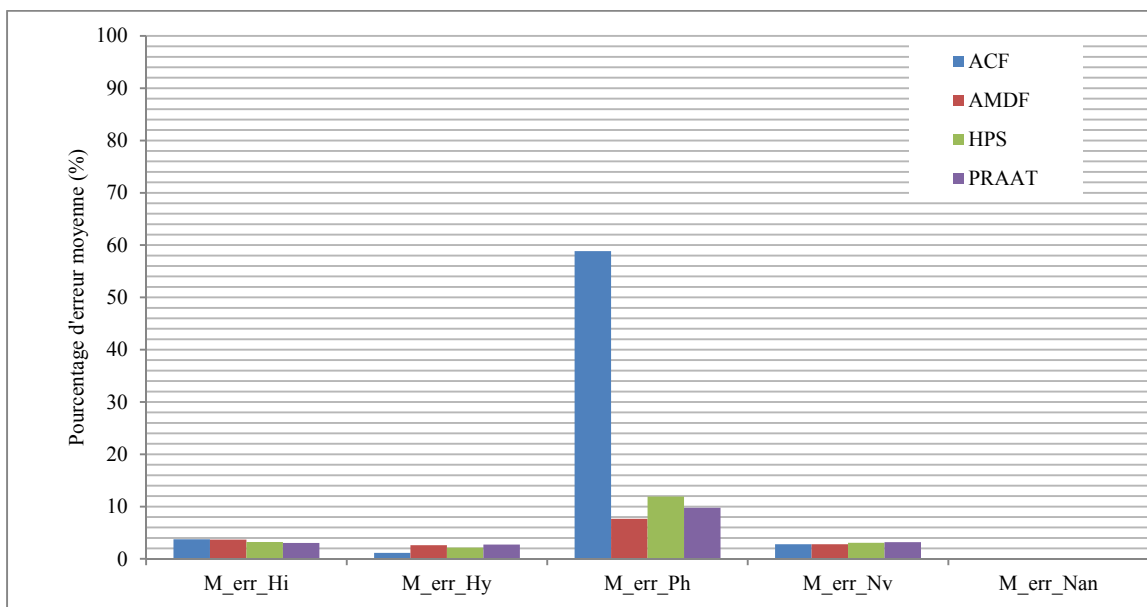


Figure 3-14 Pourcentages de la moyenne des différentes erreurs produites par rapport au SIFT modifié

D'après la Figure 3.14, nous remarquons que le taux de M_err_Hy atteint une valeur maximale de 2.5%. Ceci nous permet de considérer que l'algorithme SIFT modifié a atteint une performance de 97% dans l'extraction des segments hyperphoniques. Nous pouvons ainsi dire que c'est un algorithme ayant une haute fiabilité à reconnaître tous les segments d'un signal traité avec un M_err_Nan égal à 0%.

Le SIFT modifié possède le même taux d' M_err_Hi que l'ACF, l'AMDF, l'HPS et le Praat, variant autour de 3%. Par conséquent, le SIFT modifié présente une fiabilité de 97% pour estimer les segments haute-fréquence.

D'autre part, nous fixons une marge de variation du paramètre d' M_err_Nv entre 2% et 3%, donnant au SIFT modifié une haute précision de 97% pour estimer les segments non voisés. L'algorithme ACF possède un taux d' M_err_Ph élevé de 58.9% par rapport au SIFT modifié. Cependant, le paramètre M_err_Ph présente un taux moins élevé : de 11.9%, 9.8% et 7.7% par rapport au HPS, Praat et AMDF respectivement.

Afin de valider l'analyse primaire relative à la précision du SIFT modifié à identifier les segments du signal analysé, nous nous sommes basés sur le calcul cité au (3.6 à 3.10).

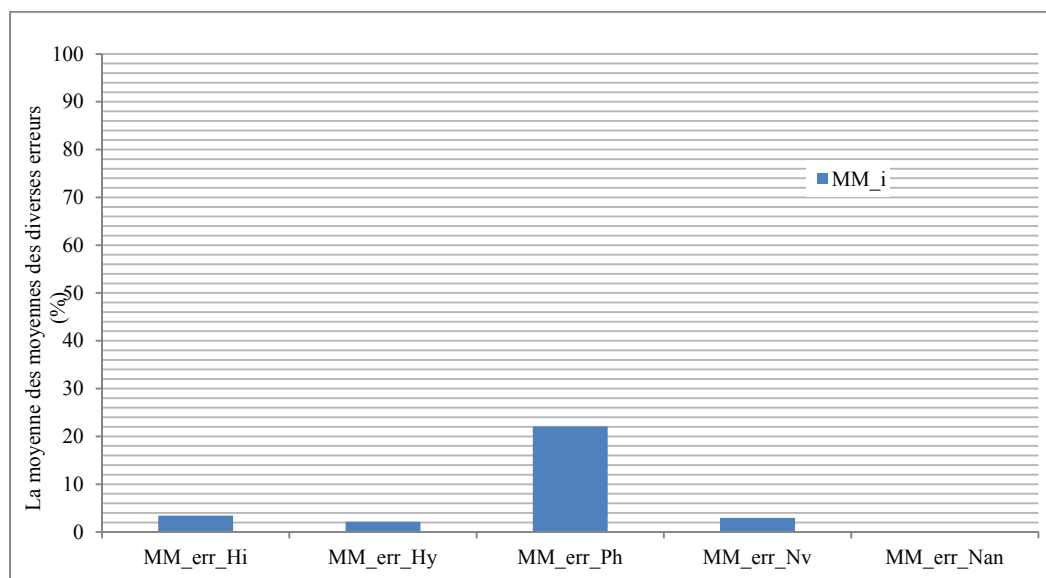


Figure 3-15 L'erreur moyenne des quatre méthodes par type d'erreur

La Figure 3.15 représente les résultats d'analyses secondaires. D'après le paramètre MM_err_Hy atteignant un maximum de 2.2%, nous constatons que le SIFT modifié possède une précision d'estimation des segments hyperphoniques qui tend vers 98%.

De même, le paramètre d'MM_err_Nv présente une valeur de 2.95%. Cela donne au SIFT modifié une performance de 97% de reconnaissance des segments non voisés.

L'évaluation du paramètre MM_err_Hi présente un maximum de 3.45%, ce qui implique que le SIFT modifié est capable de détecter les segments haute-fréquence avec une précision de 96%.

Concernant la capacité du SIFT modifié à estimer les segments phoniques, nous remarquons que le paramètre MM_err_ph est de 22%; le SIFT modifié est donc moins fiable dans ce cas.

3.3.5 Analyse des résultats du logiciel Praat

La dernière partie d'analyse consiste à comparer la performance du logiciel Praat à identifier les segments d'un signal traité par rapport aux algorithmes : AMDF, HPS, ACF et SIFT modifié.

Un calcul identique à celui des cas précédents sera réalisé. Les résultats obtenus sont présentés dans la Figure 3.16.

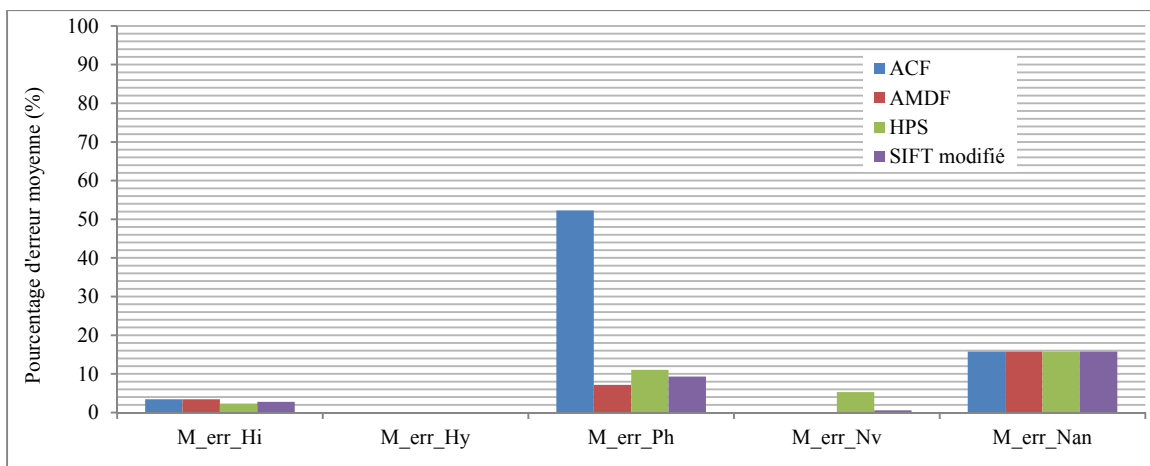


Figure 3-16 Pourcentages de la moyenne des différentes erreurs produites par rapport au Praat

D'après la Figure 3.16, nous remarquons que le paramètre d'M_err_Hy est nul. Ainsi, nous considérons que le Praat a réussi à isoler les segments hyperphoniques.

Le Praat présente un taux d'M_err_Nv inférieur à 1% par rapport aux algorithmes AMDF, ACF et SIFT modifié et une valeur de 5.3% par rapport à l'HPS.

Le paramètre M_err_Ph dans ce cas présente deux pics, un maximum de 52% avec l'ACF et un minimum de 11% avec le SIFT modifié, l'HPS et l'AMDF.

De plus, le Praat possède une capacité d'évaluation des segments haute-fréquence variant entre 96% et 98%.

Le paramètre M_err_Nan atteint un taux de 15.8% avec les algorithmes : HPS, AMDF, ACF et SIFT modifié. C'est le seul cas où ce paramètre est non nul. En effet, le Praat ne peut pas identifier toutes les parties d'un signal de cri : il y a des zones non reconnues par le Praat.

La Figure 3.17 présente les résultats d'analyses secondaires obtenus dans le but de valider l'analyse primaire.

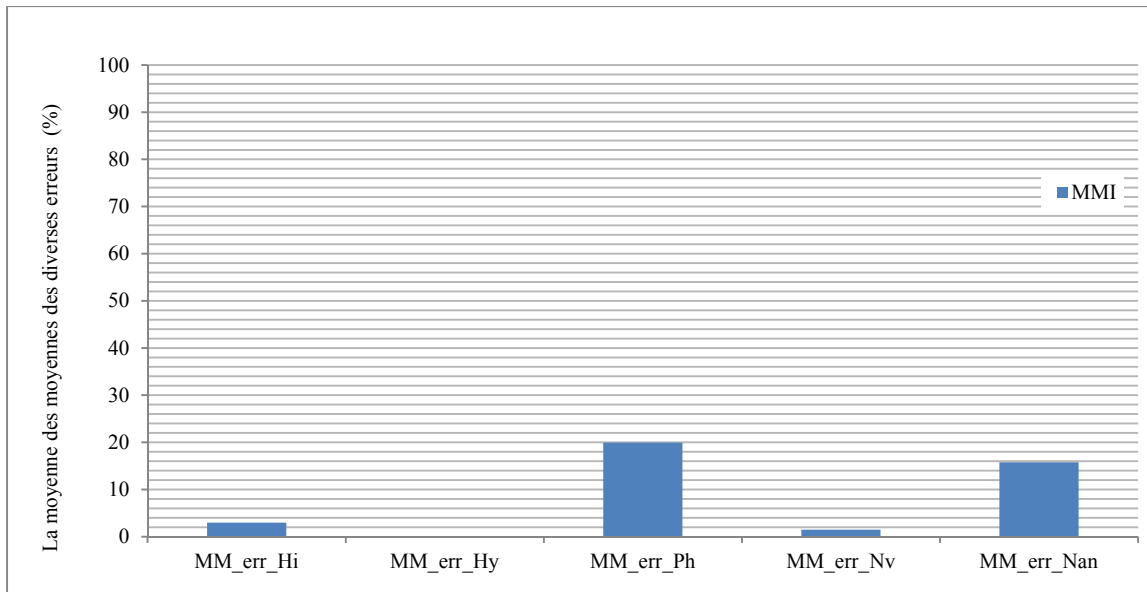


Figure 3-17 L'erreur moyenne des quatre méthodes par type d'erreur

Cette figure confirme que le Praat est capable d'estimer sans erreur les segments hyperphoniques avec un taux du paramètre MM_err_Hy nul.

Le Praat possède un taux du paramètre MM_err_Nv de 1.5% ce qui implique que sa capacité d'estimation des segments non voisés est de 98%.

Le taux 3% du paramètre MM_err_Hi a permis au Praat d'atteindre 97% pour identifier les segments haute-fréquence. Par ailleurs, la valeur du taux d'MM_err_Ph est de 20% ce qui implique que le Praat est moins fiable à isoler les segments phoniques.

Nous constatons que le taux d'MM_err_Nan est de 15.8%, confirmant la faible capacité du logiciel Praat à définir toutes les zones du signal analysé.

3.4 Conclusion

Tout au long de ce mémoire, nous avons présenté une analyse quantitative relative aux performances des différentes méthodes d'estimation de la fréquence fondamentale du signal du cri chez les nouveau-nés. D'après les résultats des simulations, nous pouvons conclure les résultats suivants.

Le logiciel Praat a donné de bons résultats d'identification des fréquences fondamentales : hyperphonique, de haute-fréquence et non voisée. Mais il a donné de moins bons résultats dans l'évaluation de la fréquence fondamentale phonique. De plus, le logiciel Praat ne peut pas être retenu comme outil d'évaluation fiable en raison de sa faiblesse à évaluer tous les segments du signal traité. Nous avons quantifié cette faiblesse par le paramètre M_err_Nan qui a donné un taux d'erreur élevé.

L'algorithme HPS, quant à lui, a montré une bonne précision d'estimation de la fréquence fondamentale phonique. Mais il a donné un résultat moyen pour la détection des fréquences hyperphoniques et de haute-fréquence. De plus, il a présenté un taux d'erreur d'évaluation un peu élevé pour les fréquences non voisées.

L'algorithme SIFT modifié a présenté une bonne précision pour l'estimation des fréquences fondamentales : hyperphonique, haute-fréquence et non voisée. Cependant nous avons observé qu'il est moins précis dans l'estimation de la fréquence fondamentale phonique.

Nous avons aussi relevé un comportement similaire des algorithmes SIFT modifié et AMDF en ce qui concerne l'estimation de la fréquence fondamentale d'un signal de cri.

Nous avons constaté que l'algorithme ACF est fiable dans la détection des hautes fréquences et des fréquences non voisées. Il a également une précision moyenne pour identifier la fréquence phonique. Il a échoué aussi à isoler la fréquence hyperphonique (constat d'un taux d'erreur très élevé).

CHAPITRE 4

DÉVELOPPEMENT ET MISE EN ŒUVRE D'UN ALGORITHME HYBRIDE CAMDA POUR L'ESTIMATION DE LA FRÉQUENCE FONDAMENTALE

4.1 Introduction

L'étude comparative réalisée dans le troisième chapitre nous a permis de connaître les avantages et les inconvénients des algorithmes d'extraction de la fréquence fondamentale appliqués sur le signal de cri des nourrissons : AMDF, ACF, HPS et le SIFT modifié ainsi que le logiciel Praat.

Nous avons remarqué que l'AMDF et le SIFT modifié sont les algorithmes les plus fiables pour détecter les segments : de haute-fréquence, hyperphoniques et non-voisés. Par contre, ils sont moins fiables à identifier les segments phoniques. L'idée principale est de minimiser l'erreur de la surestimation de la fréquence fondamentale de l'AMDF dans la zone des segments phoniques. Le SIFT modifié est un dérivé de l'ACF. Ce dernier fournit des informations complémentaires pour l'AMDF (Abdullah-Al-Mamun, Sarker et Muhammad, 2009). Cela nous a amenés à développer un nouvel algorithme permettant une combinaison entre le SIFT modifié et l'AMDF pour corriger l'erreur produite lors de la détection de la fréquence fondamentale des segments phoniques.

La réalisation du nouvel algorithme hybride CAMDA combine l'algorithme SIFT modifié et l'algorithme AMDF. Ces algorithmes nécessitent des modifications avant leurs intégrations. Ces changements sont décrits ci-dessous.

4.1.1 Description de l'algorithme hybride CAMDA

Le problème de doublement de la fréquence fondamentale de l'algorithme AMDF a été déjà corrigé sur le signal de la parole en multipliant l'AMDF aligné et inversé par l'algorithme ACF (Abdullah-Al-Mamun, Sarker et Muhammad, 2009). En effet, le choix de l'algorithme ACF est réalisé a priori puisqu'il contient des informations complémentaires à l'algorithme AMDF.

Nous allons intégrer l'algorithme AMDF dans l'algorithme SIFT modifié afin d'améliorer la précision sur l'extraction de la fréquence fondamentale. Le problème de dédoublement de la fréquence fondamentale de l'AMDF est corrigé par la haute résolution de détection de la fréquence fondamentale à l'aide de filtres passe-bandes.

4.1.2 Principe de fonctionnement de l'algorithme CAMDA

Le principe de fonctionnement de l'algorithme CAMDA que nous avons réalisé est représenté dans le schéma bloc suivant :

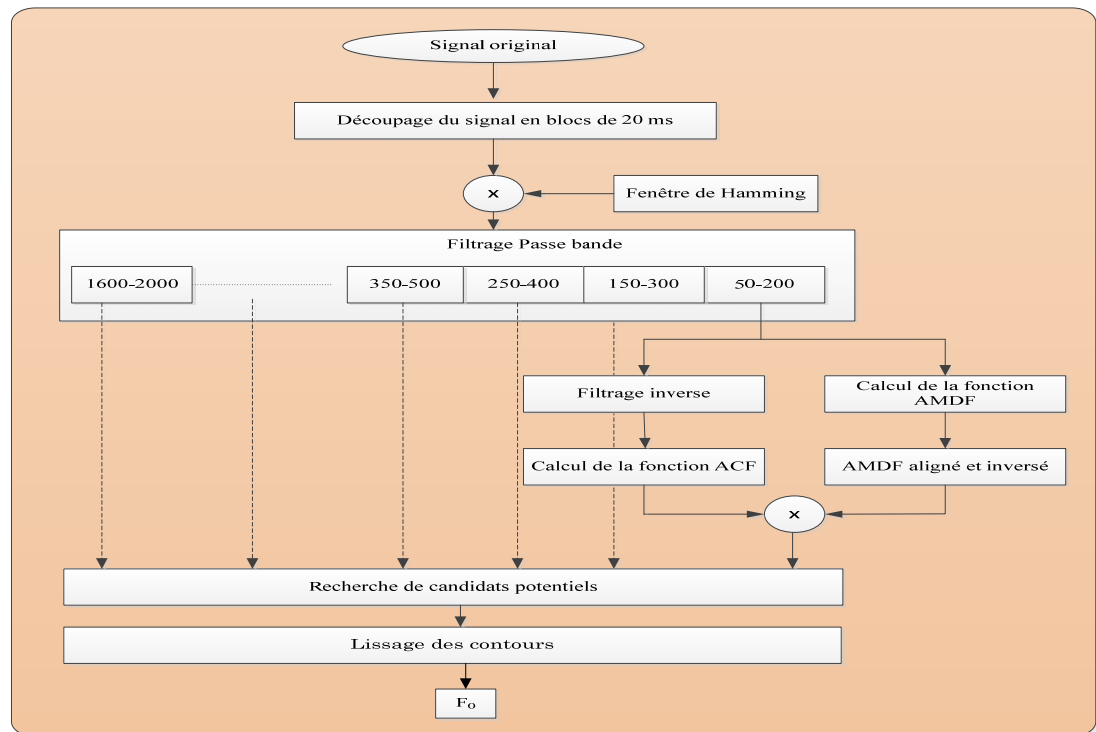


Figure 4-1 Principe de fonctionnement de l'algorithme CAMDA

1^{ère} étape : le signal original a été découpé sous forme de segments de 20ms.

2^{ème} étape : chaque segment de 20ms a été multiplié par la fenêtre de Hamming.

3^{ème} étape : cette étape consiste à réaliser un filtre passe-bandes. Ce filtre découpe le spectre de fréquence du signal de cri en onze bandes passantes. Les quatre premières bandes varient de 50 à 200Hz, de 150 à 300 Hz, de 250 à 400Hz et de 350- à 500Hz (Abdullah-Al-Mamun, Sarker et Muhammad, 2009). Les autres sept bandes de fréquence sont choisies avec une largeur de bande de 400 Hz et varient de 400 à 800 Hz, de 600 à 1000 Hz, de 800 à 1200 Hz, de 1000 à 1400 Hz, de 1200 à 1600 Hz, de 1400 à 1800 Hz, et de 1600 à 2000 Hz. En effet, la largeur des bandes de 150 Hz est choisie dans le spectre de fréquence de 50 Hz à 500 Hz pour avoir la haute résolution d'estimation de la fréquence fondamentale dans les segments phoniques.

La résolution à haute-fréquence est faible dans le spectre de fréquence variant de 600 à 2000 Hz. De cette façon, le nombre de filtres passe-bandes utilisé est faible et la durée de calcul du CPU diminue considérablement en conséquence.

4^{ème} étape : la fonction AMDF alignée et inversée est appliquée à la sortie de chaque filtre passe-bandes. En parallèle, l'ACF est appliquée sur la même bande de fréquences après la fonction de filtre inverse. Enfin, la multiplication de l'AMDF par l'ACF donne le spectre de fréquences avec différents maximums ($m1, \dots, m11$).

5^{ème} étape : pour chaque segment de 20ms, onze candidats sont détectés de $m1$ à $m11$. Le bloc de recherche de candidats potentiels sert à chercher la fréquence fondamentale optimale avec l'indice minimal parmi les fréquences $f1, \dots, f11$.

Dans le cas où l'on trouve plusieurs fréquences de même position d'indice minimum, on compare les amplitudes de ces minimums afin de choisir la plus élevée. Celle-ci correspond à la fréquence fondamentale du segment.

6^{ème} étape : cette étape est représentée par le bloc de lissage qui permet de réduire les changements brusques dans le contour de la fréquence fondamentale.

4.1.3 Discussion des résultats

La simulation de l'algorithme CAMDA sur Matlab, conformément au schéma bloc de la Figure 3.18 permet d'estimer la fréquence fondamentale, sachant que l'on se base sur les mêmes données que celles utilisées dans l'étude comparative réalisée précédemment.

Similairement, nous avons appliqué la même procédure de calcul des erreurs : $Er_{Hi_pc_i}$, $Er_{Ph_pc_i}$, $Er_{Hy_pc_i}$, $Er_{Nan_pc_i}$ et $Er_{Nv_pc_i}$ que celle décrite dans le chapitre utilisant les formules de calcul 3.1 à 3.5 afin d'évaluer la capacité de l'algorithme CAMDA à estimer la fréquence fondamentale.

Lors de cette étude comparative, nous avons fixé CAMDA comme étant le modèle de référence et nous avons testé les performances des algorithmes suivants : AMDF, ACF, HPS, SIFT modifié ainsi que le logiciel Praat.

Le calcul de la moyenne des erreurs de classification des segments a été effectué selon les deux formes : calculs primaire et secondaire. Les résultats obtenus par le calcul primaire sont présentés dans la Figure 4.2.

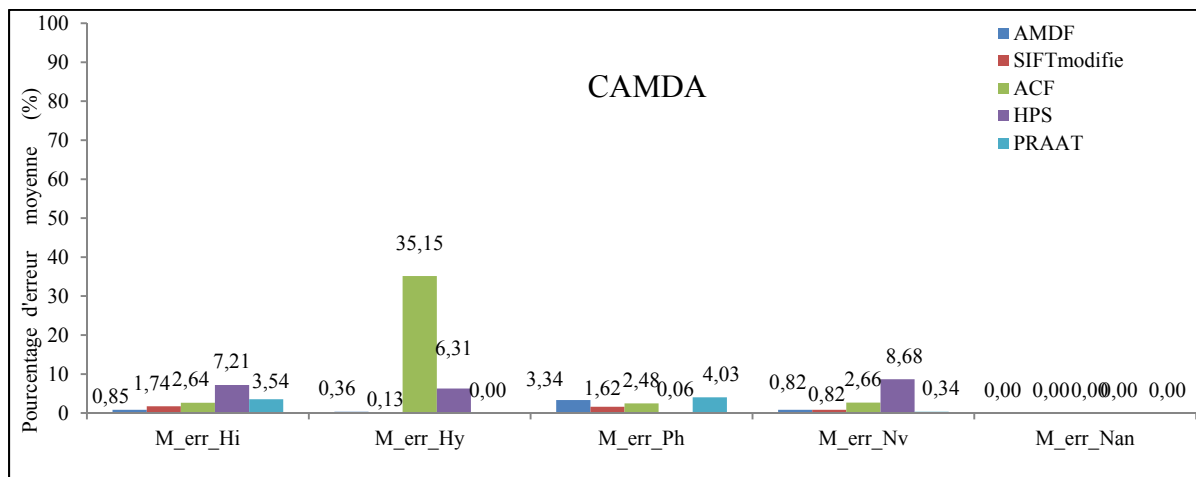


Figure 4-2 Pourcentages de la moyenne des différentes erreurs

Nous observons dans cette figure que l'algorithme CAMDA présente un taux d'erreur de classification des segments de haute-fréquence qui varie entre 0,85% et 7,21% par rapport aux algorithmes : AMDF, ACF, HPS, le SIFT modifié ainsi qu'au logiciel Praat. Ce résultat nous a permis de considérer que le CAMDA est l'une des techniques les plus fiables lors de la détection des segments de haute-fréquence.

Nous remarquons par ailleurs que le taux d'erreurs de classification des segments phoniques du CAMDA est faible comparativement aux cinq méthodes, en particulier avec l'HPS. Ceci confirme la fiabilité du CAMDA à détecter les segments phoniques.

Le taux d'erreurs de classification des segments non-voisés du CAMDA est comparable à celui de l'HPS. Par contre, le taux est meilleur qu'avec l'AMDF, l'ACF, le SIFT modifié et le Praat. Ce résultat nous permet de classer le CAMDA comme étant un algorithme fiable lors de l'estimation des segments non-voisés.

Concernant le calcul des erreurs de classification des segments hyperphoniques, nous remarquons que le taux d'erreurs présenté est mauvais par rapport à l'ACF. Étant donné que l'ACF n'est pas très efficace pour détecter les segments hyperphoniques, on se permet de négliger cette valeur. Par contre, nous avons obtenu un meilleur taux d'erreurs hyperphoniques qu'avec l'AMDF, l'HPS, le SIFT modifié et le Praat.

Le CAMDA fournit une fiabilité correcte à détecter tous les segments d'un signal traité avec un taux d'erreur non définie nul.

Nous avons refait le calcul secondaire conformément aux équations 3.6 à 3.10 pour confirmer les résultats obtenus par le calcul primaire. Les résultats obtenus sont présentés dans la Figure 4.3.

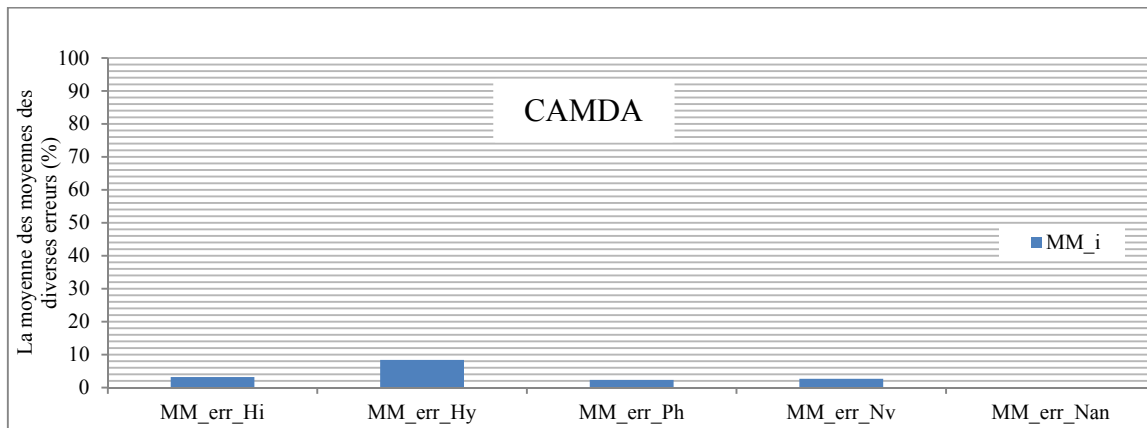


Figure 4-3 L'erreur moyenne des quatre méthodes par type d'erreurs

Selon la Figure 4.3, nous pouvons dire que l'algorithme CAMDA présente une fiabilité correcte à estimer les segments : de haute-fréquence, phoniques, non-voisés et hyperphoniques, ainsi qu'un meilleur taux d'erreur MM_err_Nan.

4.2 Conclusion

Les résultats obtenus par l'algorithme CAMDA nous ont permis de montrer que ce dernier est capable de remédier aux problèmes posés par l'AMDF et le SIFT modifié lors de la détection des segments phoniques.

De plus, nous remarquons que le CAMDA produit de bons résultats de détection des segments : hyperphoniques, non-voisés et de haute-fréquence.

CONCLUSION

La fréquence fondamentale est l'une des plus importantes caractéristiques utilisées pour distinguer les différents types de cris des bébés. Ses modèles atypiques révèlent des problèmes dans le système nerveux central. Ainsi, des mesures précises de cette fréquence et de ses variations dans le temps sont importantes pour obtenir des informations fiables sur l'état de santé des nouveaux-nés.

Dans ce mémoire, nous nous sommes intéressés à l'analyse comparative de cinq techniques d'estimation de la fréquence fondamentale d'un signal cri chez les nouveau-nés.

Pour réaliser ce travail, nous avons implémenté sur Matlab les algorithmes AMDF, ACF, HPS. Nous avons également utilisé le logiciel Praat et l'algorithme SIFT modifié. Nous avons comparé les résultats obtenus de chacune de ces techniques en terme de fréquence fondamentale par rapport aux autres méthodes afin d'identifier l'outil le plus fiable.

Nous avons également noté que l'algorithme ACF est plus fiable dans la détection des hautes fréquences et fréquences non voisées. Sa capacité à estimer la fréquence phonique est moyenne. Il a cependant échoué à estimer la fréquence hyperphonique avec un taux d'erreur très élevé. Par ailleurs, nous avons constaté dans notre analyse que le logiciel Praat ne permet pas d'identifier tous les segments du signal traité, d'où le faible intérêt de son utilisation dans notre projet.

Nous avons constaté que les algorithmes SIFT modifié et AMDF ont présenté une bonne précision pour l'estimation des fréquences fondamentales : hyperphonique, haute-fréquence et non voisée. En revanche, nous avons observé qu'ils sont moins précis dans l'identification de la fréquence fondamentale phonique.

D'autre part, nous avons remarqué que l'algorithme HPS a montré une bonne précision quant à l'évaluation de la fréquence fondamentale phonique.

Mais il a donné un résultat moyen lors de la détection des fréquences hyperphonique et haute-fréquence. De plus, il a présenté un taux d'erreur d'évaluation un peu élevé pour la fréquence non voisée.

L'algorithme proposé CAMDA a permis de détecter les différents modes de segments d'un signal traité avec un taux d'erreurs minimal. Grâce aux résultats obtenus, cet algorithme pourrait être utilisé dans des projets nécessitant l'estimation de la fréquence fondamentale d'un signal de cri avec une haute précision.

L'analyse comparative des résultats obtenus nous a permis de classer les algorithmes étudiés selon la capacité de chacun à estimer la fréquence fondamentale : hyperphonique, haute-fréquence, phonique et non voisée.

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

- Abdullah-Al-Mamun, K, F Sarker et G Muhammad. 2009. « A high resolution pitch detection algorithm based on AMDF and ACF ». *Journal of Scientific Research*, vol. 1, n° 3, p. 508-515.
- Boite, R., H. Bourlard, T. Dutoit, J. Hancq et et H. Leich. 2000. *Traitement de la Parole*. Polytechniques Universitaires Romandes.
- Cormier, Gabriel. 2013. « Analyse des signaux.Programme du premier cycle ». [Notes du cours :GELE2511], Université de Moncton 18 p.
- De La Cuadra, Patricio, Aaron Master et Craig Sapp. 2001. « Efficient pitch detection techniques for interactive music ». In *Proceedings of the 2001 International Computer Music Conference*. p. 403-406.
- Fort, Ada, et Claudia Manfredi. 1998. « Acoustic analysis of newborn infant cry signals ». *Medical Engineering & Physics*, vol. 20, n° 6, p. 432-442.
- Gargour, Christian S., et Venkat Ramachandran. 2006. *Traitement numérique des signaux*, 2nd ed. Montréal: Université du Québec, École de technologie supérieure, 392 p.
- Hui, Ding, Qian Bo, Li Yanping et Tang Zhenmin. 2006. « A Method Combining LPC-Based Cepstrum and Harmonic Product Spectrum for Pitch Detection ». In *Intelligent Information Hiding and Multimedia Signal Processing, 2006. IHH-MSP '06. International Conference on*. (Dec. 2006), p. 537-540.
- Kheddache, Yasmina, et Chakib Tadj. 2013a. « Characterization of Pathologic Cries of Newborns Based On Fundamental Frequency Estimation ». *International Conference on Bioinformatics and Biomedical Engineering*, p. 2.
- Kheddache, Yasmina, et Chakib Tadj. 2013b. « Frequential Characterization of Healthy and Pathologic Newborns Cries ». *American Journal Of Biomédical Engineering*, vol. 3, n° 6, p. 3.
- Lederman, Dror. 2002 «Automatic Classification Of Infants Cry ». Mémoire de maîtrise en génie électrique, . Ben-Gurion University of The Negev Faculty of Engeneering Sciences, 121 p.
- Lederman, Dror. 2010. « Estimation of Infants' Cry Fundamental Frequency using a Modified SIFT algorithm ». *arXiv preprint arXiv:1009.2796*.

- Manfredi, C., M. D'Aniello, P. Brusaglioni et A. Ismaelli. 2000. « A comparative analysis of fundamental frequency estimation methods with application to pathological voices ». *Medical Engineering & Physics*, vol. 22, n° 2, p. 135-147.
- Manfredi, Claudia, V. Tocchioni et L. Bocchi. 2006. « A Robust Tool for Newborn Infant Cry Analysis ». In *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*. (Aug. 30 2006-Sept. 3 2006), p. 509-512.
- MARC, DELAHAIE. 2009. *L'évolution du langage de l'enfant. De la difficulté au trouble*. Inpes, 84 p.
- Messaoud, Ali, et Chakib Tadj. 2011. « Analysis Of Acoustic Features Of Infant Cry For Classification Purposes ». In *Canadian Conference on Electrical and Computer Engineering*. (Niagaras Falls, Canada). Vol. 4.
- Preux, philippe. 2011. *Fouille de données*. Université de Lille 3, 256 p.
- Rabiner, L., M. Cheng, A. E. Rosenberg et C. McGonegal. 1976. « A comparative performance study of several pitch detection algorithms ». *Acoustics, Speech and Signal Processing, IEEE Transactions on* vol. 24, n° 5, p. 399-418.
- Truby, H. M., et J. Lind. 1965. « CRY SOUNDS OF THE NEWBORN INFANT^{1,2} ». *Acta Pædiatrica*, vol. 54, p. 8-59.
- Upadhyay, Savitha S, et Nilashree Wankhede. 2012. « Pitch Estimation using Autocorrelation Method and AMDF ». *International Conference on Advances in Computing and Management*
- Valin, J. M., et R. Lefebvre. 2000. « Bandwidth extension of narrowband speech for low bit-rate wideband coding ». In *Speech Coding, 2000. Proceedings. 2000 IEEE Workshop on*. (2000), p. 130-132.
- Valin, Jean-Marc. 2001. « Bandwidth extension of narrowband speech ». University of Sherbrooke, University of Sherbrooke, 65 p.
- Várallyay Jr, G, A Illényi et Z Benyó. 2008. « The automatic segmentation of the infant cry ». In *Proc. BUDAMED'08 Conference, Budapest*. p. 28-32.
- Várallyay Jr, G, András Illényi et Zoltán Benyó. 2009. « Automatic Infant Cry Detection ». In *Proc. of the 6th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 2009)*.

- Yasmina Kheddache, Chakib Tadj. 2013. « Acoustic measures of the cry characteristics of healthy newborns and newborns with pathologies ». *J. Biomedical Science and Engineering*, 2013, 6, 796-804.
- Ykhlef, F, R Amiar, S Hecini, W Benzaba et L Bendaouia. 2007. « Etude Comparative des Performances de Plusieurs Techniques de Détection de la Fréquence Fondamentale des Signaux Vocaux ».
- Yu-Min, Zeng, Zhen-Yang Wu, Hai-Bin Liu et Lin Zhou. 2003. « Modified AMDF pitch detection algorithm ». *Proceedings of the Second International Conference on Machine Learning and Cybernetics*,, vol. 1, p. 470-473 Vol.1.