

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

THÈSE PAR ARTICLES PRÉSENTÉE À
L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

COMME EXIGENCE PARTIELLE
À L'OBTENTION DU
DOCTORAT EN GÉNIE
Ph. D.

PAR
Yasmina KHEDDACHE

CARACTÉRISATION DES CRIS DES NOURRISSONS EN VUE DU DIAGNOSTIC
PRÉCOCE DE DIFFÉRENTES PATHOLOGIES

MONTRÉAL, LE 21 MAI 2014



Yasmina Kheddache, 2014



Cette licence [Creative Commons](https://creativecommons.org/licenses/by-nc-nd/4.0/) signifie qu'il est permis de diffuser, d'imprimer ou de sauvegarder sur un autre support une partie ou la totalité de cette œuvre à condition de mentionner l'auteur, que ces utilisations soient faites à des fins non commerciales et que le contenu de l'œuvre n'ait pas été modifié.

PRÉSENTATION DU JURY

CETTE THÈSE A ÉTÉ ÉVALUÉE

PAR UN JURY COMPOSÉ DE :

M. Chkib Tadj, directeur de thèse
Département de génie électrique à l'École de technologie supérieure

M. Chamseddine Talhi, président du jury
Département de génie logiciel et des TI à l'École de technologie supérieure

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

M. Roch Lefebvre, membre externe indépendant
Faculté de génie à l'Université de Sherbrooke

ELLE A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 12 MAI 2014

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Je tiens tout d'abord à exprimer ma gratitude à mon directeur de thèse Monsieur Chakib Tadj qui a suivi et encadré ce travail avec intérêt, disponibilité et compétence.

J'adresse mes vifs remerciements aux membres de jury pour l'attention qu'ils ont porté à mon travail et pour leurs précieuses remarques grâce auxquelles j'ai pu améliorer cette thèse. Je remercie M. Chamseddine Talhi qui assure la présidence, et les professeurs M. Christian Gargour et M. Roch Lefebvre qui ont accepté d'examiner ce travail.

Je tiens particulièrement à remercier ma famille; mon mari Brahim et mes filles, Hyba, Maissane et Lyna pour leur soutien inconditionnel. Je tiens aussi à exprimer toute ma gratitude à mes parents pour leurs encouragements tout au long de mon cursus. Je remercie également mes frères Adlene et Islame ainsi que ma sœur Nahla.

Je remercie aussi mes amis, Nacera et Amel pour leur soutien moral ainsi que toute l'équipe du laboratoire pour leur esprit de famille. Je les remercie pour tous les bons moments passés ensemble.

CARACTÉRISATION DES CRIS DES NOURRISSONS EN VUE DU DIAGNOSTIC PRÉCOCE DE DIFFÉRENTES PATHOLOGIES

Yasmina KHEDDACHE

RÉSUMÉ

L'utilisation des signaux de cris dans le diagnostic se base sur les théories qui ont été proposées par les différents chercheurs dans le domaine. Le principal objectif de leurs travaux était l'analyse spectrographique ainsi que la modélisation des signaux de cris. Ils ont démontré que les caractéristiques acoustiques des cris des nouveau-nés sont liées à des conditions médicales particulières.

Cette thèse est destinée à contribuer à l'amélioration de la précision de la reconnaissance des cris pathologiques par la combinaison de plusieurs paramètres acoustiques issus de l'analyse spectrographique et des paramètres qui qualifient les cordes et le conduit vocal. Car les caractéristiques acoustiques représentant le conduit vocal ont été largement utilisées pour la classification des cris, alors que les caractéristiques des cordes vocales pour la reconnaissance automatique des cris, ainsi que leurs techniques efficaces d'extraction n'ont pas été exploitées.

Pour répondre à cet objectif, nous avons procédé en premier lieu à une caractérisation qualitative des cris des nouveau-nés sains et malades en utilisant les caractéristiques qui ont été définies dans la littérature et qui qualifient le comportement des cordes et du conduit vocal pendant le cri. Cette étape nous a permis d'identifier les caractéristiques les plus importantes dans la différenciation des cris pathologiques étudiés.

Pour l'extraction des caractéristiques sélectionnées, nous avons implémenté des méthodes de mesures efficaces permettant de dépasser la surestimation et la sous-estimation des caractéristiques. L'approche de quantification proposée et utilisée dans ce travail facilite l'analyse automatique des cris et permet une utilisation efficace de ces caractéristiques dans le système de diagnostic. Nous avons procédé aussi à des tests expérimentaux pour la validation de toutes les approches introduites dans cette thèse. Les résultats sont satisfaisants et montrent une amélioration dans la reconnaissance des cris par pathologie.

Les travaux réalisés sont présentés dans cette thèse sous forme de trois articles publiés dans différents journaux. Deux autres articles publiés dans des comptes rendus de conférences avec comité de lecture sont présentés en annexes.

Mots clés : La neurophysiologie des cris, caractérisation acoustique des cris, classification des cris, diagnostic précoce.

CHARACTERIZATION OF NEWBORN'S CRIES FOR THE EARLY DIAGNOSIS OF VARIOUS DISEASES

Yasmina KHEDDACHE

ABSTRACT

The use of newborn cry signals in diagnosis is based on many theories proposed lately. The main objective in these researches is the cry signals modeling and spectrographic analysis. It has been shown that the newborn cry acoustics are linked to particular medical conditions.

This thesis is motivated by improvement of the accuracy of pathological cries recognition. This can be performed by the combination of several acoustic parameters from spectrographic analysis and parameters that describe the configuration of vocal tract and vocal folds. Acoustic characteristics representing the vocal tract were widely applied to the classification of the cries. However the usefulness of vocal folds characteristics in the automatic recognition, as well as their effective techniques extraction have not been exploited deeply.

In this context, we have performed a qualitative characterization of healthy and pathologic newborns cries using characteristics that have been defined in the literature and which describe vocal tract and vocal folds behavior during the cry. This step allowed us to identify the most relevant features in the differentiation of the studied pathological cries.

For the extraction of selected characteristics, we have implemented effective measure methods that avoid the overestimation and underestimation of characteristics. The proposed and used approach for characteristics quantification facilitates the automatic analysis of cries and allows efficient use of these features in the diagnostic system. We also conducted experimental tests for the validation of all proposed approaches in this thesis. The results are suitable and show an improvement of the cry-based pathology recognition.

The work presented in this thesis is a collection of three articles published/submitted in various journals. Two other papers published in conferences are presented in the annexes.

Keywords: neurophysiology of the cries, acoustic characterization of the cries, cries classification, early diagnosis.

TABLE DES MATIÈRES

	Page
INTRODUCTION.....	1
0.1 Cadre de recherche.....	1
0.2 Problématique et justification	3
0.3 Objectifs.....	5
0.4 Méthodologie	5
0.4.1 Étude de l'état de l'art.....	6
0.4.2 Enregistrement des cris	7
0.4.3 Construction de la base de données	8
0.4.4 Étude spectrographique de cris	8
0.4.5 Caractérisation qualitative	9
0.4.6 Sélection des caractéristiques	9
0.4.7 Estimation automatique des caractéristiques	9
0.4.8 Quantification des caractéristiques	10
0.4.9 Validation expérimentale.....	10
0.4.10 Caractérisation des pathologies étudiées	11
0.5 Organisation de la thèse	11
CHAPITRE 1 ETAT DE L'ART	15
1.1 Introduction	15
1.2 Le modèle physioacoustique du cri	15
1.4 Les caractéristiques du cri	19
1.5 Les caractéristiques des cris dans différents cas cliniques	23
1.6 Travaux antérieurs	26
1.6.1 L'analyse acoustique des cris.....	26
1.6.2 La perception des cris	29
1.6.3 La classification des cris	30
1.7 Modèle électrique des cris	32
1.8 Prétraitement des signaux de cris	34
1.9 Analyse par prédiction linéaire.....	36
1.10 Analyse cepstrale.....	39
1.10.1 Coefficients cepstraux.....	39
1.10.2 Coefficients MFCC.....	40
1.11 Analyse par ondelette	42
1.11.1 Transformée en ondelette discrète (DWT)	42
1.11.2 Transformée en paquet d'ondelettes	43
1.12 Analyse en composante principale	45
1.13 Techniques d'extraction de la fréquence fondamentale.....	46
1.13.1 Méthode des coefficients cepstraux	47
1.13.2 Méthode SSM	48
1.13.3 La méthode AMDF	50

1.14	Techniques de classification des cris.....	51
1.14.1	Hidden Markov Model (HMM).....	52
1.14.2	Gaussian Mixture Model (GMM).....	53
1.15	Conclusion.....	54
CHAPITRE 2 FREQUENTIAL CHARACTERIZATION OF HEALTHY AND PATHOLOGIC NEWBORNS CRIES.....		57
2.1	Introduction.....	58
2.2	Acoustic definition of cries.....	61
2.3	The characteristics of cries under certain medical conditions.....	61
2.3.1	The cries of healthy newborns.....	61
2.3.2	The cries of premature newborns.....	63
2.3.3	Cries of newborns with various pathologies.....	63
2.4	Methodology for qualitative and quantitative characterization.....	64
2.4.1	Cry database.....	64
2.4.2	Qualitative characterization of cries.....	67
2.4.3	Quantitative characterization of cries.....	69
2.4.4	Fundamental frequency estimation.....	71
2.5	Results analysis.....	72
2.6	Experimental validation.....	79
2.6.1	Features Extraction.....	80
2.6.2	Experimental results.....	81
2.7	Conclusions.....	83
CHAPITRE 3 ACOUSTIC MEASURES OF THE CRY CHARACTERISTICS OF HEALTHY NEWBORNS AND NEWBORNS WITH PATHOLOGIES.....		85
3.1	Introduction.....	86
3.2	Dysphonic cry.....	88
3.3	Cry database.....	89
3.4	Methodology.....	92
3.5	Dysphonation estimation in newborn cries.....	94
3.5.1	Normalized noise energy estimation (NNE).....	96
3.5.2	ACF-based noise estimator.....	97
3.5.3	MMSE-based noise estimator.....	99
3.6	Results and analysis.....	101
3.7	Conclusions.....	108
CHAPITRE 4 RESONANCE FREQUENCIES BEHAVIOR IN THE CRIES OF HEALTHY NEWBORNS AND NEWBORNS WITH PATHOLOGIES.....		109
4.1	Introduction.....	110
4.2	Physioacoustic Model and Neurological Basis of the Newborn Cry.....	113
4.3	Newborn Cry Characteristics in a Medical Environment.....	114
4.4	Method.....	116
4.4.1	Recording sessions.....	116

4.4.2	Procedure	117
4.4.3	Resonance frequencies estimation	121
4.4.4	Fundamental frequency estimation	123
4.5	Results and Analysis	124
4.5.1	F0, RF1, and RF2 estimation results.....	126
4.5.2	RF1:RF2 ratio and RF2 slope evaluation results	126
4.5.3	TUP and TRP evaluation results for RF1 and RF2.....	130
4.6	Discussion	133
4.7	Conclusion	137
CHAPITRE 5	SYNTHÈSE ET CONCLUSION.....	139
5.1	Synthèse des articles.....	139
5.2	Contribution.....	143
5.3	Recommandations	144
5.4	Conclusion	146
ANNEXE I	NEWBORN'S PATHOLOGICAL CRY IDENTIFICATION SYSTEM	147
ANNEXE II	CHARACTERISATION OF PATHOLOGIC CRIES OF NEWBORNS BASED ON FUNDAMENTAL FREQUENCY ESTIMATION	163
BIBLIOGRAPHIE		175

LISTE DES TABLEAUX

	Page
Tableau 1.1 Récapitulatif des caractéristiques des cris qui ont été associées à certains états de santé	24
Tableau 2.1 Cry characteristics.....	62
Tableau 2.2 The pathologies studied	66
Tableau 2.3 Characteristics of cries examined by spectrographic analysis	68
Tableau 2.4 Measured characteristics of cries	71
Tableau 2.5 Analysis of variance (ANOVA) according to gestational	79
Tableau 2.6 Results of first experiment	82
Tableau 2.7 Results of second experiment	82
Tableau 4.1 Pathologies Studied.....	117
Tableau 4.2 Cry Characteristics Measured	118
Tableau 4.3 Measured characteristics of cries	129
Tableau 4.4 Measured Characteristics of Cries	132
Tableau 4.5 Characteristics Associated with the Pathologies Studied	136
Tableau 4.6 Analysis of Variance (ANOVA) based on Gestational Age.....	137

LISTE DES FIGURES

	Page
Figure 0.1	Méthodologie.....6
Figure 0.2	Les différents niveaux d'information qui concernent les cris des bébés enregistrés8
Figure 1.1	Section du conduit vocal des adultes et des nouveau-nés16
Figure 1.2	Anatomie du cri17
Figure 1.3	Diagramme d'un spectrogramme du cri23
Figure 1.4	Le modèle de production de la parole33
Figure 1.5	Les étapes d'extractions des coefficients MFCC41
Figure 1.6	Exemple d'ondelette de Daubechies43
Figure 1.7	Décomposition discrète par ondelettes orthogonales (DWT)43
Figure 1.8	Décomposition par paquets d'ondelettes44
Figure 1.9	Principales étapes de l'application de l'ACP46
Figure 1.10	Séries Cepstraux de quelques segments de cris typiques48
Figure 1.11	Le spectre original et sa version lissée49
Figure 1.12	Reconnaissance automatique des cris des bébés51
Figure 2.1	Newborn cries characterization64
Figure 2.2	Waveform and spectrogram of healthy and pathologic cries69
Figure 2.3	Estimation of F0 using the modified SIFT algorithm73
Figure 2.4	The average percentage of hyperphonic segments and irregularity of F0 by a) Pathologic and healthy cries, b) Category of pathologies, c) Pathologies and gestational age75
Figure 2.5	The mean percentage and standard deviation of a) hyperphonic segments b) irregularity of F0 by category of pathologies76

Figure 2.6	The mean percentage and standard deviation of a) hyperphonic segments b) irregularity of F0 by pathologies.....	78
Figure 2.7	Experimental validation.....	80
Figure 3.1	Spectrograms of different cry modes a) phonation, b) hyperphonation, c) dysphonation.....	91
Figure 3.2	Dysphonation estimation in newborn cries	96
Figure 3.3	The average percentage of dysphonic segments by pathology and by measurement method.....	103
Figure 3.4	The average percentage of dysphonic and no-voice segments and the percentage of mode changes by pathology	103
Figure 3.5	Distribution of cry modes (dysphonia, hyperphonation, phonation) by pathology	106
Figure 3.6	The mean percentage and standard deviation of a) dysphonic segments; b) no-voice and c) mode change by pathologies	107
Figure 4.1	Methodology. (“Signal”; “Filtering”, “RF1, RF2”, “Slope”, “Ratio”) .	118
Figure 4.2	Fundamental frequencies and their harmonics (red lines), resonance frequencies: RF1 (black dots) and RF2 (blue dots), tuning between the RFs and F0 harmonics (green rectangles).....	125
Figure 4.3	Box-and-whiskers plots for the averages of: a) F0, b) RF1, c) RF2, d) the RF1:RF2 ratio, e) the RF2 slope, by pathology and gestational age	128
Figure 4.4	Box-and-whiskers plots for the average of : a) P_{TUP} for RF1, b) A_{TRP} for RF1, c) A_{TRP} for RF1, d) P_{TUP} for RF2, e) A_{TRP} for RF2, f) A_{TRP} for RF2, by pathology and gestational age.....	131

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

ACF	Adaptive Comb Filtering
ACP	Analyse En Composantes Principales
AMDF	Average Magnitude Différence Function
ANN	Artificial Neural Network
ANNE	Adaptative Normalised Noise Energy.
ANOVA	ANalysis Of VAriance
BSAT	Best Structure Abstract Tree
CWT	Continuous Wavelet Transform.
DCT	Discrete Cosinus Transform
DME	Dynamic Mean Evaluation
DTW	Dynamic Time Warping
DWT	Discrete Wavelet Transform
FFT	Fast Fourier Transform.
FP	Faux Positif
FN	Faux Négatif
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model.
HPS	Harmonic Product Spectrum
LPC	Linear Predictive Coding
MFCC	Mel-Frequency Cepstrum Coefficients
MMSE	Minimum Mean-Squared Error
PCPP	Periodic Component Probability Presence
PCIS	Pathological Cry Identification System
PSD	Power Spectral Density.
SIFT	Simple Inverse Filtering Tracking.
SPP	Speech Probability Presence
SVD	Singular Value Decomposition
SSM	Smoothed Spectrum Method

XX

STD	Standard de deviation
STE	Short Time Energy.
SVM	Support Vector Machine
TBCS	Mel Scale Admissible Tree by Cry's Samples
TBCC	Mel Scale Admissible Tree by Cry's Classes
VAD	Voice Activity Detector
VP	Vrais Positif
VN	Vrais Négatif
WPT	Wavelet Packet Transform
ZCR	Zero Crossing Rate

LISTE DES SYMBOLES ET UNITÉS DE MESURE

A_{F0}	Average F0
A_{TUP}	Average Duration of Tuning
A_{TRP}	Average Transition Duration
A_{RF1}	Average RF1
A_{RF2}	Average RF2
A_{ratio}	Average RF1:RF2 ratio
A_{SLP}	Average slope
$AP_{ch.mo}$	Average Percentage of changes in cry mode
AP_{dys}	Average Percentage of dysphonic segments
AP_{hyp}	Average Percentage of hyperphonic segments
AP_{irr}	Average Percentage of irregularity of F0
AP_{ph}	Average Percentage of phonic segments
AP_{unv}	Average Percentage of unvoiced segments
$F0$	Fréquence fondamentale
P_{TUP}	Percentage of Tuning
$RF1, RF2$	Fréquences de résonance

INTRODUCTION

0.1 Cadre de recherche

La plupart des nouveau-nés atteints de maladies semblent en bonne santé à la naissance. Ils courent le risque d'avoir de graves problèmes de santé si leurs maladies ne sont pas diagnostiquées au bon moment. Le diagnostic précoce permet un traitement immédiat, réduisant le risque de graves problèmes de santé plus tard dans la vie.

Tous les nouveau-nés subissent le test d'Apgar, dès les premiers jours de leur vie. Ce test est rapide et indolore. Il se base sur l'observation et l'évaluation de cinq états du bébé (couleur de la peau, rythme cardiaque, réflexes, tonicité musculaire, respiration). Le score obtenu est utilisé pour mesurer la santé du bébé pendant ses premiers moments de vie et déterminer si le nouveau-né a besoin d'une assistance particulière ou d'une intervention rapide. Néanmoins, ces scores ne peuvent prévoir les futurs problèmes de santé chez le bébé.

Le cri des nouveau-nés est le premier signe de vie à la naissance. C'est aussi un moyen pour signaler la détresse et la douleur. C'est un comportement dynamique multimodal qui implique des vocalisations caractérisées et des expressions faciales ainsi que des mouvements du corps qui changent à travers le temps. Ce cri inclut d'innombrables caractéristiques temporelles et acoustiques servant de sirène biologique. L'exploitation de ces caractéristiques peut aider les médecins à identifier un bébé avec des dommages dans le système nerveux central (CNS) en l'absence d'autres symptômes (Lester et al. 2008).

Depuis quelques années, les médecins et les phonéticiens analysent les sons des cris dans le but d'améliorer le monitoring des premiers jours des bébés. Ainsi, deux types de recherches ont été réalisés; d'une part l'analyse des sons des cris afin de dresser une symptomatologie acoustique de différentes pathologies et, d'autre part la classification et la reconnaissance des cris en utilisant les techniques de traitement du signal initialement utilisées dans la reconnaissance de la parole.

Cependant, les deux axes de recherche ont évolué séparément, ce qui explique l'absence d'un système opérationnel pour le diagnostic précoce de différentes pathologies. Ces travaux sont restés au stade expérimental sans aucune réalisation utilisable. A cet effet, nous proposons dans ce travail de conduire nos efforts sur les deux axes de recherches pour le développement d'un système de diagnostic fiable. Ce système utilisera les caractéristiques les plus pertinentes qui ont été associées aux différentes pathologies en complément des caractéristiques qui sont généralement utilisées dans les classificateurs.

A cet effet, notre intérêt dans ce projet de recherche est d'analyser les sons des cris des nouveau-nés en vue de développer un système de diagnostic précoce. Ce système ne remplacera pas le travail réalisé par le personnel médical. Mais il sera un support appréciable pour eux, surtout dans les hôpitaux mal équipés.

Pour cela, nous proposons une méthode automatique de caractérisation des cris. Elle se base principalement sur la quantification des caractéristiques les plus pertinentes qui ont été associées à des troubles médicaux sévères. Ces caractéristiques une fois obtenues seront utilisées comme entrée du système de diagnostic en plus des caractéristiques liées à la configuration du conduit vocal.

Nous présentons dans ce qui suit le but principal de notre projet de recherche et les objectifs identifiés pour atteindre ce but. Ce chapitre présente également la méthodologie suivie pour atteindre ces objectifs.

0.2 Problématique et justification

Le cri des nouveau-nés est un comportement précoce qui joue un rôle fondamental dans la survie, la santé et le développement de l'enfant. C'est aussi un moyen de communication moins explicite pour signaler la détresse et la douleur. Il a donné lieu à de nombreuses études tant sur l'interprétation du cri lui-même que sur les techniques d'analyse de ce cri.

Plusieurs chercheurs ont analysé les sons des cris des bébés en se basant sur l'analyse spectrographique pour définir et qualifier certaines caractéristiques en lien avec des pathologies sévères. Ils ont formulé plusieurs hypothèses importantes reprise dans la présente recherche. Selon celles-ci, le cri du bébé est le résultat de son état physique et psychologique ainsi que des stimulations internes et externes liées à son environnement.

D'autres chercheurs ont évalué la performance de différentes méthodes de reconnaissance de la parole pour différencier les cris des bébés en bonne santé de ceux des bébés malades. Ils se sont basés particulièrement sur les paramètres destinés à caractériser la contribution et la dynamique du conduit vocal sans utiliser les caractéristiques acoustiques qui qualifient les cris pathologiques telles que les modes de vibration des cordes vocales et la présence du bruit dans le son des cris.

Face au constat de ces deux axes de recherches existants et en absence d'un système de diagnostic qui permet une évaluation précoce de l'état de santé des nouveau-nés en exploitant leurs cris, ce présent projet vise à utiliser les caractéristiques des cris pour l'identification de certaines pathologies. Pour cela, des techniques avancées de traitement du signal sont d'une part utilisées pour mesurer ces caractéristiques et d'autre part pour établir des relations quantitatives et qualitatives entre les pathologies étudiées et les caractéristiques les plus pertinentes.

Donc, notre problématique de recherche consiste à développer un outil automatique de caractérisation des cris qui sera une partie intégrante du système de diagnostic global qui présentera les avantages suivants :

- possibilité d'extension pour d'autres pathologies non étudiées;
- installation ne nécessitant pas beaucoup de ressources et facilité de déploiement à moindre coût partout dans le monde;
- identification des pathologies qui nécessitent des traitements précoces pour préserver la vie du nouveau-né et réduire les séquelles sur sa santé future.

Pour atteindre nos objectifs, nous avons cherché à répondre à ces interrogations:

1. Quels sont les niveaux d'information que doit inclure la base de données des cris?
2. Comment représenter la base de données de test?
3. Quelles sont les caractéristiques les plus pertinentes dans la différenciation des cris pathologiques étudiés et des cris des nouveau-nés sains?
4. Quelles sont les méthodes de mesure les plus adéquates à l'estimation des caractéristiques sélectionnées ?
5. Comment représenter les caractéristiques estimées?
6. Comment utiliser ces caractéristiques dans le système de diagnostic global?

0.3 Objectifs

Ce travail de thèse s'inscrit dans le prolongement de plusieurs autres études effectuées sur les cris des bébés depuis les années soixante. Nous avons cherché à améliorer la précision de la reconnaissance des cris pathologiques par l'utilisation combinée de différentes sources d'information. Le travail réalisé est centré sur l'exploitation des caractéristiques acoustiques représentant la configuration du conduit vocal ainsi que les modes de vibration des cordes vocales.

Nous avons plusieurs objectifs:

- a. Créer une base de cris des nouveau-nés, les uns en bonne santé et les autres atteints de certaines pathologies;
- b. Identifier les paramètres prépondérants dans la différenciation des cris normaux et pathologiques;
- c. Établir des relations qualitatives et quantitatives entre ces paramètres et les différents cris pathologiques des nouveau-nés;
- d. Catégoriser les pathologies étudiées selon la variabilité des caractéristiques des cris des nouveau-nés;
- e. Développer un outil d'analyse dédié aux bébés. Il nous permettra d'effectuer des mesures précises des différents paramètres du cri qui seront utilisés comme entrée du système de diagnostic global.

0.4 Méthodologie

Pour la réalisation de nos objectifs, nous avons adopté la démarche de recherche qui est représentée par un diagramme en bloc sur la Figure 0.1.

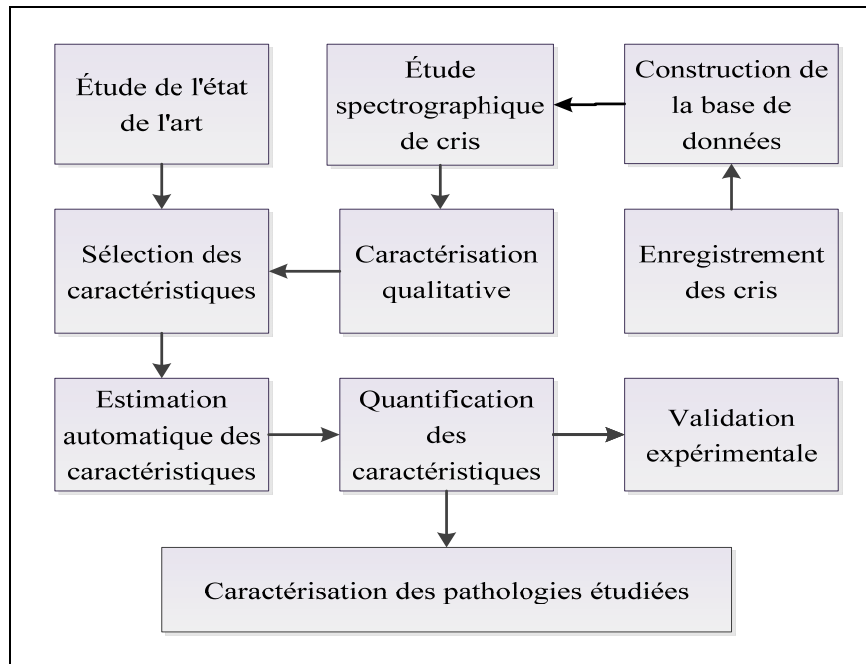


Figure 0.1 Méthodologie

0.4.1 Étude de l'état de l'art

Dans cette partie du travail, il s'agit de dresser un état de l'art sur les recherches concernant les cris des bébés. L'analyse et la critique des travaux qui ont été réalisés dans le domaine nous ont permis de déterminer nos objectifs et notre plan de travail pour apporter une contribution significative dans le domaine médical et le domaine du traitement du signal. L'état de l'art est détaillé dans le chapitre 1 et dans les trois articles qui sont inclus dans cette thèse.

Nous avons également étudié et identifié: 1) le comportement physioacoustique des cris; 2) les caractéristiques qui ont été standardisées; 3) les méthodes de mesure de ces caractéristiques. Ce travail nous a permis de faire une présélection des caractéristiques qui serviront pour la caractérisation des pathologies étudiées.

0.4.2 Enregistrement des cris

La première phase du projet a consisté à enregistrer les cris des nouveau-nés. Ces enregistrements ont été réalisés au sein du département de pédiatrie / service de néonatalogie de l'hôpital Sainte-Justine sous la supervision du responsable médical du projet et avec l'aide de collaborateurs médicaux.

Le choix des pathologies étudiées a dépendu de la disponibilité des sujets et des recommandations de notre spécialiste médical à l'hôpital Sainte-Justine. L'enregistrement des cris a été effectué par l'infirmière de recherche chargée de la collecte des cris.

Compte tenu de l'augmentation spectaculaire du nombre de naissances prématurées un peu partout dans le monde, l'enregistrement des cris concerne aussi les nouveau-nés prématurés en plus des nouveau-nés à terme, parmi eux des bébés en bonne santé et des bébés malades.

L'âge des bébés lors de l'enregistrement ne dépasse pas 30 jours parce qu'à ce moment-là les nouveau-nés ne contrôlent pas leurs cris. Pour chaque enregistrement plusieurs niveaux d'information sont collectés (*voir* Figure 0.2).

Dans le but de fiabiliser la base des cris, un suivi sur 6 mois (pathologies non détectées / diagnostiquées) a été effectué.

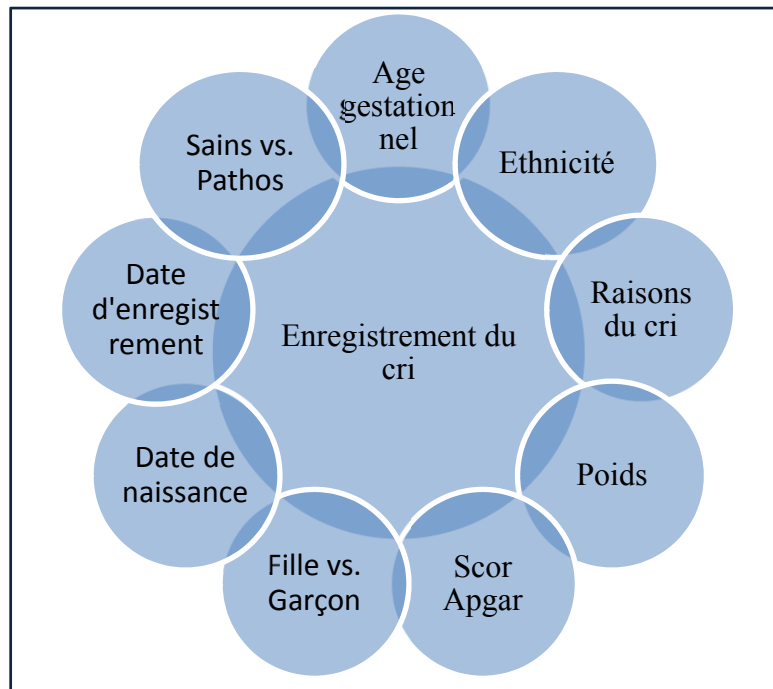


Figure 0.2 Les différents niveaux d'information qui concernent les cris des bébés enregistrés

0.4.3 Construction de la base de données

Une fois les signaux des cris collectés, il est nécessaire d'effectuer un prétraitement pour trouver les parties les plus utiles à l'analyse des cris dans la totalité des signaux.

Les enregistrements des cris contiennent des sons indésirables, comme le bruit ambiant, l'écho, les sons d'appareils, etc. qui peuvent affecter les résultats de l'analyse. À cet effet, nous avons utilisé le logiciel libre-service Praat (P. Boersma) pour effectuer une segmentation et un filtrage manuel du bruit. Après le prétraitement, les signaux sont répartis par pathologie et par âge gestationnel.

0.4.4 Étude spectrographique de cris

L'analyse spectrographique des cris en utilisant Praat nous a permis d'identifier les différentes caractéristiques des cris qui ont été définies dans la littérature. La visualisation

des signaux de cris montre une différence significative entre les caractéristiques des cris des bébés sains et celles de ceux qui sont malades. Ce travail nous a permis de compléter la caractérisation qualitative des cris pathologiques par rapport à ceux des bébés sains.

0.4.5 Caractérisation qualitative

Sachant que la variabilité des caractéristiques des cris a été toujours liée aux différentes conditions de santé des bébés, nous avons jugé utile son utilisation pour la caractérisation des cris pathologiques étudiés. Ce travail nous a permis de distinguer les différents types de cris, d'établir des relations qualitatives entre les caractéristiques acoustiques et les pathologies étudiées et pour finir d'identifier les caractéristiques les plus pertinentes.

0.4.6 Sélection des caractéristiques

La sélection des caractéristiques des cris s'est faite d'une part sur la base des caractéristiques des cris qui ont été déjà identifiées dans la littérature et associées à des conditions médicales sévères et d'autre part, sur l'étude de la caractérisation qualitative des cris étudiés, qui a été citée précédemment.

Les caractéristiques sélectionnées une fois estimées et quantifiées serviront comme une entrée du système de diagnostic global.

0.4.7 Estimation automatique des caractéristiques

Étant donné que la précision des mesures des caractéristiques sélectionnées et leurs variations dans le temps sont déterminantes pour obtenir des informations fiables sur l'état de santé des nouveau-nés, nous avons commencé par identifier les méthodes de mesure les plus adéquates aux caractéristiques des cris des nouveau-nés.

Pour cela, nous avons comparé les différentes méthodes de mesures qui sont proposées dans la littérature selon les caractéristiques sélectionnées. Notre choix s'est porté sur les algorithmes qui permettent de minimiser la surestimation et la sous-estimation des caractéristiques ainsi que sur les méthodes qui ont été testées avec succès sur des voix pathologiques. Le détail des méthodes utilisées se trouve dans les chapitres 2, 3 et 4 ainsi que dans l'article de conférence en annexe II.

0.4.8 Quantification des caractéristiques

La quantification des caractéristiques consiste à faire correspondre pour chaque signal de cri, une valeur concernant une certaine caractéristique donnée. Sachant que le signal du cri est considéré stationnaire sur une très courte durée, l'estimation des caractéristiques est faite principalement pour chaque segment de cri de 20ms. À cet effet, pour chaque signal de cri d'une durée allant jusqu'à 3 minutes, correspond un nombre de valeurs estimées pour chaque caractéristique.

Pour cela, nous avons adopté des méthodes mathématiques et statistiques permettant de quantifier les caractéristiques estimées pour chaque signal de cri. Les méthodes de quantification utilisées sont : la moyenne, le pourcentage, l'écart-type ainsi que la médiane et l'écart interquartile.

0.4.9 Validation expérimentale

Au cours de notre travail de caractérisation des cris, nous avons entrepris des tests de validation concernant la faisabilité du système de diagnostic ainsi que l'apport de certaines caractéristiques acoustiques quantifiées dans la classification des cris pathologiques.

La première série d'expériences a été conduite pour différencier deux classes de cris (pathologiques et sains). Pour cela nous avons utilisé les paramètres tels que : les coefficients MFCC (Mel-Frequency Cepstrum Coefficients) et les coefficients WPT (entropie) (Wavelet

Packet Transform). Nous avons testé aussi l'apport de l'application de l'ACP (analyse en composantes principales) et l'algorithme BSAT (Best Structure Abstract Tree) pour réduire la dimension des coefficients WPT. La méthodologie et le détail des résultats obtenus se trouvent dans l'article de conférence qui est présenté en annexe I.

La deuxième série d'expériences réalisées concerne principalement l'utilisation de quelques caractéristiques acoustiques qui ont été quantifiées comme le pourcentage de l'hyperphonation et le pourcentage de l'irrégularité de la fréquence fondamentale. Deux types de tests ont été conduits : 1) différencier deux classes de cris (pathologiques et sains) avec ou sans utilisation des caractéristiques quantifiées pour les deux catégories de bébés (prématurés et à terme). 2) distinguer les cris selon les pathologies étudiées avec ou sans utilisation des caractéristiques quantifiées pour les deux catégories de bébés (prématurés et à terme). La méthodologie et le détail des résultats trouvés, sont présentés dans le chapitre 2.

0.4.10 Caractérisation des pathologies étudiées

Nous finalisons notre travail par la caractérisation des pathologies étudiées. Pour chaque pathologie étudiée nous associons la variabilité des caractéristiques quantifiées les plus discriminatives. Cette variabilité est évaluée par rapport aux caractéristiques des cris des bébés en bonne santé.

0.5 Organisation de la thèse

Cette thèse s'organise en cinq chapitres. Dans l'introduction, nous avons présenté une description de la problématique de notre projet de recherche ainsi que nos objectifs et la méthodologie suivie pour atteindre ces objectifs. Le chapitre 1 fait une synthèse de l'état de l'art. Les chapitres 2, 3, et 4 rassemblent les articles publiés ou soumis à des journaux scientifiques avec comité de lecture. Le chapitre 5 présente la synthèse des résultats et la conclusion.

Le premier chapitre est consacré à la synthèse bibliographique où nous présentons des définitions sur les cris des bébés et leurs caractéristiques ainsi que les méthodes de mesures utilisées dans ce domaine. Ce chapitre est consacré aussi aux travaux effectués sur les cris des bébés.

Le deuxième chapitre présente le premier article intitulé : « Frequential Characterization of Healthy and Pathologic Newborns Cries ». Cet article publié dans le journal « American Journal of Biomedical Engineering » décrit la caractérisation qualitative et quantitative des cris pathologiques et des cris des nouveau-nés en bonne santé en utilisant deux principales caractéristiques fréquentielles telles que l'hyperphonation et l'irrégularité de la fréquence fondamentale. Ce chapitre présente aussi une analyse statistique des résultats obtenus ainsi qu'une validation expérimentale concernant l'utilisation des caractéristiques étudiées dans le système de diagnostic des bébés malades.

Le troisième chapitre présente le deuxième article intitulé : « Acoustic Measures of The Cry Characteristics of Healthy Newborns and Newborns with Pathologies » qui est publié dans le journal « J. Biomedical Science and Engineering ». Dans cet article nous introduisons l'utilisation de nouvelles caractéristiques acoustiques dans la discrimination des cris pathologique telles que la dysphonation, le changement de mode et les cris non voisés. Dans le cadre de l'estimation de la dysphonation et dans le but de remédier à la surestimation du bruit dans les sons des cris, nous avons proposé et implémenté une nouvelle méthode de mesure adaptée aux cris des nouveau-nés. L'analyse statistique des résultats obtenus a été effectuée aussi dans cet article.

Le quatrième chapitre concerne le troisième article intitulé: « Resonance Frequencies Behavior in Pathologic Cries of Newborns ». Cet article qui est publié dans le «Journal of Voice» introduit de nouveaux paramètres pour la caractérisation des cris. Ces paramètres qualifient principalement l'interaction des fréquences de résonance avec les harmoniques de la fréquence fondamentale. Les résultats obtenus par la caractérisation des cris en utilisant

les paramètres étudiés permettent la discrimination des cris pathologiques, car la distribution de ces paramètres dépend de la pathologie qui touche le nouveau-né.

Nous terminons cette thèse par le chapitre 5 qui présente la synthèse des articles et les principaux résultats obtenus. Nous présentons également dans ce chapitre les contributions de notre recherche et nos recommandations.

CHAPITRE 1

ETAT DE L'ART

1.1 Introduction

La première partie de ce chapitre est consacrée à une revue de la littérature sur le comportement du cri en commençant par le modèle physioacoustique et la neurophysiologie des cris ainsi que sur les différentes définitions des caractéristiques des cris qui ont été associées à des conditions pathologiques sévères. Puis sont exposés les travaux antérieurs menés sur les cris des bébés, principalement les travaux d'analyses acoustiques, travaux de perceptions des cris ainsi que les travaux de classification et de reconnaissance des cris. Enfin sont décrites les différentes techniques de base ainsi que les moyens d'analyse et de classification des cris des nouveau-nés.

1.2 Le modèle physioacoustique du cri

Le cri des bébés est défini comme l'alternation rythmique de sons (énoncés) et d'inspiration. Donc, il fait partie de la phase d'expiration, pendant laquelle le larynx produit un son ou une phonation (LaGasse, Neal et Lester, 2005).

Le modèle physioacoustique de la production du cri comprend deux parties : une partie acoustique qui décrit la génération du son et une partie physiologique qui décrit le mouvement et la configuration du conduit vocal.

Le mécanisme de production des cris chez les nourrissons ressemble au processus de production de la parole chez les adultes. Mais la forme du conduit vocal des nouveau-nés est différente de celle des adultes. Elle ressemble plutôt à celle des « chimpanzés ». Chez les nourrissons, le larynx est positionné très haut dans le conduit vocal qui est plus court que celui de l'adulte (*voir* Figure 1.1). Dans ce cas, l'espace pharyngé qui constitue un des résonateurs les plus importants dans la phonation est réduit. C'est pour cette raison que les

bébés ne respirent pas par leur bouche quand leur nez est bouché. La position du larynx commence à descendre à partir de 6 mois, elle atteint la position des adultes à deux ans (Laitman JT, 1980; Laitman, Heimbuch et Crelin, 1978).

La production du cri est un processus complexe dans lequel sont impliqués de nombreux organes (*voir* Figure 1.2). La source du cri provient des poumons qui émettent un flux d'air. Ce flux d'air va traverser le larynx pour faire vibrer ou non les cordes vocales à une fréquence fondamentale F_0 . Le son des cris est déterminé alors par la vibration des cordes vocales qui se trouvent dans le conduit vocal et par d'autres vibrations d'ondes à des fréquences différentes nommées ondes complexes. L'onde sonore devient plus complexe en traversant les voies respiratoires produisant une concentration de l'énergie sonore à des niveaux élevés nommés formants (Lester et LaGasse, 2008). En d'autres termes, la dimension et le contour du conduit vocal supérieur qui est constitué par les cavités nasales, la bouche et le pharynx forment le son des cris en produisant des fréquences de résonances ou des formants RF1 et RF2 au-dessus de F_0 (Lester et LaGasse, 2008).

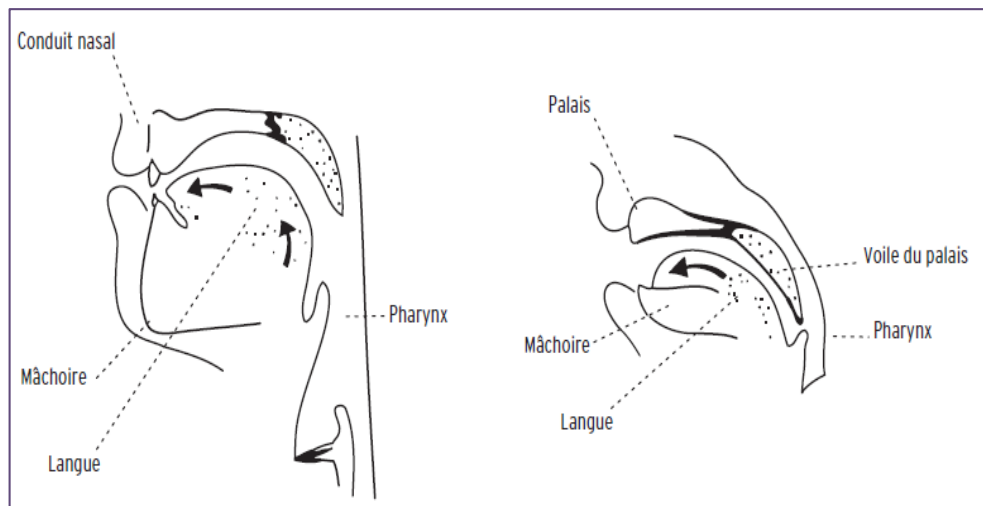


Figure 1.1 Section du conduit vocal des adultes et des nouveau-nés
Tirée de Delahaie (2009, p. 21)

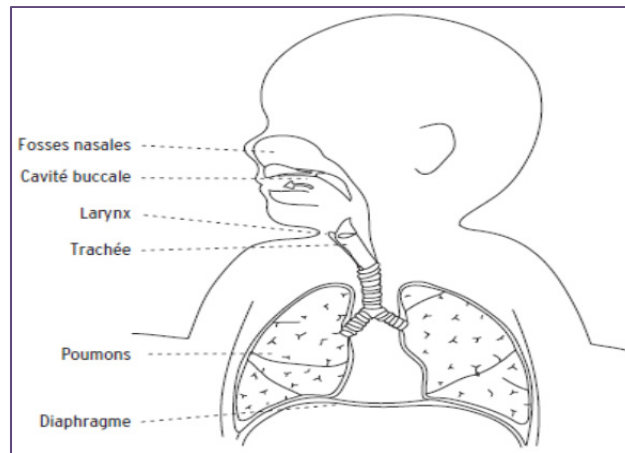


Figure 1.2 Anatomie du cri
Tirée de Delahaie (2009, p. 21)

Le cri des bébés change avec l'âge grâce à la maturation de l'appareil vocal et à l'évolution du système de contrôle neural et aussi grâce au développement des muscles de la cage thoracique qui induit une réduction dans la vitesse de respiration durant les cris. Chaque cri d'enfant est acoustiquement unique. Il est caractérisé par des sons voisés et des sons non voisés ainsi que des F0 plus élevées que celles des adultes. Pour la plupart des enfants, F0 varie entre 250 Hz et 450 Hz avec un premier formant RF1 de 1200 Hz et un deuxième formant RF2 de 3200 Hz. Les sons non voisés ou turbulences sont produits par le passage de l'air à travers les constriction du conduit vocal et les sons voisés sont produits par le passage de l'air à travers la glotte (Manfredi C., 1998).

1.3 La neuropsychologie du cri

Le cri est une réaction du bébé à une stimulation aversive interne ou externe (faim, douleur...). Il est premièrement initié dans le cerveau du bébé qui traduit cette stimulation en une série de commandes à travers le système nerveux aux organes de production de la parole (Jam et Sadjedi, 2009). Donc, le cri est le résultat d'une interaction complexe entre plusieurs structures anatomiques et des mécanismes physiologiques. Cette interaction implique le système nerveux central (CNS), le système respiratoire, le système nerveux périphérique et une variété de muscles (Soltis, 2004). C'est le CNS qui contrôle la capacité, la stabilité et la

coordination des mouvements de ces muscles (Michelsson et Michelsson, 1999). Donc, le cri des nouveau-nés peut nous informer sur le niveau de fonctionnement du CNS.

Le signal du cri est aussi le résultat d'une coordination de plusieurs régions du cerveau y compris le tronc cérébral, mésencéphale (Midbrain) et le système limbique qui contrôle la respiration et la vibration des cordes vocales (LaGasse, Neal et Lester, 2005; Michelson Katarina 2007; Zeskind et Lester, 2001).

Le tronc cérébral inférieur du cerveau contrôle les muscles du larynx, du pharynx, la poitrine et le cou supérieur à travers les nerfs crâniens et les nerfs phréniques et thoraciques (Lester et Boukydis, 1992).

D'après (LaGasse, Neal et Lester, 2005):

- les éléments responsables de la formation de F0 sont la variation de tension des muscles du larynx, cricothyroïde vocale et les muscles respiratoires abdominaux;
- le tronc cérébral contrôle également le contour des voies respiratoires transversales du système supra-glottal, qui façonne les fréquences des formants;
- les autres régions du crâne telles que le système limbique et l'hypothalamus participent dans l'initialisation du cri.

Selon (Michelson Katarina 2007) : 1) un déficit dans le fonctionnement du cerveau peut affecter le contrôle du cri; 2) les modèles atypiques de F0 sont dus aux dommages dans le nerf complexe crânien vagal; 3) le décalage rapide (shift) et la haute variabilité de F0 sont dus à une instabilité dans le système de contrôle neural.

1.4 Les caractéristiques du cri

Une certaine terminologie des manifestations acoustiques du signal du cri a été établie dans différentes études (*voir* Figure 1.3). Dans (Michelson Katarina 2007; LaGasse, Neal et Lester, 2005; Lester et LaGasse, 2008) les auteurs identifient et définissent les caractéristiques et les modes des cris des bébés suivantes :

Temps de latence

Intervalle de temps entre le stimulus et la première réponse de l'enfant c.-à-d. le début du cri.

Seuil

Nombre de stimulations pour provoquer le cri.

Nombre d'énoncés

Nombre de phonations expiratoires qui se produisent durant le cri.

Période inspiratoire

L'intervalle en secondes entre le premier et le deuxième énoncé.

Amplitude

L'énergie moyenne en dB durant un énoncé.

Durée

Tous les cris se produisent entre deux inspirations. Donc la durée du cri correspond à toute la vocalisation durant une seule expiration. Dans les cris induits par la douleur, la durée du premier cri inclut le temps de latence. Sur le spectrogramme il est facile de déterminer le début et la fin du cri, car l'intensité de la voix est élevée au début du signal et à la fin elle est souvent caractérisée par la diminution graduelle de l'intensité du son et par la disparition de la fondamentale et de ses harmoniques.

Il est difficile de comparer la durée des cris dans les différentes études, car les auteurs ne mentionnent pas quel segment ils incluent dans leurs analyses.

Seconde pause

Elle est définie comme l'intervalle de temps entre la fin d'une phonation et la prochaine inspiration.

Fréquence fondamentale (F0)

Le spectrogramme du son affiche la tonalité voisée comme des lignes noires. La fréquence fondamentale est la plus basse ligne, ses harmoniques apparaissent en dessus comme des lignes parallèles. La fréquence fondamentale maximale correspond au point le plus élevé de la fréquence fondamentale mesurable sur le spectrogramme et la fréquence fondamentale minimale correspond au point le plus bas de la fréquence fondamentale mesurable sur le spectrogramme.

Décalage (shift)

Cette caractéristique dénote l'augmentation ou la diminution rapide de la fréquence fondamentale. Dans le cri normal, la partie « shift » est observée comme une hyperphonation à intensité élevée au commencement de la phonation.

Mélodie

Les différents types de mélodies sont classés comme suit : décroissante, croissante-décroissante, plate ou sans type avec un cri complètement instable représentant un son inaudible ou « glottal plosives ».

Continuité du signal

Un cri est défini comme le son entre deux inspirations. Il peut être continu ou constitué par plusieurs parties séparées par des silences.

Voix

Les cris sont voisés ou dysphoniques. Dans le cri voisé, l'onde sonore est périodique et la fondamentale et ses harmoniques sont visibles sur le spectrogramme. Dans les cris dysphoniques inaudibles, le spectrogramme affiche une bavure de turbulence ou un bruit

apériodique avec une fondamentale qui n'est pas visible ni mesurable. Le signal semi-voisé comprend deux parties « voisé et inaudible ».

« **Vibrato** »

Il est composé d'au moins quatre mouvements de montées et de descentes de la fondamentale.

« **Glottal roll** »

Il représente la fin d'un cri avec un son de faible intensité. La fondamentale et les harmoniques sont visibles, mais difficilement mesurables parce qu'elles sont faibles.

Harmoniques doubles pauses

Elles sont visibles entre la fréquence fondamentale et ses harmoniques comme une série d'harmoniques parallèles.

Bi-phonation

Elle est caractérisée par la présence d'une double série de fréquences fondamentales induites par la vibration de deux sources. En comparaison avec les harmoniques double pause les harmoniques de la bi-phonation n'ont pas une forme de mélodie parallèle.

Glissement

Il est défini comme une rapide augmentation ou diminution continue de la fréquence fondamentale. Le changement dans ce cas est d'au moins de 600 Hz en 0.1 secs dans la fréquence fondamentale et pas dans ses harmoniques.

Division « Burcation »

C'est une division soudaine dans la F0. Elle se traduit par la division du signal de cri qui est fort en une série de signaux plus faibles, avec leurs propres contours fondamentaux.

« Glottal plosive »

C'est une rapide fermeture et ouverture des cordes vocales qui cause une brève turbulence d'air.

Phonation

C'est un son voisé généré par la vibration des cordes vocales à une F0 variant entre 350 et 750Hz. Elle est définie comme un mode basique d'oscillation de la glotte. Sur le spectrogramme, elle est caractérisée par des harmoniques bien définies et équidistantes.

Hyperphonation

Son voisé caractérisé par une F0 très élevée variant de 1000 à 2000 Hz. Généralement, c'est une modification brusque du cri basique.

Dysphonation

Cri bruyant dû à un mouvement apériodique des cordes vocales. Ce mode implique une altération significative du mode basique des cris. Sur le spectrogramme, il est généralement caractérisé par des harmoniques mal définies non équidistantes.

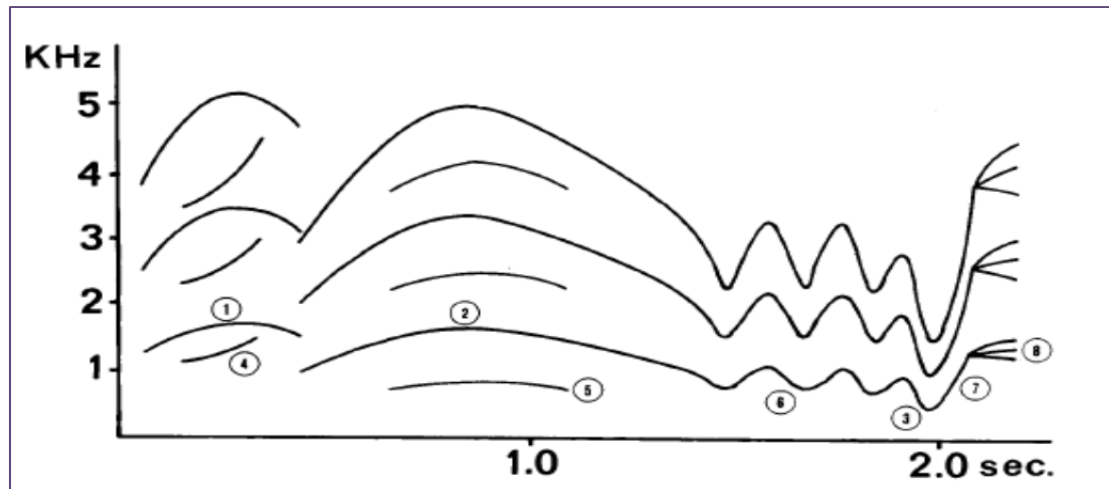


Figure 1.3 Diagramme d'un spectrogramme du cri
 1) Shift et le max de F0, 2) Max de F0 moyen, 3) Min F0, 4) Bi-phonation
 5) Harmonics double pauses, 6) Vibrato, 7) Glide, 8) Division
 Tirée de Michelsson et Michelsson (1999, p. 298)

1.5 Les caractéristiques des cris dans différents cas cliniques

À la naissance les bébés possèdent deux types de cris, le cri basique et le cri de douleur. Ces cris sont clairs et différemment audibles à l'oreille humaine et correctement distingués par l'analyse acoustique. Le cri de douleur est généralement fort et débute soudainement avec une fréquence fondamentale élevée et de longues détentions de souffle. Le cri basique est typiquement entendu quand l'enfant a faim. La fréquence fondamentale est faible et la durée de l'inspiration est courte. Les deux cris basiques et celui de douleur du nouveau-né de 4-6 semaines, possèdent une qualité réfléchie et automatique. À ce moment-là, le système nerveux de l'enfant arrive à la maturité : le bébé acquies un contrôle volontaire de ses cordes vocales (Lester et LaGasse, 2008).

Les analyses spectrographiques effectuées par plusieurs chercheurs sur les cris des bébés en bonne santé indiquent que la fréquence fondamentale varie entre 400 et 600 Hz, souvent avec des mélodies décroissantes ou croissantes-décroissantes. La fréquence fondamentale est totalement stable. Selon les techniques de mesure utilisées par les différents chercheurs, la durée moyenne de latence varie considérablement de 0.6 à 3.6 secs et la durée de phonation

varie entre 1.5 à 6.5 secs (Verduzco-Mendoza et al., 2009). Le décalage « shift » est fréquent au début du signal dans les cris de douleur et les cris spontanés, mais il est rare durant les premiers jours de vie de l'enfant. Contrairement à l'harmonique double pause et le « glottal roll » qui sont assez fréquents dans les cris des enfants en bonne santé, la biphonation, le glissement et la division « bifurcation » ne se produisent jamais ou rarement.

Dans (LaGasse, Neal et Lester, 2005; Lester et LaGasse, 2008; Michelson Katarina 2007), les auteurs ont reporté l'association des caractéristiques des cris aux différentes conditions pathologiques (voir Tableau 1.1). Cette association a été établie par différents chercheurs en utilisant l'analyse spectrographique. D'après ces recherches, les bébés qui ne sont pas en bonne santé spécialement ceux qui souffrent de désordres dans le système nerveux central, ont des cris qui se caractérisent par une fréquence fondamentale élevée, un contour de mélodie irrégulier et une longue latence entre le stimulus de douleur et le début du cri.

Les auteurs dans (LaGasse, Neal et Lester, 2005) indiquent aussi que toutes les études effectuées sur les cris ont montré malgré des différences dans les mesures, qu'une fréquence fondamentale élevée ou des cris aigus sont systématiquement des indices d'un problème neurologique.

Tableau 1.1 Récapitulatif des caractéristiques des cris qui ont été associées à certains états de santé

Conditions médicales	Caractéristiques des cris
Asphyxie	↑F0, ↑ variabilité de F0, bi-phonation, ↑ pause sous-harmonique, ↓ durée
Domage dans le cerveau	↑F0, ↑ variabilité de F0, bi-phonation, ↑ seuil, ↓ durée, ↑ latence, ↑ énoncés courts
Cri du chat	↑F0
Syndrome bas	↑F0, ↑ variabilité de F0, ↓amplitude

Conditions médicales	Caractéristiques des cris
Hydrocéphalie	↑F0, ↑ variabilité de F0, ↓ latence
Hypothyroïdie	↓F0
Maladie de krabbe	↑F0
Méningite bactérienne	↑F0, ↑ variabilité de F0, bi-phonation, ↓ durée
Trisomie 13,18, 21	↓F0
Hyperbilirubine	↑F0, ↓ durée, ↓ latence, ↑ variabilité de F0 et RF1, ↑ changement de mode.
Bébé Prématuré	↑F0, ↓ durée ↓ variabilité de RF1, ↑ variabilité de F0, ↑ nombre d'énoncés courts
Poids faible à la naissance	F0, ↑ variabilité de F0, bi-phonation
Excitabilité	↓énoncés, ↑Latence, ↓Amplitude (énergie), ↑dys-phonation
Dépression	↑durée, ↑F0, ↑RF1, ↑Variabilité de RF1
Faible Contrôle respiratoire	↓durées des énoncés, ↑période d'inspiration, ↓nombre d'énoncés, ↑ nombre d'énoncés courts, ↓Amplitude, ↑variabilité dans l'amplitude, ↑dys-phonation
Faible contrôle du conduit vocal	↑Fo, ↑RF1, ↑RF2, ↑Hyperphonation, ↑Changement de mode, ↑Variabilité dans F0, ↑Variabilité dans RF1, RF2
Réactivité atypique faible	↑latence, ↑seuil
Réactivité atypique élevé	↓latence

1.6 Travaux antérieurs

1.6.1 L'analyse acoustique des cris

La méthode spectrographique du son est le premier moyen utilisé dans l'analyse acoustique des cris. Les premières études menées sur le cri néonatal avec cette méthode ont débuté entre les années soixante et les années quatre-vingt. Les différents paramètres utilisés dans ces recherches sont mesurables dans le spectrogramme du cri. La fréquence fondamentale et ses harmoniques sont obtenus à l'aide d'un filtre à bande étroite. Elles apparaissent sur le spectrogramme comme des lignes noires. Dans le cas des cris voisés, la fréquence fondamentale est représentée par la plus basse ligne horizontale. Les harmoniques sont vues comme un multiple précis de cette fondamentale (Michelson Katarina 2007).

Parmi les recherches qui ont utilisé le spectrogramme, nous retrouvons (Fisichelli et Karelitz, 1966; Thodén et Michelsson, 1979; Michelsson, 1971; Michelsson, Tuppurainen et Aula, 1980; Michelsson, Sirviö M.A et Wasz-Höckert, 1977; Michelsson, Sirviö et Wasz-Höckert, 1977; Lind et al., 1970). Ils se basaient principalement sur l'identification auditive des cris et l'analyse spectrographique visuelle pour définir et qualifier les différents phénomènes acoustiques des cris. Ils ont réussi à établir certaines terminologies des manifestations des signaux des cris. Leur but était de mettre en évidence la relation qui existe entre les signaux des cris des bébés et leurs statuts neurologiques. Ils examinaient d'ailleurs les cris des bébés déjà diagnostiqués de dommages neurologiques (maladie de Krabbe, syndrome bas, asphyxie, méningite) pour qualifier les modèles atypiques de ces cris. La conclusion de leurs travaux est que les variations de la fréquence fondamentale (F0) sont étroitement liées aux statuts neurologiques des bébés. Une F0 très élevée qui présente une très grande variabilité est un indice de problèmes neurologiques.

D'autres travaux menés par (Michelsson, Järvenpää et Rinne, 1983; Lester, 1987; Vohr et al., 1989; Corwin et al., 1992) étaient focalisés sur les bébés à risque avec un potentiel neurologique faible tel que la prématurité, l'Hyperbilirubine ainsi que l'exposition prénatale à la drogue. Ils ont trouvé que le statut médical de ces bébés influe sur certaines

caractéristiques comme la variabilité de F0, RF1 ainsi que les durées des énoncés « utterances » des cris.

L'étude menée sur les cris de douleur par (Lester et al., 1992) a montré que les bébés ont une réelle capacité à signaler leur douleur à leurs parents qui la perçoivent comme une urgence.

Une autre application remarquable de l'analyse des cris a été faite par (Golub et Corwin, 1982 ; Corwin et al., 1995) sur des enfants victimes d'une mort subite connue par le syndrome SIDS (Sudden Infant Death Syndrome). Ils déduisent que le syndrome SIDS est caractérisé par une élévation importante de RF1 et le changement de modes dans les cris.

Par ailleurs, d'autres études ont été faites sur des nouveau-nés en bonne santé et plus particulièrement celle de (Michelsson et al., 2002) qui a concerné 172 nouveau-nés en bonne santé âgés de 1 à 7 jours. Ces travaux ont permis d'identifier la plage de variation de F0 ainsi que la durée moyenne des cris des bébés sains.

La recherche sur les cris des bébés a évolué au fil des années avec le développement des équipements d'enregistrements et des outils de traitement du signal. Certaines études plus récentes ont utilisé des outils plus modernes de traitement numérique du signal pour analyser les signaux des cris. Ces outils sont des systèmes d'analyses désignés initialement pour le traitement du langage des adultes. Dans (Wermke et al., 2005), les auteurs ont utilisé *CSL-Speech 4400/MDVP* pour étudier l'interaction entre la mélodie et les fréquences de résonances chez des bébés âgés de 3 à 4 mois. Dans (Buder et al., 2008), les auteurs ont exploité le système d'analyse nommé *TF32* pour définir et classifier les régimes de vibrations phonatoires des cris des bébés. Dans (Verduzco-Mendoza et al., 2009), les auteurs se sont servis du système d'analyse nommé *Cool Edit 2000* pour réaliser une étude qualitative et quantitative sur les cris des bébés avec gestation à haut risque. Dans (Wermke et Robb, 2010), les auteurs ont utilisé le système d'analyse nommé *CSL-Speech 4400/MDVP* pour étudier l'effet de la taille du bébé sur la fréquence fondamentale de ses cris. Dans

(Cecchini, Lai et Langher, 2010), les auteurs ont exploité le système d'analyse nommé *Cool Pro* pour étudier la perception des cris dysphoniques.

Toujours dans le but d'améliorer l'analyse des cris des bébés, d'autres types de travaux ont été réalisés. L'objectif principal était de développer des algorithmes qui s'appuient sur l'analyse numérique du signal pour permettre une analyse efficace des cris. Parmi ces études, nous trouvons les études menées par Varallyay dans (Várallyay Jr et al., 2005; Várallyay, 2006; Varallyay Jr György, 2008; Várallyay, 2007). Les sons des cris des bébés sourds et les sons des cris des bébés en bonne santé ont été utilisés. L'âge des bébés variait de 1 à 15 mois. Ce chercheur propose la méthode (HPS) pour la segmentation des cris et la méthode SSM pour l'estimation de F0 ainsi qu'une nouvelle méthode pour déterminer la mélodie des cris.

Nous retrouvons aussi l'étude menée par (Lederman, 2010) qui a proposé l'algorithme SIFT modifié pour l'estimation de la fréquence fondamentale (F0). Il a testé cet algorithme sur des signaux de cris réels. Ses résultats indiquent que le SIFT modifié permet d'éviter la surestimation et la sous-estimation de F0. Une autre étude aussi a été faite par (Manfredi, Tocchioni et Bocchi, 2006) qui propose pour le traitement automatique des cris la méthode NNE pour l'estimation du bruit dans les cris dysphoniques et la méthode AR-PSD pour l'estimation des fréquences de résonances, ainsi que le SIFT pour l'estimation de la fréquence fondamentale.

Enfin, d'autres chercheurs ont développé leurs propres systèmes d'analyses, dédiés spécialement aux cris des bébés. Dans (Lester et LaGasse, 2008), les auteurs ont utilisé un outil nommé « criogram ». (LaGasse, Neal et Lester, 2005) se sont servis d'un système nommé « CRI » et (Manfredi et al., 2009) ont utilisé un outil de mesure nommé « BioVoice ». Ces systèmes d'analyses numériques permettent d'extraire les caractéristiques couramment utilisées dans l'évaluation des cris. Contrairement aux logiciels qui permettent l'analyse acoustique du langage et des sons, les systèmes développés ne sont pas

commerciallement disponibles et nécessitent une expertise pour l'interprétation des résultats obtenus.

1.6.2 La perception des cris

Dans les études faites sur les cris des bébés, certains auteurs se sont concentrés sur la fonction communicative des cris des bébés pour éclaircir le processus production/perception du cri néonatal. Ils ont analysé la relation entre la perception des adultes et les caractéristiques acoustiques des cris avec l'objectif d'identifier les caractéristiques atypiques qui interfèrent avec la réponse des personnes qui prennent soin du bébé.

Les auteurs de (Leger et al., 1996) ont examiné la relation entre les caractéristiques des cris et la perception des adultes dans le cas des bébés de 1 et 6 mois. Ils trouvent que :

- l'évaluation des adultes est systématiquement liée aux caractéristiques des cris;
- l'âge du bébé et les caractéristiques de son cri sont les facteurs les plus importants dans la perception des cris par rapport aux caractéristiques du perceuteur comme le sexe et l'expérience dans la garde d'enfant;
- les adultes réalisent un jugement concernant la nature du cri avant même d'entendre toutes les pulsations du cri;
- les caractéristiques des cris comme la vitesse de l'impulsion, le modèle de pause, la durée du cri, le ratio signal sur bruit et la tendance linéaire de la fréquence fondamentale influencent la perception des adultes.

Julia montre dans son étude (Irwin, 2003) sur les bébés âgés de 3, 6, 8 et 12 mois que les cris sont perçus avec une haute détresse lorsque leurs durées sont plus longues. Pour beaucoup de chercheurs, F0 est considérée comme le déterminant le plus important dans la perception des

cris des bébés par les adultes. Généralement les cris avec une F0 très élevée sont évalués comme indices d'urgences, de détresses et de maladies.

Les auteurs de (LaGasse, Neal et Lester, 2005) ont démontré que les cris caractérisés par : une longue durée de la phase d'expiration, une courte pause, une fréquence fondamentale élevée, un grand pourcentage de cri dysphonique et hyperphonique ainsi que la diminution de l'amplitude et le rapport signal sur bruit sont estimés comme aversifs par les adultes qui surveillent les bébés.

Une autre étude statistique effectuée par (Cecchini, Lai et Langher, 2010) pour déterminer la relation qui existe entre les caractéristiques des cris et les trois niveaux d'évaluations (douleur, faim et recherche de soin) a démontré :

- une corrélation positive entre le cri dysphonique, le pourcentage d'expiration vocale et la perception de la faim et de la douleur;
- une corrélation négative entre le pourcentage de l'expiration silencieuse et la perception de la faim et de la douleur;
- un constat concernant le cri dysphonique qui serait l'indice le plus fort de la perception de la faim et de la douleur.

1.6.3 La classification des cris

En parallèle des travaux menés sur les caractéristiques des cris normaux et pathologiques, plusieurs auteurs ont focalisé leurs investigations sur la reconnaissance et la classification des cris des bébés. Cette classification rentre dans le cadre de l'automatisation de la perception naturelle des cris ainsi que sa modélisation grâce aux techniques avancées de traitement de signal.

Les auteurs de (Orozco et García, 2003) ont présenté un système de reconnaissance automatique pour différencier les bébés sourds des bébés sans déficience auditive. Ils ont utilisé des caractéristiques acoustiques obtenues par prédiction linéaire et le réseau de neurones. Le test du système sur 31 bébés âgés de 6 mois a donné un meilleur résultat avec 91% de reconnaissance.

Dans (Cano Ortiz, Escobedo Beceiro et Ekkel, 2004), les auteurs ont testé la performance d'un classificateur utilisant le réseau RBF (Radial Basic Function) et le réseau de neurones artificiel (ANN) pour classer les cris des bébés en deux catégories (normale et anormale). La base de données utilisée pour les tests inclut 35 bébés. Parmi eux des bébés en bonne santé et des bébés atteints de bilirubinémie, hypoxémie et retard intra-utérin. Le taux de reconnaissance trouvé est de 85%.

Dans le travail proposé par (Cano et al., 2006), ils ont utilisé la combinaison de deux classificateurs pour identifier trois catégories de bébés (malade, modérément malade et en bonne santé). L'un des classificateurs se base sur le réseau de neurones artificiel (ANN) avec les coefficients MFCC et l'autre est un classificateur décisionnel qui considère 4 caractéristiques acoustiques (mélodie, décalage, « voicedness », « stridor »). L'identification des cris est faite à partir d'un bloc décisionnel qui fait la moyenne des scores obtenus des deux classificateurs. Leurs résultats indiquent une nette amélioration de la reconnaissance des cris avec la combinaison des deux classificateurs (90% de reconnaissance).

L'expérience réalisée par (Amaro-Camargo et Reyes-García, 2007) consistait à comparer l'efficacité de plusieurs méthodes de classification (support vector machine (SVM), réseaux de neurones (NN), J48, Random Forest, Naive Bayes) en utilisant les coefficients MFCC pour différencier trois catégories de bébés (normaux, sourds et asphyxiés). Ils trouvent de meilleurs résultats (91.86%) en utilisant conjointement les réseaux de neurones et « Random Forest ».

Dans (Jam et Sadjedi, 2009), les auteurs ont développé un classificateur pour différencier les cris de douleurs et les cris normaux (260 échantillons de cris dans chaque cas). Ils ont utilisé les caractéristiques basées sur les bancs de filtres à l'échelle de Mel et l'entropie combinée au réseau de neurones artificiel (ANN) pour l'apprentissage et la reconnaissance. Ils ont obtenu 88.5% de bonne reconnaissance.

Les auteurs, de (Feng-Ling et Xiang-Nian, 2010) ont testé la performance d'un système de reconnaissance des cris basé sur les coefficients MFCC et IMFCC combinés au DTW (dynamic time warping) pour l'apprentissage et la reconnaissance des signaux de cris et de parole (137 échantillons de cris et 100 échantillons de parole). Le meilleur résultat trouvé est de 86.2% de bonne reconnaissance.

1.7 Modèle électrique des cris

Le modèle source-filtre de la production de la parole est le plus utilisé dans les analyses acoustiques de production de la parole et des cris. Le modèle proposé se base sur l'interconnexion de sections de filtres qui représentent les différentes cavités vocales (*voir* Figure 1.4). Dans ce modèle, L'ensemble des organes impliqués dans la production des cris agit comme un filtre linéaire. Les sources qui agissent sur ce filtre peuvent être voisées ou/et non voisées. Dans le cas d'une source voisée, le signal provenant de la glotte est quasi périodique et dans le cas d'une source bruitée non voisée, le bruit provient d'une turbulence dans le conduit vocal.

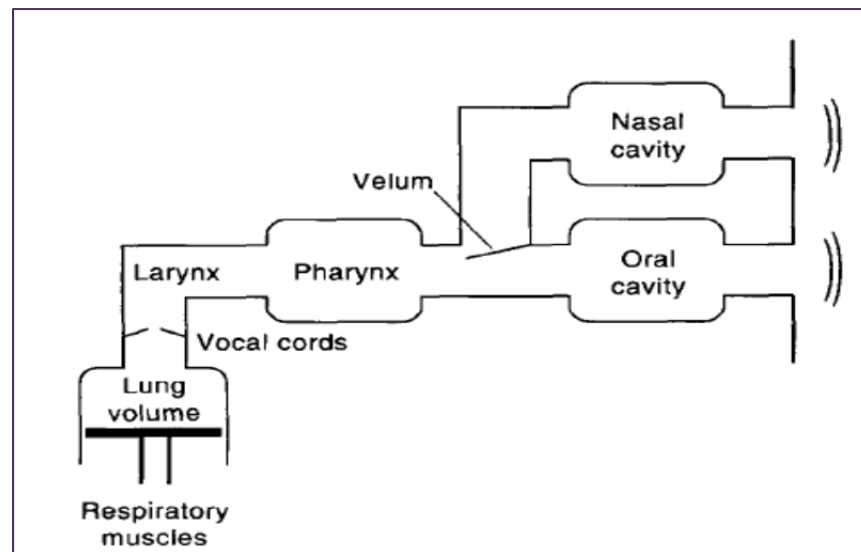


Figure 1.4 Le modèle de production de la parole
Tirée de Fort et al. (1996, p. 678)

Dans le cas d'une source voisée, le flux d'air causé par le changement de pression dans les muscles respiratoires est modulé périodiquement par les cordes vocales. Ce flux d'air passe ensuite à travers les cavités orales qui agissent comme un filtre variant dans le temps. Les cavités nasales agissent aussi sur les caractéristiques de ce filtre.

Dans le cas d'un son non voisé avec une source de bruit, le flux d'air est modulé par des turbulences dans le conduit vocal. Ces turbulences sont causées par un changement brusque de ce dernier. Une analyse séparée peut être appliquée pour l'excitation voisée et non voisée parce que la source et le modèle du conduit sont indépendants (Fort et al., 1996).

Donc, les sons voisés sont générés par une source d'excitation périodique $U(z)$ qui passe à travers un filtre glottal dont la fonction de transfert est $G(z)$. Le signal résultant de cette excitation glottale est ensuite utilisé pour exciter le système supra-glottal qui est composé de la fonction de transfert du conduit vocal $H(z)$ et la composante de radiation $R(z)$. Donc :

$$S(z) = U(z).G(z).H(z).R(z) \quad (1.1)$$

Dans le cas d'une excitation non voisée, $E(z)$ est généralement considéré comme un bruit blanc. Dans ce cas :

$$S(z) = E(z).H(z).R(z) \quad (1.2)$$

Donc, la sortie du système de production de la parole est le résultat du filtrage d'une excitation appropriée par plusieurs filtres linéaires séparés (*voir* équation 1.1 et 1.2).

Étant donné que la production de la parole est caractérisée par un changement de la forme du conduit vocal, le modèle le plus réaliste est considéré comme une concaténation d'une série de N tubes acoustique avec un diamètre et une longueur qui varient. Chacun d'eux est représenté par la fonction de transfert tous pôles suivante :

$$H(z) = \frac{1}{[1 + \sum_{k=1}^P a_k z^{-k}]} \quad (1.3)$$

a_k : Coefficients de prédiction.

P : nombre de pôles réels ou complexes.

Les pôles de $H(z)$, définissent les fréquences de résonance (RFs) du model du N-tube.

1.8 Prétraitement des signaux de cris

Il est nécessaire avant de faire toute analyse pour obtenir les caractéristiques acoustiques et les paramètres du signal de parole ou du cri, de procéder à certaines opérations sur les fichiers audio. Dans plusieurs types de paramétrisation, il est nécessaire en premier lieu de segmenter les signaux pour trouver le point de départ et de terminaison des segments utiles à l'analyse acoustique des cris. Dans la majorité des études sur les signaux des cris des bébés, la segmentation du son enregistré est faite manuellement parce qu'elle fournit une meilleure

précision. Néanmoins, il existe plusieurs méthodes qui permettent la détection d'activité de la voix. Parmi ces méthodes, nous retrouvons :

- la méthode STE (Short Time Energy) qui identifie les segments utiles par un rapport signal sur bruit élevé (Várallyay, 2006). L'inconvénient de cette méthode est qu'elle ne détecte pas les segments à basse amplitude (Várallyay Jr, Illényi et Benyó, 2009);
- la méthode ZCR (Zero Crossing Rate) qui se base sur le calcul du changement de signe des échantillons du signal. Le nombre du ZCR est élevé dans les segments non voisés par rapport aux segments voisés (Várallyay, 2006). L'inconvénient de cette méthode c'est qu'elle traite la plupart des sons respiratoires et les bruits comme des segments utiles (Várallyay Jr, Illényi et Benyó, 2009).

Du fait de la quasi-stationnarité du signal de parole ou du cri sur des courtes périodes, ce signal est analysé sur des segments découpés par une fenêtre de pondération avec un certain taux de recouvrement (30% à 50%) pour éviter le repliement du spectre ainsi que l'effet des lobes aux extrémités (effet de bord). Dans le cas d'un signal échantillonné à 44.1Khz, une fenêtre d'analyse de $N=1024$ points correspond à une longueur de 23.22ms. La fenêtre de pondération la plus utilisée est celle de Hamming. Elle est définie par la fonction suivante:

$$w(n) = 0.54 + 0.46 \cos(2\pi n / N) \quad - N/2 \leq n < N/2 \quad (1.4)$$

Dans la plupart des techniques de paramétrisation, il est nécessaire d'appliquer un filtre de préaccentuation au signal de parole pour éliminer efficacement les contributions spectrales du larynx et des lèvres. Ce filtrage permet d'effectuer l'analyse seulement sur les paramètres correspondants au conduit vocal. La fonction de transfert de ce filtre est donnée par :

$$H(z) = 1 - 0.97 z^{-1} \quad (1.5)$$

1.9 Analyse par prédiction linéaire

Le signal de cris ou de parole présente une corrélation à court terme induite principalement par les cavités buccales et aussi une corrélation à long terme induite par la périodicité du signal d'excitation. La corrélation à court terme est traduite dans le spectre par la structure des formants (l'enveloppe du signal) et la corrélation à long terme est traduite dans le spectre par une structure fine en peigne dite harmonique (Fort et al., 1996).

L'analyse par prédiction linéaire ou analyse LPC (Linear Predictive Coding) utilise donc la propriété de la corrélation à court terme qui conduit à une dépendance des échantillons en fonction des précédents. Le signal d'excitation est alors remplacé par un train d'impulsions périodiques pour les sons voisés ou bruit blanc pour les sons non voisés (Oppenheim V. Alan 1975) et (Atal, 1974).

La valeur prédite du $n^{\text{ième}}$ échantillon du cri est alors :

$$\hat{s}(n) = \sum_{k=1}^P \hat{a}_k s(n-k) \quad (1.6)$$

La fonction de transfert numérique du filtre linéaire tout pôle qui représente le conduit vocal, la radiation et le flux glottal est :

$$H(z) = \frac{\sigma}{A(z)} \quad (1.7)$$

Où le polynôme $A(z)$ est noté:

$$A(z) = 1 + \sum_{k=1}^P a_k z^{-k} \quad (1.8)$$

Donc les P coefficient a_k , le facteur d'échelle et les caractéristiques de l'excitation (telle que la période de F_0) fournissent l'information nécessaire pour la synthèse du signal de parole pendant le temps où le conduit vocal peut être considéré stationnaire. Le signal de parole sera donc décomposé en une séquence de tranches d'échantillons $s(n)$, appelées segments

acoustiques, dont la durée varie de 10 à 30 ms. Le passage de l'excitation $u(n)$ dans le filtre $H(z)$ produit le signal de parole $S(z)$ dont la fonction est donnée par :

$$S(z) = \frac{\sigma}{A(z)} \cdot U(z) \quad (1.9)$$

Avec :

$$u(n) = \sum_k \delta(n-k.P) \quad (1.10)$$

$H(z)$ est appelé filtre de synthèse. Dans ce modèle autorégressif, encore appelé modèle LP, la sortie du signal à l'instant n est fournie par la récurrence suivante :

$$s(n) = \sum_{k=1}^P a_k s(n-k) + \sigma \cdot u(n) \quad (1.11)$$

$u(n)$ est le $n^{\text{ième}}$ échantillon de l'excitation.

Ce modèle est appelé autorégressif, parce qu'il correspond dans le domaine temporel à une régression linéaire où chaque échantillon est obtenu en ajoutant un terme d'excitation à une prédiction obtenue par combinaison linéaire de P échantillons précédents. La réponse impulsionnelle de ce modèle comporte des fréquences de résonances représentées par des pics dans le spectre du signal de sortie.

Pour estimer les paramètres du modèle autorégressif, nous disposons uniquement de l'observation du signal $s(n)$ et non de son excitation. Dès lors, ce signal peut être prédit par la récurrence suivante (équation. 1.6). Dans ce cas, l'erreur commise appelée erreur de prédiction ou signal résiduel est donnée par :

$$e(n) = s(n) - \hat{s}(n) = \sigma \cdot u(n) \quad (1.12)$$

L'estimation $\hat{a}_k$ se base sur la minimisation de la variance de l'erreur de prédiction qui est définie par l'équation suivante :

$$\sigma^2 = E^2 = \sum_n (s(n) - \hat{s}(n))^2 = \sum_{i,j} a_i \cdot a_j \cdot r(i-j) \quad (1.13)$$

Ce critère de minimisation donne lieu à la résolution du système d'équation linéaire suivant :

$$\sum_{j=1}^P r(i-j) \cdot a_j = -r(i), \forall i = 1, 2, \dots, P \quad (1.14)$$

La fonction d'autocorrélation r du signal est définie par :

$$r(i) = r(-i) = \sum_n s(n) \cdot s(n+i) \quad (1.15)$$

L'algorithme de Levinson-Durbin permet de résoudre efficacement l'équation (1.14) par une récurrence sur l'ordre de prédiction. Les relations de récurrence sont les suivantes :

$$E(0) = r(0)$$

$$k_i = \frac{r(i) + \sum_{m=1}^{i-1} a_m^{(i-1)} \cdot r(i-m)}{E(i-1)}, i = 1, 2, \dots, P \quad (1.16)$$

$$a_j^{(i)} = a_j^{(i-1)} + k_i a_{j-1}^{(i-1)}, j = 1, 2, \dots, i-1$$

$$E(i) = (1 - k_i^2) \cdot E(i-1)$$

Les paramètres $a_j^{(i)}$, $j = 1, \dots, i$, sont les coefficients d'un filtre du i ème ordre. Ainsi, les coefficients du filtre d'ordre P sont donnés par :

$$a_k = a_k^P, k = 1, 2, \dots, P \quad (1.17)$$

Les coefficients k_i sont appelés coefficients de réflexions ou coefficients PARCOR (PARTIAL CORrelation).

Les pôles du filtre de synthèse sont les zéros du filtre inverse $A(z)$. La présence des zéros proches du cercle unité correspond à des sommets dans le spectre à court terme du signal de parole.

$$\text{PSD}(f) = \frac{1}{\left| 1 + \sum_{k=1}^P a_k e^{-2j\pi k T} \right|^2} \quad (1.18)$$

1.10 Analyse cepstrale

1.10.1 Coefficients cepstraux

La modélisation de la production de la parole implique une convolution temporelle de la réponse impulsionnelle du conduit vocal $h(n)$ avec la fonction d'excitation $u(n)$.

$$s(n) = u(n) * h(n) \quad (1.19)$$

Cette convolution devient une multiplication dans le domaine fréquentiel.

$$S(e^{j\omega}) = U(e^{j\omega}) H(e^{j\omega}) \quad (1.20)$$

Cependant, l'application de la fonction logarithme dans le domaine des fréquences permet d'effectuer la séparation de $h(n)$ et de $u(n)$. Étant donné que l'oreille humaine est pratiquement insensible à la phase relative entre deux composantes sonores, nous pouvons nous limiter au module de la transformée.

$$\log(|S(e^{j\omega})|) = \log|U(e^{j\omega})| + \log|H(e^{j\omega})| \quad (1.21)$$

Les coefficients cepstraux sont alors donnés par la transformée de Fourier inverse où l'indice k correspond aux raies spectrales (Bogert B.P, 1963) :

$$cc(k) = \text{DFT}^{-1} \log |U(e^{j\omega})| + \text{DFT}^{-1} \log |H(e^{j\omega})| \quad (1.22)$$

Le nombre de coefficients cepstraux calculés détermine le niveau de lissage de l'enveloppe spectrale estimée. Les coefficients cepstraux d'ordre faible sont très utilisés en reconnaissance automatique du locuteur (RAL). Ils caractérisent un trait anatomique de l'individu, principalement le conduit vocal.

À partir du cepstre, il est possible aussi de définir F0 de la source $u(n)$ en détectant les pics périodiques (harmoniques) au-delà d'un certain nombre N de coefficients. En effet, les N premiers points du cepstre contiennent l'information la plus pertinente sur le spectre et permettent d'obtenir un spectre lissé, débarrassé des harmoniques dues à la contribution de la source.

1.10.2 Coefficients MFCC

Les coefficients MFCC (Mel-scaled Frequency Cepstral Coefficients) représentent les traits les plus utilisés dans les systèmes les plus récents de reconnaissance du locuteur. Leur principe de calcul est issu des recherches psychoacoustiques sur la perception des différentes bandes de fréquences par l'oreille humaine. Le calcul des coefficients MFCC est en effet basé sur une échelle de Mel qui est linéaire pour les fréquences inférieures à 1000 Hz et logarithmique pour les fréquences supérieures. L'échelle de Mel redistribue les fréquences en fonction de la fréquence perçue (Stevens et Volkmann, 1940). Elle peut être approximée par la formule de conversion suivante :

$$f_{mel} = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (1.23)$$

Les cinq étapes de l'extraction des coefficients MFCC sont représentées sur la Figure 1.5.

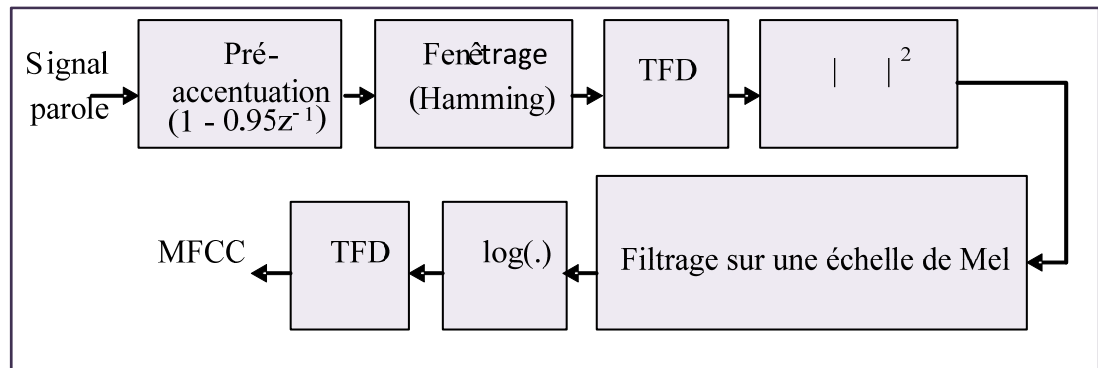


Figure 1.5 Les étapes d'extractions des coefficients MFCC

La procédure de calcul pas à pas des MFCC est la suivante :

- découpage du signal en trames chevauchées pour assurer la stationnarité et éviter les transitions brusques de trame en trame;
- préaccentuation qui permet de renforcer la contribution des hautes fréquences avec un filtre passe-haut de la forme $(1-0.9z^{-1})$;
- fenêtrage de Hamming qui permet d'assurer la continuité aux bords;
- calcul de la TFD sur chaque trame;
- filtrage par un banc de filtres triangulaires répartis le long de l'échelle de Mel;
- calcul du logarithme du module de l'énergie en sortie du banc de filtres;
- application de la Transformée en Cosinus Discrète inverse (qui joue le rôle d'une TFD inverse).

Seuls les premiers coefficients sont conservés. Par exemple, dans la reconnaissance de la parole les 12 premiers coefficients sont utilisés. Le premier coefficient c_0 représente l'énergie, mais peut être remplacé par le log de l'énergie moyenne des échantillons.

1.11 Analyse par ondelette

1.11.1 Transformée en ondelette discrète (DWT)

La transformation en ondelettes permet de caractériser un signal dans les domaines temporels et fréquentiels. Elle décompose le signal d'entrée en une série de fonctions d'ondelettes $\psi_{a,b}(t)$ qui dérivent d'une fonction mère $\psi(t)$ donnée par des opérations de dilatation et de translation (Meyer et al, 1987).

Les ondelettes sont des fonctions élémentaires très particulières; ce sont les vibrations les plus courtes et les plus élémentaires que l'on peut envisager (voir Figure 1.6). Les coefficients d'ondelette $C_{a,b}$ contiennent des informations concernant le signal $s(t)$ étudié à différentes échelles.

$$C_{a,b} = \sum_t s(t) \Psi_{a,b}(t) \quad (1.24)$$

$$\Psi_{a,b}(t) = n \Psi\left(\frac{t-b}{a}\right) \quad (1.25)$$

n : Facteur de normalisation (généralement $n = 1/\sqrt{a}$).

b : C'est le facteur de translation en temps.

a : C'est le facteur de dilatation en fréquence.

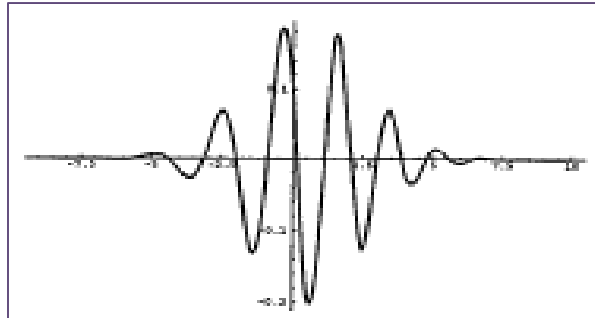


Figure 1.6 Exemple d'ondelette de Daubechies
Tirée de Moreau (1995, p. 29)

L'auteur de (Mallat, 1989) définit la transformée en ondelette discrète (DWT) comme une séquence d'application de filtres (*voir* Figure 1.7). Le signal noté S_0 est décomposé en i étapes en coefficients d'ondelettes C_i et en restes S_i . Les C_i sont calculés par application du filtre passe-haut h (*voir* équation 1.26). Le signal est alors filtré par le filtre passe-bas g pour calculer les restes successifs S_i (*voir* équation 1.27).

$$C_{i+1}[p] = S_i * h[2p] \quad (1.26)$$

$$S_{i+1}[p] = S_i * g[2p] \quad (1.27)$$

Les coefficients et les restes sont sous-échantillonnés en enlevant un point sur deux. Le signal ainsi est représenté avec un minimum de coefficients nécessaires.

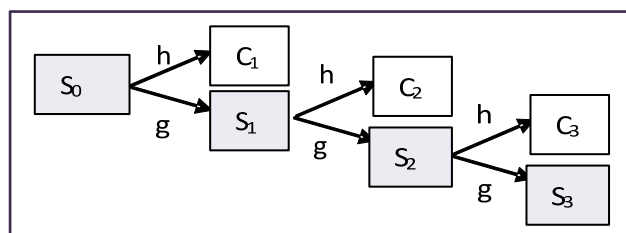


Figure 1.7 Décomposition discrète par ondelettes orthogonales (DWT)
Transformée en paquet d'ondelettes

La transformée en paquet d'ondelette (WPT) a l'avantage de fournir une décomposition complète niveau par niveau du signal étudié où chaque sous-bande (appelé Père) d'un niveau i engendre deux sous-bandes fils. Ces deux fils deviennent à leur tour des pères pour engendrer d'autres fils. Nous obtenons donc une décomposition multiéchelle du signal de départ en séparant à chaque niveau de résolution les basses fréquences (approximations) et les hautes fréquences (détails) du signal. La décomposition du signal en paquet d'ondelettes peut être réalisée par un filtrage successif en utilisant des bancs de filtres (g et h précédemment définis) et un sous-échantillonnage. Donc pour avoir plusieurs choix de résolutions, les coefficients d'ondelettes obtenus (voir Figure 1.8) sont décomposés à nouveau en appliquant à ceux-ci les mêmes filtres ainsi qu'un sous-échantillonnage.

Cette décomposition est intéressante dans la mesure où l'on peut ne conserver que les paquets qui reflètent bien le signal initial pour le critère choisi. Donc, elle permet de choisir la meilleure base d'ondelettes.

Il existe des algorithmes, fondés sur le calcul de l'information (au sens de Shannon par exemple), qui permettent de déterminer l'arbre représentant le mieux le contenu fréquentiel d'un signal ou d'un groupe de signaux, ce qui permet de choisir, en fonction du signal, sa représentation optimale.

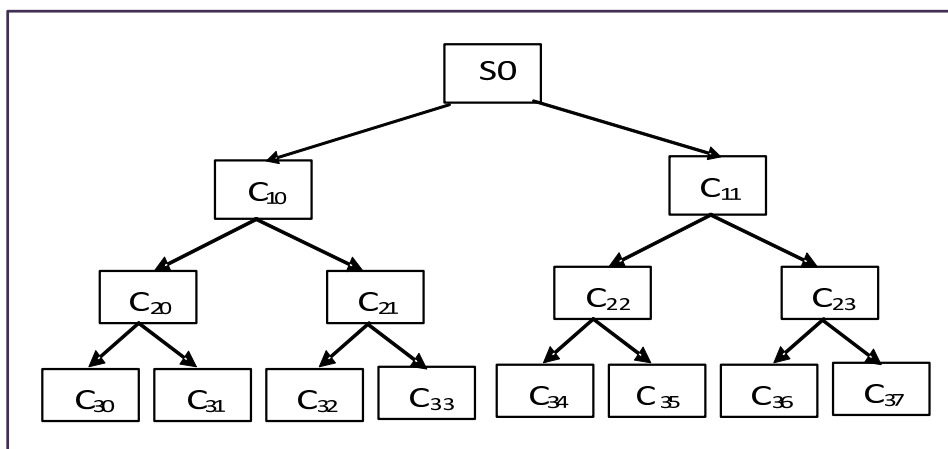


Figure 1.8 Décomposition par paquets d'ondelettes

1.12 Analyse en composante principale

L'Analyse en composantes principales (ACP) est une technique classique de l'analyse statistique multivariable (Jolliffe, 2005). Elle est utilisée pour la réduction de la dimension des données. Son principe consiste à remplacer une famille de variables par de nouvelles variables de variances maximales, non corrélées deux à deux et qui sont des combinaisons linéaires des variables d'origine. Ces nouvelles variables sont appelées composantes principales.

Les principales étapes de l'application de l'ACP sont montrées sur la Figure 1.9. Pour effectuer une réduction de la dimension par la méthode ACP, il faut effectuer une décomposition en valeur singulière (SVD) de la matrice de covariance S (Jackson, 2005).

Pour se faire, étant donné que la matrice est carrée, les valeurs propres λ ainsi que les vecteurs propres v de la matrice de covariance S de l'ensemble des données X sont obtenus par l'équation suivante :

$$(S - \lambda I)v = 0 \quad (1.28)$$

I : est la matrice identité.

Les valeurs propres et leurs vecteurs propres associés une fois obtenus sont classés par ordre décroissant pour former les colonnes de la matrice A_p , seulement les plus grandes valeurs propres sont retenues. La représentation des données par la méthode ACP consiste alors à projeter les données X sur le nouvel espace à l'aide de l'équation suivante :

$$X' = A_p^T X \quad (1.29)$$

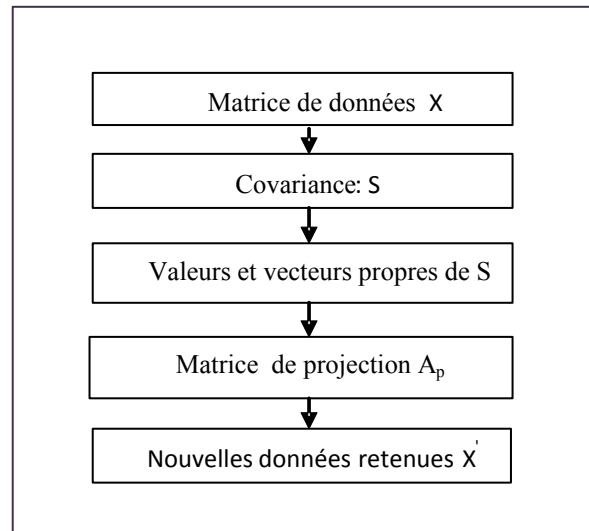


Figure 1.9 Principales étapes de l'application de l'ACP

1.13 Techniques d'extraction de la fréquence fondamentale

Les principales méthodes proposées pour la mesure de F0 dans les signaux des cris sont : les coefficients cepstraux, SSM (Smoothed Spectrum Method), AMDF (Average Magnitude Difference Function) et le SIFT modifié (Simple Inverse Filtering Tracking).

Ces méthodes peuvent être classées en trois catégories :

1. Les méthodes qui utilisent les propriétés temporelles du signal, comme celle de la fonction d'autocorrélation et la fonction AMDF;
2. Les méthodes qui utilisent les propriétés spectrales, comme celle des coefficients cepstraux et la méthode SSM;
3. Les méthodes qui utilisent à la fois les propriétés temporelles et spectrales comme l'algorithme SIFT. (Le détail d'implémentation de ce dernier se trouve dans les chapitres qui suivent).

Les principales étapes de l'estimation de F0 dans la plupart des algorithmes sont :

- prétraitement qui consiste à pondérer le signal en trames entrelacées. Cette étape peut inclure aussi des techniques telles que la décimation et le filtrage;
- traitement qui consiste à extraire F0 selon l'algorithme utilisé;
- post traitement qui consiste à faire corriger les erreurs d'estimation par exemple le lissage.

Les études effectuées sur l'estimation de F0 dans les signaux des cris révèlent des problèmes d'identification incorrecte des sous-harmoniques comme une duplication d'une fréquence fondamentale dominante induisant une classification incorrecte du segment comme un segment hyperphonique. La sous-estimation de la fréquence fondamentale de l'hyperphonation peut arriver aussi en induisant une classification incorrecte du segment en tant que phonique (Lederman, 2010).

1.13.1 Méthode des coefficients cepstraux

L'analyse cepstrale est utilisée pour séparer la contribution du conduit vocal de celle des cordes vocales. Les valeurs cepstrales qui représentent l'excitation sont extraites à partir du RC (Real Cepstrum) en multipliant les fréquences de valeur faible par 0 et le reste par 1 (Manfredi et al., 2000). F0 est obtenue à partir de la périodicité du signal traité.

Le cri des bébés est principalement voisé. Il peut être classé en un cri phonique avec une fréquence fondamentale qui varie de 350Hz à 750Hz et un cri hyperphonique avec une fréquence fondamentale au-dessus de 1000Hz. Dans le cas du segment phonique, l'estimation de la fréquence fondamentale en utilisant les coefficients cepstraux n'est pas significativement différente de celle de la parole chez les adultes. Mais dans le cas du segment hyperphonique, l'élargissement de la structure des harmoniques rend difficile la séparation de la contribution spectrale de l'excitation de celle du conduit vocal. Par conséquent, les coefficients cepstraux peuvent échouer dans la dé-convolution de l'excitation de la source et du conduit vocal (Lederman, 2010).

Nous remarquons sur la Figure 1.10, que dans le cas d'une phonation il est facile de séparer la contribution spectrale de la source d'excitation et du conduit vocal. Mais dans les deux cas d'hyperphonation, nous visualisons clairement qu'il est difficile d'estimer la fréquence fondamentale et de distinguer la contribution du conduit vocal

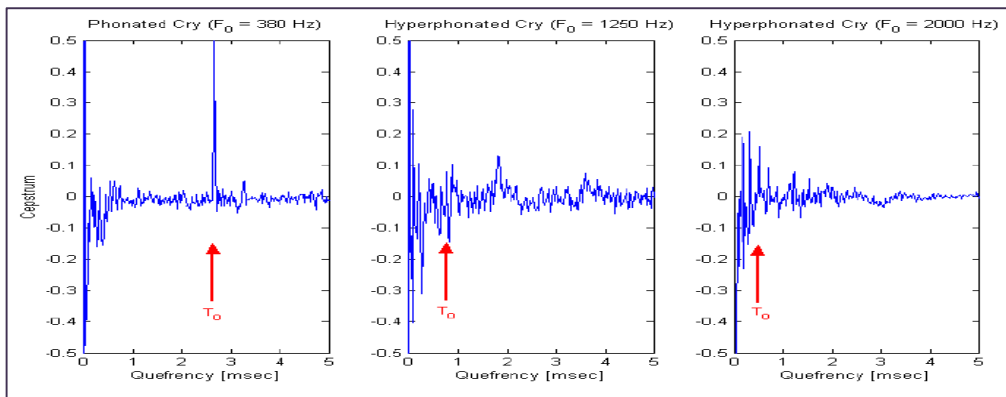


Figure 1.10 Séries Cepstraux de quelques segments de cris typiques
 À gauche cri phonique avec $F_0=380$ Hz. Au milieu cri hyperphonique avec $F_0= 1250$ Hz.
 À droite cri hyperphonique avec $F_0= 2000$ Hz
 Tirée de Lederman (2010, p. 1)

1.13.2 Méthode SSM

Várallyay (Várallyay, 2007) propose une méthode nommée (SSM) « Smoothed Spectrum Method » pour détecter F_0 . Cette méthode se base sur deux étapes consécutives. La première est le lissage du spectre et la détection de pics significatifs. La seconde étape consiste à calculer les plus probables valeurs de la fréquence fondamentale par des méthodes statistiques. Donc dans la première étape, le spectre de la FFT est lissé par la fonction symétrique de 'Kernel' (Figure 1.11) pour faire apparaître les pics et les composantes harmoniques et éliminer le bruit à large bande. Ce lissage est obtenu par la pondération d'addition dans une bande passante prédéterminée.

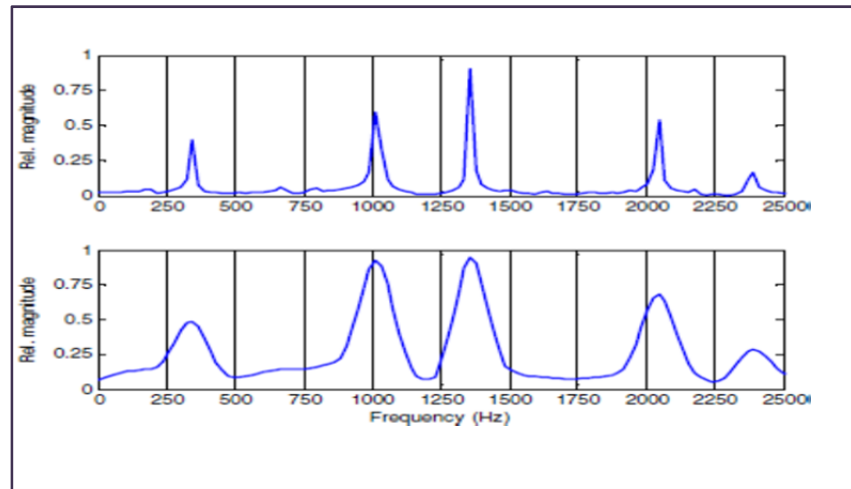


Figure 1.11 Le spectre original et sa version lissée
Tirée de Várallyay (2007, p. 3)

L'auteur prouve que la fréquence fondamentale peut être déterminée avec une petite erreur à travers la détection des harmoniques.

$$F_n / n = F_0 \pm B / 2n = F_0 \pm F_s / 2nN = F_0 \pm h 0.4 \quad (1.30)$$

F_n est la n ème sous harmonique de la fréquence fondamentale F_0 . B est la fréquence de résolution de spectre numérique qui est égale à la fréquence d'échantillonnage F_s divisée par la longueur de la fenêtre N .

Les pics d'harmoniques sont localisés dans le spectre à une distance $\pm B/2$ des valeurs réelles. La division des valeurs des fréquences correspondant aux pics sur leurs ordres correspondants, nous donne avec une marge d'erreur des valeurs autour de F_0 . Pour trouver l'ordre correspondant, la méthode SSM génère une séquence de tests par la combinaison de nombre d'ordres possible et calcule le ratio des pics obtenus par cette séquence. La séquence choisie est celle qui donne le plus petit des standards de déviation (STD) des ratios. La séquence une fois choisie, la valeur de la fréquence fondamentale est déterminée selon la fréquence du pic le plus élevé sur son ordre correspondant. Cependant, cette méthode n'a pas été testée sur des signaux réels, en plus elle s'avère moins efficace en présence de bruit à

large bande. Pour identifier les composantes du bruit, la méthode nécessite de générer plus de séquences avec plus de divisions.

1.13.3 La méthode AMDF

L'analyse AMDF (Average Magnitude Difference Function) est souvent utilisée pour détecter la fréquence fondamentale dans les signaux de paroles et dans les signaux de cris de bébés. Elle est intéressante parce qu'elle est facile à implémenter, mais elle est moins robuste en présence d'un signal bruité (Manfredi et al., 2000).

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

$$\text{AMDF}(m) = \frac{1}{N} \sum_{n=0}^{N-1-m} |x(n) - x(n+m)|, 0 \leq m \leq M_0 \quad (1.31)$$

M_0 est le nombre de points de la fonction AMDF.

Cette fonction présente un minimum aux multiples de la période fondamentale. Les principales étapes de cette méthode sont :

- filtrage passe-bande [50-300 Hz] dans le cas de la parole pour augmenter la précision;
- évaluation, pour chaque trame de longueur M , de L 'AMDF selon l'équation (1.31);
- sélection de la première valeur minimale absolue selon le seuil défini.

L'estimation de la fréquence fondamentale en utilisant l'algorithme AMDF génère deux types d'erreurs telles que F_0 estimée est le double ou la moitié de la valeur réelle. Ces deux erreurs sont dues à la baisse des pics d'énergie de la fonction AMDF (Abdullah-Al-Mamun, Sarker et Muhammad, 2009).

1.14 Techniques de classification des cris

Les systèmes de classification automatique des cris permettent de distinguer différents types de cris selon des classes prédéfinies (bébés sains/bébés malades ou sains/ pathologie1/ pathologie2/ ...).

Dans les systèmes supervisés, la classification se base sur trois phases : phase d'extraction des caractéristiques, phase d'apprentissage et phase de reconnaissance (voir Figure 1.12). Ces trois tâches demandent la contribution d'outils techniques aussi divers que puissants : traitement du signal, modèles mathématiques, algorithmiques, etc.

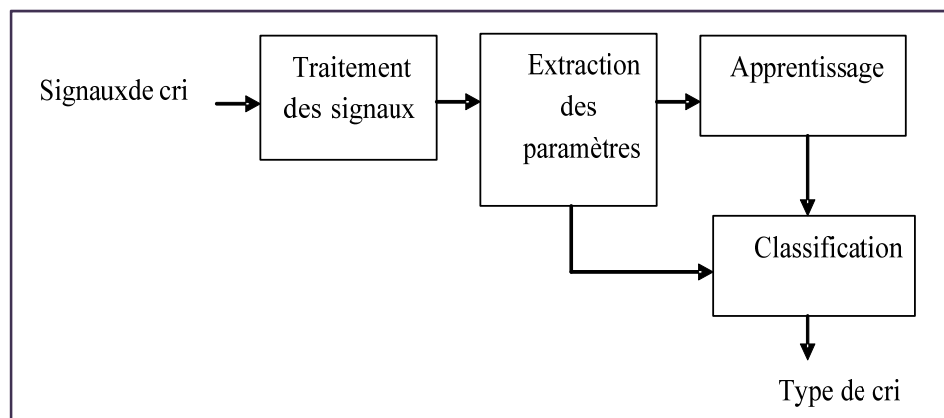


Figure 1.12 Reconnaissance automatique des cris des bébés

Les caractéristiques sélectionnées (MFCC, WPT, LPC,.....) dans la première étape sont représentées en vecteurs pour chaque signal de cri. Une optimisation de ce vecteur est effectuée par la suite en utilisant certaines techniques comme par exemple l'analyse en composantes principales (ACP), qui permet d'optimiser les performances de classification en ne retenant que les informations pertinentes et en supprimant la redondance.

Une fois la base de données constituée, avec une indication de la classe associée à chaque ensemble de coefficients, l'apprentissage des modèles selon les différentes classes se fait en deuxième étape.

La troisième étape de reconnaissance consiste à classer de nouveaux signaux de cris selon les modèles de classes entraînés dans l'étape d'apprentissage. Les classificateurs les plus utilisés dans les études les plus récentes sont : HMM (Lederman et al., 2008), PNN (Kheddache et Tadj, 2012, Hariharan et al., 2012), ANN (Jam et Sadjedi, 2009), GMM (Alaie et Tadj, 2012).

Nous expliquons dans ce qui suit le principe de fonctionnement des deux méthodes de classification HMM et GMM.

1.14.1 Hidden Markov Model (HMM)

Un HMM résulte de l'association de deux processus stochastiques : un processus interne, non observable $Q(t)$ et un processus observable $O(t)$, d'où le nom de modèle caché. La génération d'un modèle HMM approprié à une séquence d'observation donnée $V = \{v_1, v_2, \dots, v_M\}$, nécessite la spécification de trois mesures de probabilité liées au modèle $\lambda = (A, B, \pi)$.

A: Ensemble de probabilités de transition d'un état s_j à un état s_i

$$A = [a_{ij}], a_{ij} = P(q_t = s_j | q_{t-1} = s_i) \quad (1.32)$$

B : Ensemble de probabilités de production de l'observation v_k par un état s_i

$$B = [b_i(k)], b_i(k) = P(x_t = v_k | q_t = s_i) \quad (1.33)$$

π : Probabilité initiale.

$$\pi = [\pi_i], \pi_i = P(q_1 = s_i) \quad (1.34)$$

$Q=q_1, \dots, q_T$ une séquence d'états fixes de longueur T , correspondant à une séquence d'observation $O=O_1, \dots, O_T$. La probabilité de l'observation O pour une séquence d'états spécifiques Q est :

$$P(O|Q, \lambda) = \prod_{t=1}^T P(o_t | q_t, \lambda) = b_{q_1}(o_1) \times b_{q_2}(o_2) \dots b_{q_T}(o_T) \quad (1.35)$$

Et la probabilité des séquences d'états est:

$$P(Q|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \dots a_{q_{T-1} q_T} \quad (1.36)$$

La probabilité des observations selon le modèle est donnée par :

$$P(O|\lambda) = \sum_Q P(O|Q, \lambda) P(Q|\lambda) = \sum_{q_1 \dots q_T} \pi_{q_1} b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) \dots a_{q_{T-1} q_T} b_{q_T}(o_T) \quad (1.37)$$

Pour déterminer les paramètre du modèle λ , le MLE (maximum likelihood estimate) peut être utilisé (Rabiner, 1989).

1.14.2 Gaussian Mixture Model (GMM)

Une densité de probabilité d'un ensemble de données $X= \{x_1, x_2, \dots, x_M\}$, peut être approximée par une combinaison de gaussiennes (Kim, Kent et Weismer, 2011). Sous forme de l'équation suivante :

$$p(x_i | \lambda) = \sum_{j=1}^N p(x_i | N_j) \omega_j \quad (1.38)$$

sous la
contrainte:

$$\sum_{j=1}^N \omega_j = 1 \quad (1.39)$$

$$p(x_i | N_j) = \frac{1}{2\pi^{(\frac{D}{2})} \sqrt{|\Sigma_j|}} \exp\left(-\frac{\Delta_{ij}^2}{2}\right) \quad (1.40)$$

$$\Delta_{ij} = \sqrt{(x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j)}$$

D est la dimension du vecteur d'entraînement. Δ_{ij} est la distance de Mahalanobis, λ est le modèle GMM d'une classe de son de cri. Ce modèle est représenté par un ensemble de moyennes, de covariances et de poids comme suit :

$$\lambda = \{\mu_j, \Sigma_j, \omega_j\}, j = 1, \dots, N \quad (1.41)$$

N est le nombre de gaussiennes

Les paramètres du modèle λ sont estimés en utilisant la méthode du maximum de vraisemblance MLE (Maximum Likelihood Estimation) via l'algorithme EM (Expectation Maximization).

1.15 Conclusion

Ce chapitre permet principalement d'exposer au lecteur les concepts théoriques de la production du cri des bébés et les outils mathématiques nécessaires pour son analyse et son traitement numérique.

Dans ce chapitre, nous avons passé en revue les différentes études effectuées sur les cris des bébés telles que les travaux menés sur l'identification des caractéristiques acoustiques qui ont un lien avec certains cris pathologiques, les différentes méthodes de mesure de ces caractéristiques, ainsi que les travaux d'identification et de classification des cris pathologiques.

Il nous apparaît à partir de ces travaux que l'évaluation des cris des nouveau-nés est un outil de dépistage efficace pour détecter le statut neurologique du bébé. L'utilisation des caractéristiques acoustiques des cris pour le diagnostic précoce de différentes pathologies

offre un potentiel largement inexploité. Ainsi, nous avons exploré dans notre recherche certaines voies pour développer un outil d'analyse utilisant des algorithmes dédiés au conduit vocal du bébé permettant l'exploitation des caractéristiques des cris pour la détection de désordres chez le nouveau-né.

CHAPITRE 2

FREQUENTIAL CHARACTERIZATION OF HEALTHY AND PATHOLOGIC NEWBORNS CRIES

Yasmina Kheddache¹, Chakib Tadj¹

¹Electrical Engineering Department, École de technologie supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, H3C 1K3 Canada

Article publié dans le journal « American Journal of Biomedical Engineering»
Vol.3, No 6, Octobre 2013, pp.182-193

Résumé

Dans cet article, nous avons présenté de récentes avancées dans la caractérisation des cris des nouveau-nés malades et en bonne santé. Nous avons identifié et quantifié les caractéristiques acoustiques qui apparaissent les plus pertinentes pour distinguer les cris des bébés sains de ceux qui sont malades; comme la fréquence fondamentale (F0), l'irrégularité de F0 et la présence des modes hyperphoniques. Les résultats obtenus sont très encourageants, puisque les caractéristiques mesurées différencient réellement les cris pathologiques des cris de bébés en bonne santé.

Mots clés : cri des nouveau-nés, caractérisation des cris, fréquence fondamentale

Abstract

In this paper, we present recent developments in the characterization of healthy and pathologic cries of newborns. We have identified and quantified acoustic characteristics that appear the most relevant in differentiating between pathological and healthy cries; such as fundamental frequency (F0), irregularity of F0 and presence of hyperphonic modes. The results obtained are very encouraging, since the characteristics measured actually differentiate pathological cries from the cries of healthy babies.

Keywords: Newborns' Cry, Cry Characterization, Fundamental Frequency.

2.1 Introduction

Crying is the first sign of life at birth. It is an innate behavior that plays a fundamental role in the survival, health and development of the child. It is also the least explicit expression of distress or pain. Tests for evaluating the health of the newborn, such as Apgar score (Appearance, Pulse, Grimace, Activity, and Respiration) are now routine during birth. However, these tests do not incorporate quantitative measurement of acoustic characteristics of cries, providing only a qualitative summary based on observation. Our goal is therefore to utilise cries in order to improve monitoring of the first days of infant life (Wasz-Höckert, Michelsson et Lind, 1985; Lester et LaGasse, 2008; J. D. Newman, 1994).

Cry analysis allows identification of infant with medical syndromes when no symptoms are present (Lester et LaGasse, 2008; LaGasse, Neal et Lester, 2005). Early diagnosis before the onset of clinical symptoms will increase the likelihood of successful intervention before the illness has serious impact on the health of the infant. In the case of pathologies that are not detectable without in-depth examination and specialized tests, if medical treatment is begun soon enough, infant can heal completely.

Early diagnosis of various pathologies that can afflict newborn using the spontaneous cries during the first four weeks of life is crucial because from 4 weeks, infants acquired a voluntary control of these vocal tracts (Lester et LaGasse, 2008).

The cry itself contains a wealth of information. Its acoustic properties reflect the conditions that caused it to be produced (LaGasse, Neal et Lester, 2005). In order to evaluate infant cries, acoustic parameters must be measured quantitatively and deviations from “normality” must be interpreted.

The classification of the cries is a basic technique used to design a newborn’s PCIS (pathological cry identification system). It falls within the automation of a natural perception of cries and its modelling using advanced signal processing techniques. It contains two main parts namely a cry signal characterization and modelling.

There have been numerous fundamental studies related to this research area. The thrust of one of these lines of study has been to analyze infant cries in order to catalogue acoustic manifestations of pathological symptoms (Wasz-Höckert, Michelsson et Lind, 1985; Verduzco-Mendoza et al., 2009; Manfredi et al., 2009; Várallyay Jr et al., 2005). Our study is based on several hypotheses that these researchers have formulated. These hypotheses assert essentially that infant cries stem from physical and psychological status as well as both internal and external stimuli (LaGasse, Neal et Lester, 2005). Both temporal and acoustic characteristics contain biological alarms. By observing spectrograms and spectra of audio signals of cries, associations have been made with pathological conditions.

This visual approach is applied after the fact. To our best knowledge, there is no software tool that alerts health-care workers to the development of pathologies in newborns by direct monitoring of cries. Therefore, we propose to conduct a basic experimental research through the following:

- review the most pertinent characteristics of infant cries, based on scientific consensus

- identify the most promising characteristics and formally define them;
- associate these characteristics with pathologies of interest;
- automatic measurement of these characteristics;
- test the influence of use the studied characteristics in addition to other features in performance of PCIS.

One of the aims of this research is therefore to define as precisely as possible these characteristics, to create a tool for estimating them without human intervention and then quantify them for each pathological condition examined. The principal advantage of this approach is that once the tool is developed, it will evolve through updates as other pathologies are characterized. Our main contribution is therefore the quantification of acoustic characteristics associated with pathologies and ultimately provides a basis for alerting health-care workers to intervene. Of course the developed tool and system is not being proposed to replace health-care personnel. However, it could be highly appreciated for alerting personnel, especially in the hectic hospital setting. Once developed, the operation of the system is simple and does not require many resources. It can be set up quickly and at minimal cost.

This paper is divided into seven sections. In section 2.2, we present a succinct review of the scientific literature providing acoustic definitions of infant cries and their characteristics and modes. We describe in section 2.3 known associations between infant medical conditions and cry characteristics, with focus on healthy infants, premature births and certain health problems. Section 2.4 presents the database used in this study and discusses our adopted methodology for qualitative and quantitative characterization of cries with pathological implications and cries without such implications among the healthy infants studied. Section 2.5 is devoted to the analysis of the results obtained for an initial quantitative characterization of healthy and pathological cries. The next section presents experimental

validation of use the proposed characteristics concatenated with MFCC coefficients (Mel Frequency Cepstral Coefficients) in PNN classifier (Probabilistic Neural Network). We conclude this paper with a conclusion.

2.2 Acoustic definition of cries

Infant crying consists primarily of rhythmic alternation of phonation and breaths. The cry coincides with exhaling, the phonation being produced by the larynx (LaGasse, Neal et Lester, 2005). Each infant cry is acoustically unique. In most children, the fundamental frequency F0 varies from 250 Hz to 450 Hz, with a first formant at a frequency RF1 of 1100 Hz and a second formant at a frequency RF2 of 3100 Hz (LaGasse, Neal et Lester, 2005).

Lester et al. define three identifiable modes of cries due to vocal cord vibrations: a). Basic cry or phonation with F0 (350–750) Hz, b) High pitched cry F0 (750–1000Hz) or Hyperphonation F0 (1000–2000Hz) and c) Noisy, turbulent or dysphonic cries.

A certain terminology has been developed in various studies to describe the acoustic characteristics of cries (K. Michelson, 2007; Lester et LaGasse, 2008; Verduzco-Mendoza et al., 2009). The generally accepted terms are shown in Table2.1.

2.3 The characteristics of cries under certain medical conditions

2.3.1 The cries of healthy newborns

During the first months of life of an infant, cries are short with simple melodic forms and become longer and more complex with age (Varallyay et Benyo, 2007). At the age of one year, cries of infants born healthy are characterized by F0 varying from 400 to 600 H z with the average value of 450 Hz, often with a decreasing or increasing-decreasing melody shape (K. Michelson, 2007), with super-imposed symmetric harmonics and an average duration of 1–1.5s (Manfredi et al., 2009). The variation of F0 is regular. A shift is frequent at the

Table 2.1 Cry characteristics

Characteristic	Definition
Fundamental Frequency F0	The average vibratory frequency (in Hz) of the vocal folds
Hyperphonation	The average percentage of 25ms blocks having an F0 >1000 Hz.
Phonation	The average percentage of 25ms blocks having an F0 in the 350 –750 Hz range
F0 Irregularity	Sudden change in F0 > 100 Hz within 20ms
Dysphonation	The average percentage of 25ms blocks containing noise or aperiodic sound
Utterances	The number of vocal sounds produced by exhaling during the cry
Number of changes in cry mode	The number of blocks that change between phonation and dysphonation
Shift	Sudden change in F0 >100 Hz
Glide	Very fast increase or decrease in F0 of 600Hz or more during a time of 0.1s
First and second formant RF1, RF2	The average resonance frequencies produced by filtering the upper vocal passage
Duration of Inhalation	The interval in seconds between the first and second vocalization
Amplitude	The average energy in dB during a vocalization
Bi-phonation	Characterized by the presence of two F0
F0, RF1, RF2, amplitude variability	Inter-quartile spread of each parameter

beginning of the signal in the case of cries produced by pain and spontaneous cries, but rare during the first days of the newborn (Michelsson et al., 2002).

2.3.2 The cries of premature newborns

The differences in the characteristics of these cries are proportional to the number of weeks by which the birth was premature, the more premature, the higher fundamental frequency (Wasz-Höckert, Michelsson et Lind, 1985). Some cries are short and piercing, while others may be of duration and F0 similar to the cries of full-term newborns. This can be due to some disorders not detected at birth (Wasz-Höckert, Michelsson et Lind, 1985; K. Michelson, 2007).

2.3.3 Cries of newborns with various pathologies

The cries of a newborn with a pathological condition are persistent with little punctuation, reflecting high irritability and poor physiological stability. When a CNS disorder (central nervous system) is involved, the cry exhibits auditory abnormalities. However, its F0 is high and its melodic contour is irregular (Lester et LaGasse, 2008). In spite of differences in measurement procedures, all cry studies have shown that high F0 is an indicator of a neurological problem (LaGasse, Neal et Lester, 2005). Other markers also associated with neonatal risk include hyperphonic cries, very high-pitched cries, noisy or dysphonic cries as well as changing mode between phonation, dysphonic cries and variability of F0 and RF1 (Manfredi, Tocchioni et Bocchi, 2006).

Spectrographic analysis studies carried out on the cries of infants with disorders such as meningitis, hydrocephaly, congenital abnormalities, chromosomal aberrations or metabolic dysfunction indicate abnormal cry features (K. Michelson, 2007). In the syndrome known as cat cry, the cry is intense and monotone with a flat melodic contour, while in infants with a hypothyroid condition, the cries are of weak intensity. In newborns with herpetic encephalitis, the cry is significantly dysphonic and marked by increased noise concentration (LaGasse, Neal et Lester, 2005).

2.4 Methodology for qualitative and quantitative characterization

Our approach for characterization of healthy and pathologic cries is represented on the simplified blocks diagram illustrated in Figure 2.1. It consists of four steps: 1) choosing of the pathologies to be studied and recording of cries of newborns in order to build an experimental database, 2) qualitative characterization of the cries, based on identification of important characteristics in studied cries, 3) quantitative characterization based on estimation of selected characteristics and 4) establishing quantitative relationships between these characteristics and the pathologies studied.

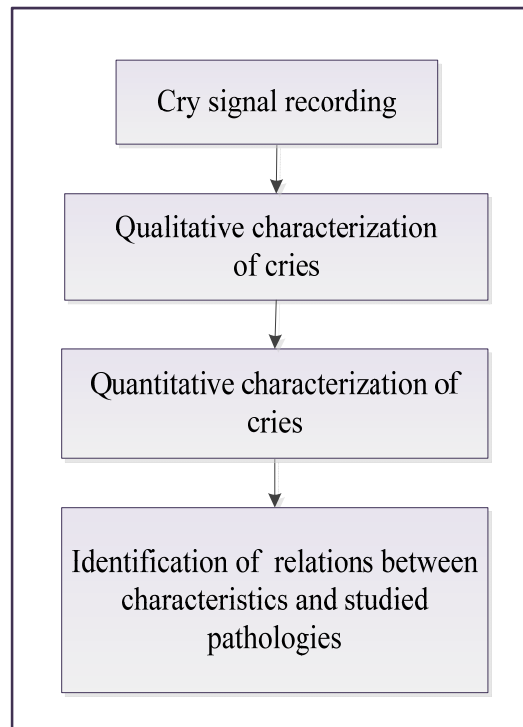


Figure 2.1 Newborn cries characterization

2.4.1 Cry database

The database used contains 3396 cry samples of 1s duration from 68 newborn babies aged 1 day to 4 weeks. 1774 cry samples from 31 healthy newborn (among them 764 are premature) and 1622 from 37 newborn who

present some diseases (among them 1079 are premature). The Table 2.2 shows the pathologies studied by categories of diseases. These cries were collected with the aid of medical collaborators of neonatology department at Saint-Justine Hospital in Montreal. Each infant is recorded two or three times, with at least one hour between each recording over a period of not more than ten days. The date and time of recording, gender identification, birth date, diagnosis, ethnicity are noted for each crying episode.

The recording of cries is done using a small recorder, placed at a distance of 10 cm of babies' mouth with a sampling rate of 44.1 kHz. The conditions in which the cries are recorded are: hunger, sampling blood and change of diapers. In order to confirm the health status of the infant or to determine that an undiagnosed health problem was in fact present during re-cording; we consult the follow-up examination conducted six months after recording of the cries.

In view of the spectacular increase in the numbers of pre-mature births recorded throughout much of the world, we decided to compare this category of cries to the cries of full-term newborns. Pre-mature newborns face major risks associated with organ functional immaturity. They are also more vulnerable to infection and face an increased risk of cerebral lesion when a significant jaundice is present. Early diagnosis of the various pathologies that can afflict this category of infant is crucial.

Table 2.2 The pathologies studied

Category	Pathology	Sample size
Heart Defect	Tetralogy of Fallot	53
	Thrombosis in the vena cava	77
Respiratory	RDS (Respiratory Distress Syndrome)	270
Infectious	Meningitis	115
	Peritonitis	20
Neurological	Asphyxia	190
	IUGR - asphyxia Intra-uterine Growth Retardation)	148
	IUGR- microcephaly	78
	Hyperbilirubinemia	250
Congenital Malformation	Gastroschisis	280
	Lingual frenum	141
Preterm newborn	Healthy	1010
Full-term newborn	Healthy	764

2.4.2 Qualitative characterization of cries

The principal aim of this part of the work was to identify the acoustic characteristics that best serve to characterize the cries studied and to establish qualitative and relationships between these characteristics and the various pathologies studied. This work is completed using PRAAT, a freeware program for the analysis and the reconstruction of acoustic speech signals (P. Boersma) .

The spectrographic analysis based on observation of pathological cry signals as well as cries of healthy infants to identify acoustic characteristics of these cries, allows us to establish the initial qualitative relationships between acoustic characteristics of cries and pathologies shown in Table 2.3. This analysis leads us to the conclusion that cries of infants with pathological conditions are indeed quite different from those of healthy infants.

Examples of spectrogram and estimated F0 for full-term, premature, healthy and sick newborns are shown in Figure 2.2. We note raising-falling pitch contour for healthy newborn's cries (preterm and full-term) and lower spectral intensity, height and irregular pitch contour in pathologic cries.

Table 2.3 Characteristics of cries examined by spectrographic analysis

Category	Pathology	Cry characteristics
Heart Defect	Tetralogy of Fallot	Shift
	Thrombosis in the vena cava	↑ Hyperphonation, ↑ F0 irregularity, ↑ dysphonation, ↑ shift
Respiratory	RDS	↑ Hyperphonation, ↑ F0 irregularity, ↑ dysphonation
Infectious	Meningitis	↑ F0, Hyperphonation , F0 irregularity
	Peritonitis	↑ Hyperphonation, ↑ F0 irregularity, ↑ dysphonation,
Neurological	IUGR - asphyxia	↑ Hyperphonation, ↑ F0 irregularity, ↑ dysphonation
	IUGR- microcephaly	Hyperphonation, F0 irregularity
	Hyperbilirubinemia	↑ Hyperphonation, ↑ F0 irregularity, dysphonation
	Asphyxia	↑ Hyperphonation, ↑ F0 irregularity , ↑ dysphonation
Congenital malformation	Gastroschisis	↑ F0, ↑ Hyperphonation , F0 irregularity, dysphonation
	Lingual frenum	↑ Hyperphonation, ↑ F0 irregularity, ↑ dysphonation
Healthy	-	↓ Hyperphonation, ↓ F0 irregularity, ↓ dysphonation

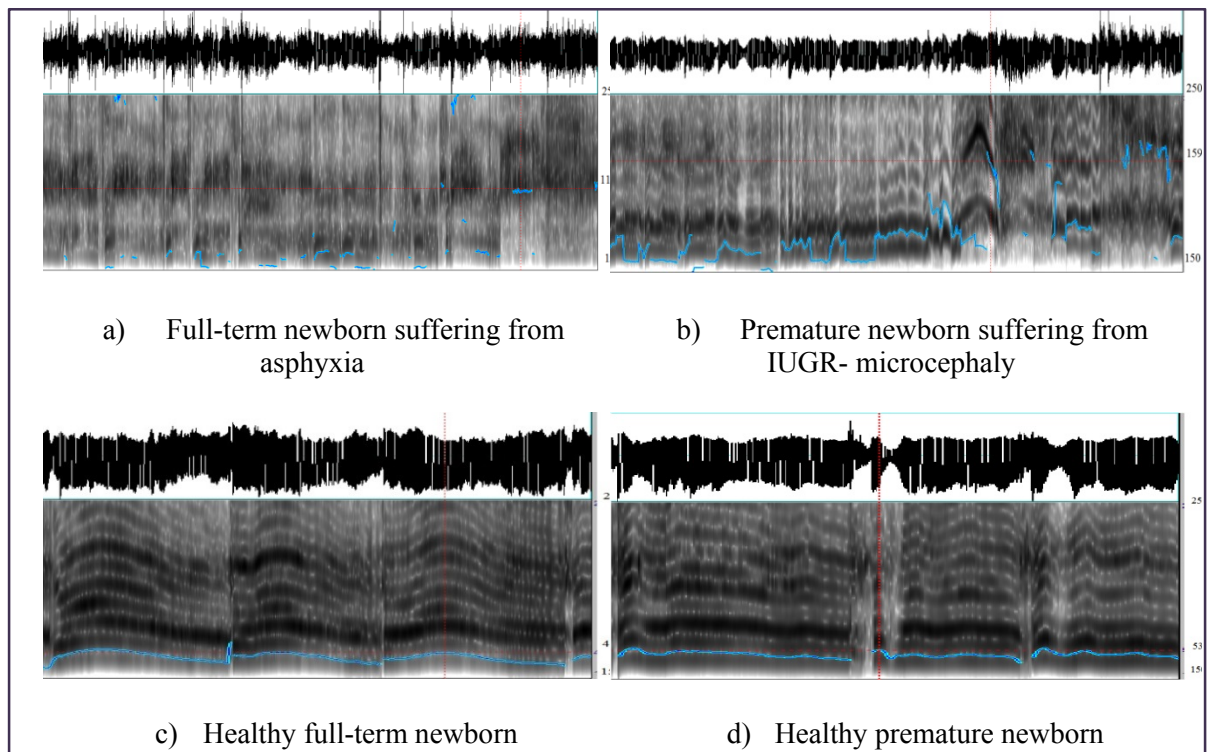


Figure 2.2 Waveform and spectrogram of healthy and pathologic cries

2.4.3 Quantitative characterization of cries

The quantitative characterization of cries contains an automated measurement of acoustic characteristics of cries that make major contributions to the differentiation between pathological cries and healthy cries.

In order to establish quantitative relationships between cry characteristics and the pathologies, we began by estimating or measuring certain acoustic characteristics cited in the literature in association with severe medical conditions, such as F0, the presence of hyperphonation and irregularity of F0.

The principal means used by researchers to analyze infant cries involve software initially dedicated to adult voice analysis. Since the adult and infant vocal passages differ in shape, these tools should be used with precaution (LaGasse, Neal et Lester, 2005). We developed

our own measuring tool using Matlab. This tool estimates F0 and percentage of hyperphonic segment as well as the F0 irregularity, as defined in Table 2.4.

The following approach was taken when estimating these characteristics:

- noise filtering and segmenting recordings into useful and non-useful portions;
- estimation of F0 in short segments typically of 20ms;
- identification of hyperphonic segments, as well as irregularity of F0 as defined in Table 2.4;
- calculation of average percentage of hyperphonic segments (AP_{hyp}) and irregularity of F0 (AP_{irr});

This has been performed: 1) for healthy and pathologic cries, 2) by category of pathologies and finally 3) by pathology and gestational age. We used the following formulas:

$$AP_{hyp} = \frac{N_{hyp}}{N_{total}} \quad (2.1)$$

$$AP_{irr} = \frac{N_{irr}}{N_{total}} \quad (2.2)$$

Where N_{total} , N_{hyp} and N_{irr} are the total number of segments, the total number of hyperphonic segments and the total number of F0 irregularity respectively.

- utilization of standard deviation for estimated characteristics by category of pathologies and application of ANOVA variance analysis for healthy and pathologic cries to carry out the analysis of the estimated characteristics by gestational age (Full-term, Preterm).

Table 2.4 Measured characteristics of cries

Characteristic	Definition
Fundamental frequency F0	Vibratory frequency (in Hz) of the vocal folds
Hyperphonation	Blocks of 20ms having an F0 >1000 Hz
F0 irregularity	Sudden change in F0 >100 Hz within 20ms

2.4.4 Fundamental frequency estimation

Since F0 is one of the most widely used characteristics for distinguishing cries and measurement of most other characteristics stems from it, precise measurement of F0 and its variations over time is viewed as an essential component of reliable information on the health status of newborns and its variations over time is viewed as an essential component of reliable information on the health status of newborns.

In this paper, the modified SIFT algorithm (Simple Inverse Filtering Tracking) is used for estimating the fundamental frequency. It was demonstrated that this algorithm includes the autocorrelation properties and the cepstral pitch analysis technique (Markel, 1972). In addition, the performance of this algorithm has been tested on a real newborn's cry database (Manfredi et al., 2009; Lederman, 2010).

The main steps of this algorithm are:

- division of the signal into overlapping frames of 20ms and n samples with 10ms recovering and multiplying each frame by hamming window;
- performing glottal inverse filtering to attenuate the influence of vocal tract;

- estimation of the autocorrelation sequence;
- performing a peak picking and decision algorithm in which the peak value is compared to voiced threshold;
- the fundamental frequency is estimated using: $T0 = 1/F0 = \arg\max_{\eta} \{r_i(\eta)\}$, where $r_i(\eta)$ is the autocorrelation sequence with $i=[1,..,n]$;
- smoothing of the result fundamental frequency contour using median filter.

Details about the SIFT algorithm can be found in (Markel, 1972; Lederman, 2010).

2.5 Results analysis

Examples of estimated F0 for full-term, premature, healthy and sick newborns are shown in Figure 2.3. Figure 2.3(a) shows the estimation of F0 for the cry of a healthy, full-term newborn. The frequency range is 400–500 Hz and no hyperphonic segments, shifts or gliding are noted. Figure 2.3(b) shows the estimated F0 for the cry of a full-term newborn suffering from undiagnosed vena cava thrombosis during the days following birth. The measured values characterize a pathological cry. The pathological condition was confirmed six months after birth. Variation of F0 in this case is irregular. We note the presence of long hyperphonic segments with gliding and shifts of F0 within the 700–2200 Hz range. Figure 2.3(c) shows the estimated F0 for the cry of a premature but otherwise apparently healthy newborn. We did not note the presence of hyperphonic segments. F0 varies somewhat more and with less regularity than in the case of the full-term healthy infant. Figure 2.3(d) shows the estimated F0 for the cry of a premature newborn suffering from Gastroschisis.

As in case (b), the measured values correspond to a cry with pathological implications. The presence of several hyperphonic segments is noted. The F0 variations are less regular than are those of the healthy premature infant cries.

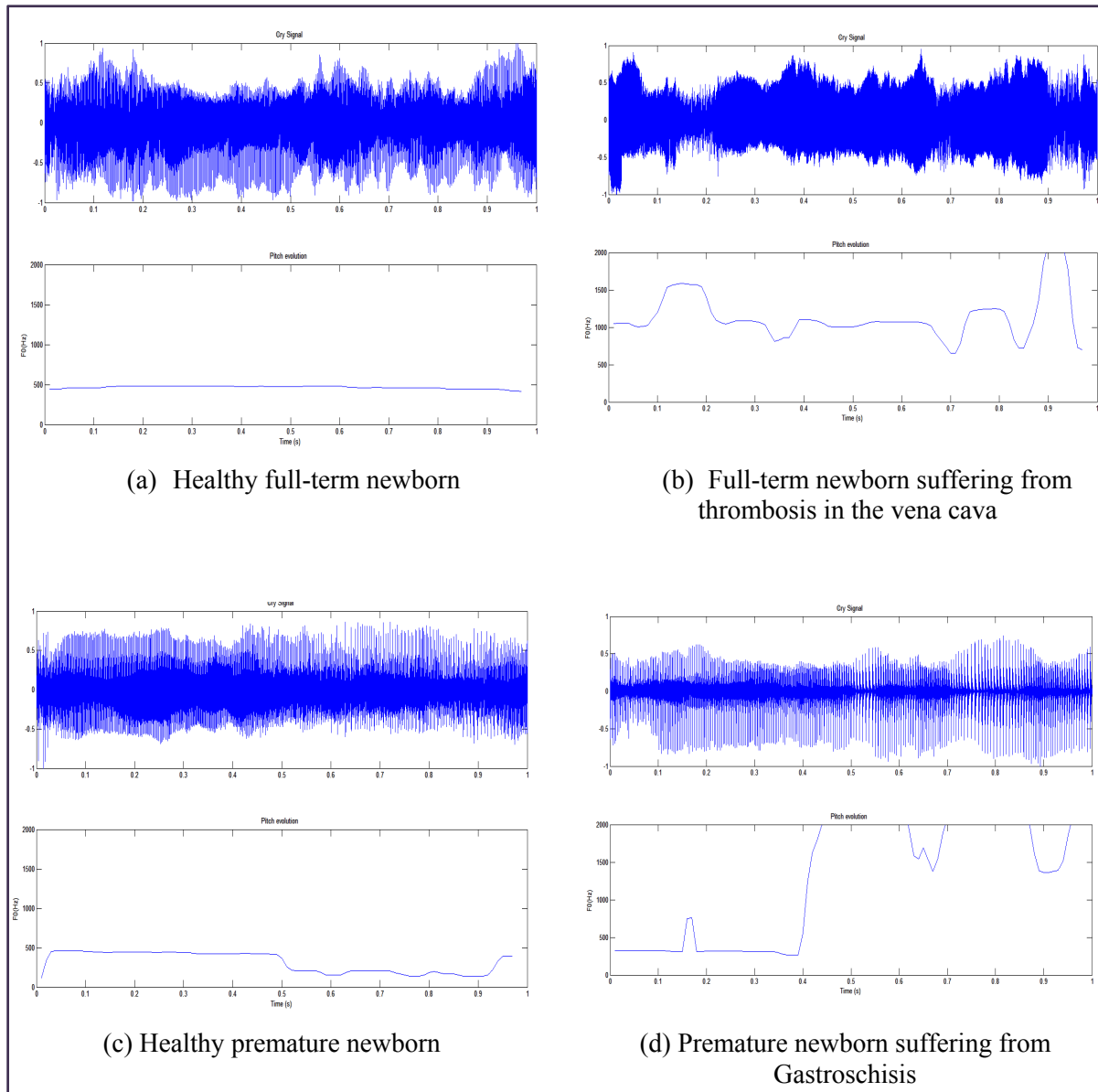


Figure 2.3 Estimation of F0 using the modified SIFT algorithm

The estimated percentages of hyperphonic segments as well as the irregularity of F0 are shown in Figure 2.4. The recordings used for this estimation are of premature infants (p) and full-term infants (t), including healthy infants and infants suffering from pathologies listed in Table 2.2. Figure 2.4(a) indicates that the estimated characteristics allow well an initial distinction between a healthy newborn and a sick newborn, because the average percentage

of the hyperphonation and the irregularity of F0 are clearly higher for sick babies compared to healthy ones.

These results show that the cries of healthy newborns (full-term or preterm) contain around 6.5% of hyperphonic segments and around 6.5% of irregularity of F0. On the other hand, the cries of sick newborns contain around 11% of hyperphonic segments and around 9% of irregularity of F0.

These preliminary results led us to estimate the characteristics by category of pathologies. The estimated characteristics according to the category of pathologies are presented in Figure 2.4(b). We notice that the average percentage of the hyperphonation and the irregularity of F0 are higher for all categories compared to that of healthy babies. Similarly, the highest average percentage of the hyperphonation and the irregularity of F0 are found in the category of the newborn with neurological problems.

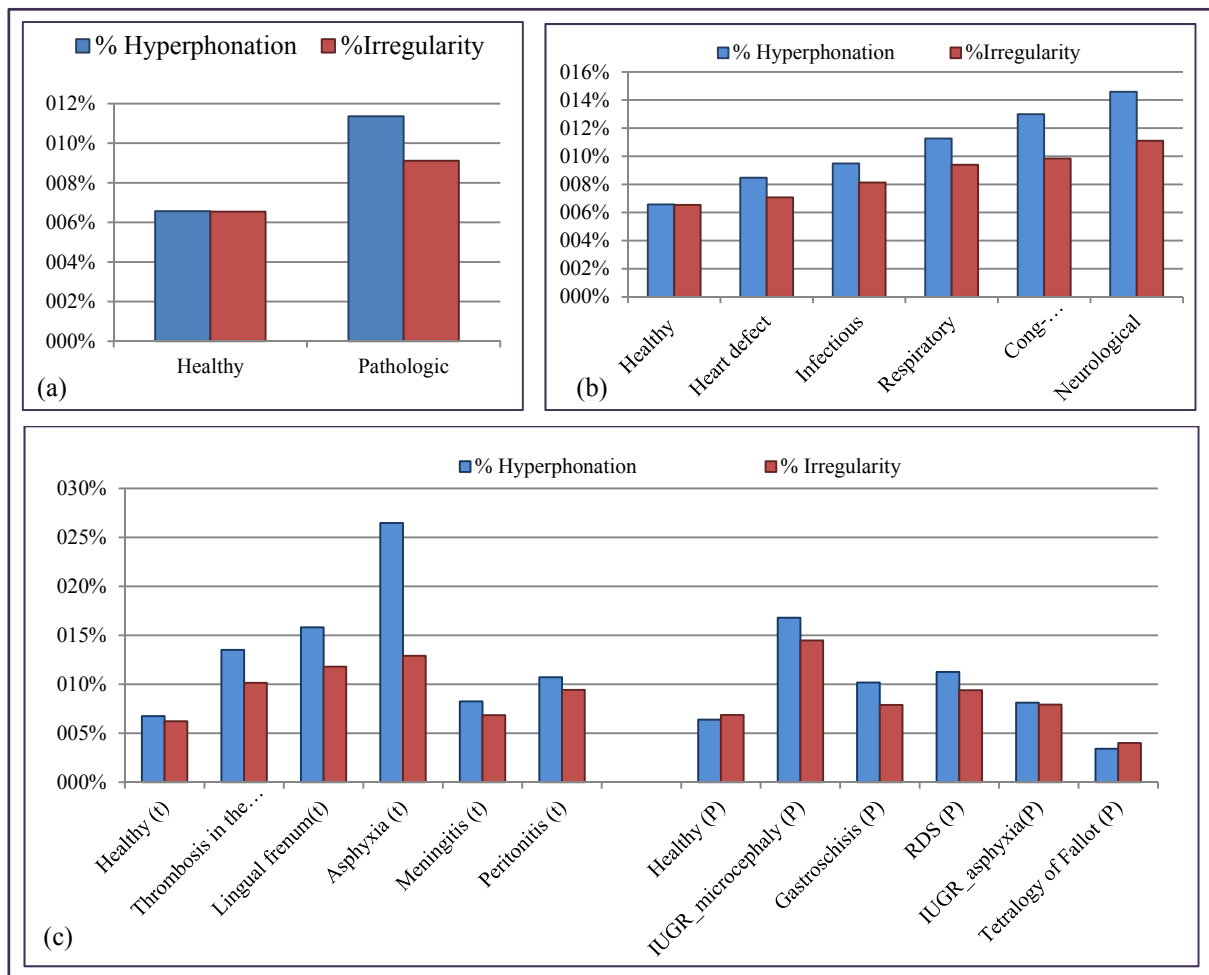


Figure 2.4 The average percentage of hyperphonic segments and irregularity of F0 by
a) Pathologic and healthy cries, b) Category of pathologies,
c) Pathologies and gestational age

As shown in Figure 2.5, the calculation of the standard deviation for estimated characteristics by category of pathologies shows a large dispersion of measured characteristics for Heart Defect and Neurological categories in both case of average percentage of hyperphonic segments and irregularity of F0.

The estimation of studied characteristics by pathologies and gestational age indicates a better variation of the newborns cries characteristics. This estimation is presented in Figure 2.4(c). It shows clearly that the average percentages of hyperphonic segments in the cry samples are similar for both healthy premature infants and healthy full-term newborns. It also shows that

the average percentages of irregularity of F0 are slightly more important for premature infants than healthy full-term newborns. This result is consistent with spectrographic studies of crying newborns (Lester et LaGasse, 2008; K. Michelson, 2007).

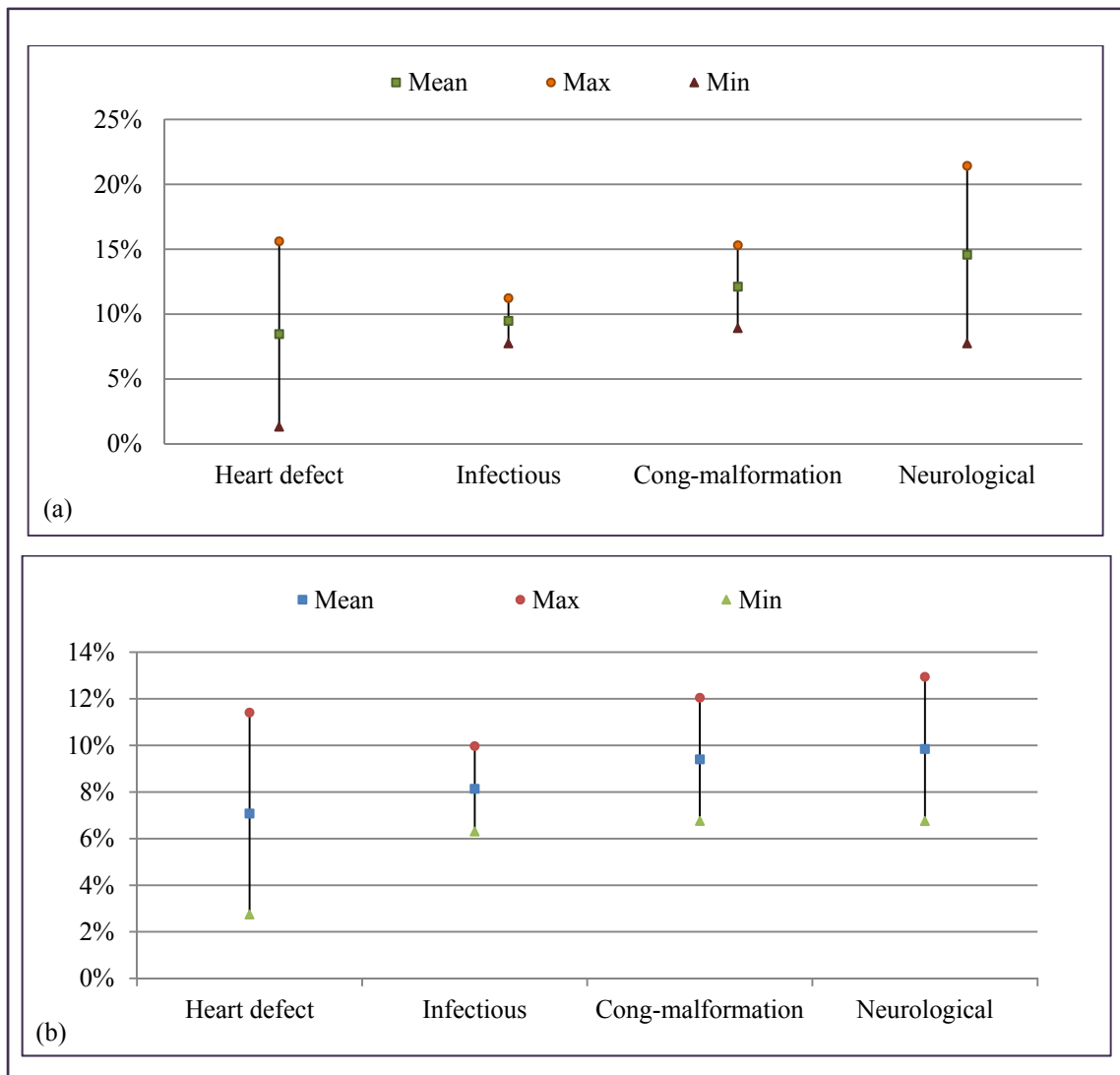


Figure 2.5 The mean percentage and standard deviation of a) hyperphonic segments
b) irregularity of F0 by category of pathologies

The results shown in Figure 2.4(c) also indicate that the average percentages of hyperphonic segments and irregularity in the cry samples from infants with a pathological condition are clearly higher compared to healthy newborns. These percentages vary from one pathological condition to another and are dependant to level of prematurity. We thus are inferring that the characteristics of the cries do not vary according to the categories of diseases but the pathology itself. For example, in the case of the category Heart Defect, where the percentages of the characteristics estimated for pathology Thrombosis in the Vena Cava is much higher compared to the percentages of the characteristics for pathology Tetralogy of Fallot.

The calculation of the standard deviation for estimated characteristics by pathologies is shown in Figure 2.6. The results indicate a large dispersion for measured hyperphonic segments in the case of Lingual frenum and Gastroschisis. They also indicate a best estimation of average percentage of hyperphonic segment in the case of healthy preterm newborn, thrombosis in the vena cava, asphyxia, RSD and Tetralogy of Fallot diseases.

The large dispersion for measured irregularity of F0 is found in asphyxia, IUGR – microcephaly, Gastroschisis and, IUGR–asphyxia diseases. The best estimation for this characteristic is found in full term and preterm newborn, Lingual frenum, RSD, Tetralogy of Fallot diseases.

With this study, we should also be able to demonstrate that some pathological conditions do not manifest themselves in cries and are therefore not detectable using the characteristics studied such as Tetralogy of Fallot. More characteristics may be necessary to better distinguish and/or detect these pathologies.

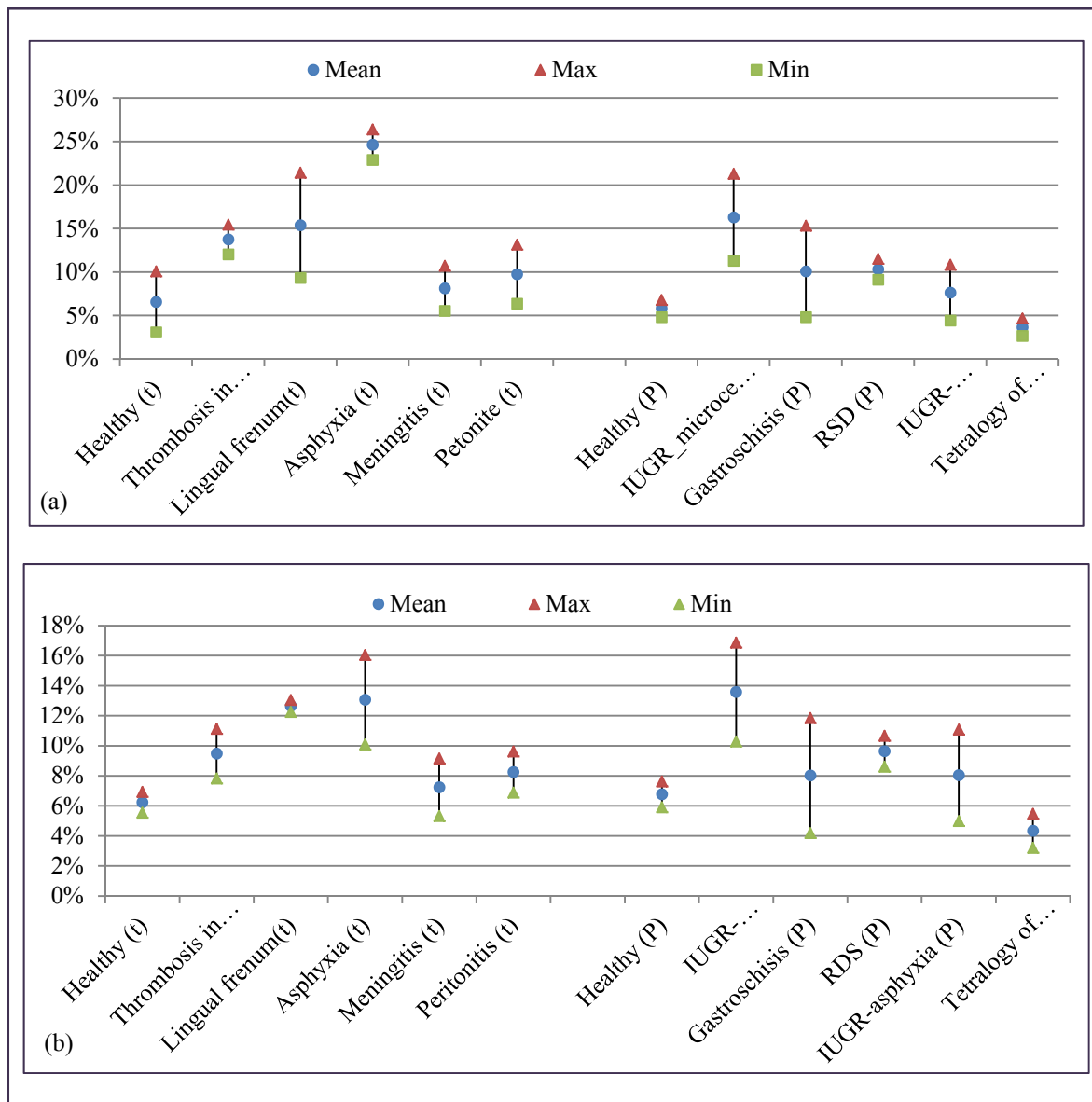


Figure 2.6 The mean percentage and standard deviation of a) hyperphonic segments
b) irregularity of F0 by pathologies

We performed ANOVA variance analysis for healthy and pathologic cries to compare behavior of the estimated characteristics (percentage of hyperphonic segments and irregularity of F0) with respect to gestational age (preterm, full term). We notice from these

results in Table 2.5, significant statistical differences ($F=7.27$, $p<0.05$) between average percentage of hyperphonic segment of healthy premature cries and full-term healthy cries. For all other cases no significant difference were found by gestational age.

Hence, the percentage of hyperphonic segments is dependent on gestational age for healthy cries. For pathologic cries, this characteristic does not depend on gestational age but it's depending on pathology itself. The percentage of irregularity of F0 does not depend on gestational age in both cases of healthy and pathologic cries.

Table 2.5 Analysis of variance (ANOVA) according to gestational

Characteristics	Healthy		Pathologic	
	F	<i>p</i>	F	<i>p</i>
% Hperphonation	7.27	0.0071	2.116	0.176
% Irregularity	1.425	0.2327	1.233	0.29

2.6 Experimental validation

The adopted methodology to test the influence of use the studied characteristics in addition to other features in performance of PCIS is represented on the simplified diagram blocks illustrated in Figure 2.7. Matlab is used for the development of all blocks of the system.

In this context, we used studied characteristics (AP_{hyp} , AP_{irr}) concatenated with the most used features in the most recent speaker recognition systems to characterize acoustic features of cries signals, namely MFCC (Mel Fre-quency Cepstral Coefficients) witch based on Fourier analysis and filter bank on a Mel scale. These characteristics once obtained are used as input of PNN classifier (Probabilistic Neural Network). PNN is generally used for classification problems in the medical domain (Sweeney, Musavi et Guidi, 1994; Kheddache et Tadj, 2012).

The use of the PNN classifier is motivated by its speed and simplicity of the training process (Kheddache et Tadj, 2012). It is obtained using function `newpnn ()` in Matlab.

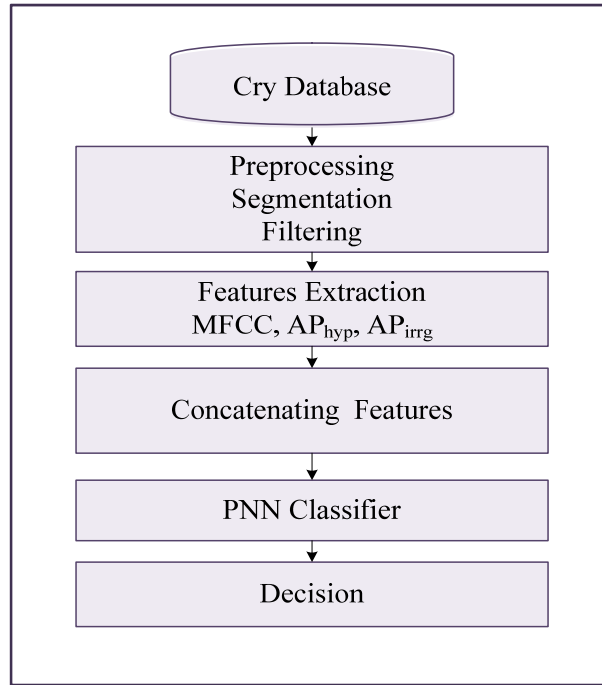


Figure 2.7 Experimental validation

2.6.1 Features Extraction

MFCC coefficients are extracted using 20ms interlaced frame with 10ms recovering, providing 99 windows for one second cry samples. We extracted 12 MFCC parameters for each 20ms of a cry signal. We obtained an MFCC matrix of 12 lines $\times$ 99 columns for each sample of 1second. Thereafter we calculate the average percentages of hyperphonic segments and irregularity for each sample of 1second according to the methodology explained in section 2.4. The two sets of characteristics once obtained are used as an input of the PNN classifier.

The structure of this input consists of converting each matrix of MFCC coefficients specific to a cry sample into only one vector of $12 \times 99 = 1188$ dimension for one second sample.

Thereafter we concatenate each vector with his corresponding percentages of hyperphonic segments and irregularity of F0.

2.6.2 Experimental results

To evaluate the efficiency of proposed PCIS using studied characteristics, two experiments were performed in case of full term and preterm newborns.

The first experiment consists of testing the proposed PCIS to separate pathologic cries from healthy ones, with or without studied characteristics (AP_{hyp} , AP_{irr}). The samples of cries are distributed in two classes, healthy cries and pathological cries, 120 samples for each class.

In the case of pathologic cries of full term newborns, we used the following pathologic cries: (20 samples of asphyxia, 20 samples of lingual frenum, 20 samples of meningitis and 60 samples of healthy cries) for training and (20 samples of samples peritonitis, 40 samples of thrombosis in the vena cava and 60 samples of healthy cries) for tests. In the case of pathologic preterm newborns cries, we used follow pathologic cries: (20 samples of IUGR – asphyxia, 20 samples of gastroschisis, 20 samples of Tetralogy of Fallot and 60 samples of healthy cries) for training and (40 samples of RDS, 20 samples of IUGR- microcephaly and 60 samples of healthy cries) for tests.

The second experiment consists of testing the proposed PCIS to separate cries by pathologies from healthy ones with or without studied characteristics (AP_{hyp} , AP_{irr}). The samples of cries are distributed in five classes, 70 samples for each class. In case of full term newborn: healthy cries, asphyxia cries, lingual frenum cries, meningitis cries thrombosis in in the vena cava cries are used. In the case of preterm newborns cries: healthy cries, IUGR–asphyxia cries, IUGR-microcephaly cries, gastroschisis cries, RDS cries are used. For the test of proposed PCIS, 1/5 of each subset is used for test and 4/5 for training with cross validation method.

It is important to note that the samples of healthy cries used for training are different from once used for testing for both experiment. The results of the correct identification rate (Overall accuracy) for all experiments are shown in Table 2.6 and Table 2.7. The results obtained from the experimental PCIS using studied characteristics concatenating to MFCC coefficients as input of PNN classifier, indicate that the use of AP_{hyp} and AP_{irrg} improve performance of a PCIS particularly for identification system by pathologies in second experiment.

Table 2.6 Results of first experiment

Gestational age	Features	Overall accuracy
Full term	MFCC	85.83%
	MFCC+ AP_{hyp} , AP_{irrg}	86.67%
Preterm	MFCC	64.17%
	MFCC+ AP_{hyp} , AP_{irrg}	65.00%

Table 2.7 Results of second experiment

Gestational age	Features	Overall accuracy
Full term	MFCC	72.86 %
	MFCC+ AP_{hyp} , AP_{irrg}	82.86%
Preterm	MFCC	65.71%
	MFCC+ AP_{hyp} , AP_{irrg}	77.14%

2.7 Conclusions

The acoustic characteristics of infant cries reflect CNS integrity. The potential for their use in the early diagnosis of various pathologies in newborns remains largely undeveloped. Our aim in this study was to analyse and characterize the healthy and pathologic cries of newborns. This characterization is important for the development of a robust system of early diagnosis for the identification of selected pathological conditions.

In this paper, we have presented the three principal steps of our research project. We have specified the methodology followed for characterizing the cry. We have also identified and quantified the acoustic characteristics that appear the most apt for betraying the pathological conditions such as percentage of hyperphonic modes (AP_{hyp}) and irregularity of F0 (AP_{irr}). And finally we have tested the performance of proposed PCIS using MFCC coefficients and studied characteristics (AP_{hyp} , AP_{irr}).

The results obtained are very encouraging and even exceed our expectations. They showed that these characteristics are depending on pathology itself not on gestational age. In fact, the use of AP_{hyp} , AP_{irr} improve performance of the proposed PCIS to separate cries by pathologies from healthy ones. We conclude that the studied characteristics are very effective for the development of a pathological cry detection system.

We also expect to improve the results by using other classifier and other acoustic characteristics in addition to AP_{hyp} , AP_{irr} .

For further characterization of the pathological cries studied, we shall proceed with measurement and quantification of other relevant characteristics. In order to achieve a reliable evaluation of characterization performances, we are still recording a larger database with a greater variety of pathologies and more subjects for each pathology.

Acknowledgements

We would like to thank Dr. Barrington and members of neonatology group of Saint-Justine Hospital in Montreal (QC) for their dedication of the collection of the Infant's cry database. This research work has been funded by a grant from the Bill & Melinda Gates Foundation through the Grand Challenges Explorations Initiative.”

CHAPITRE 3

ACOUSTIC MEASURES OF THE CRY CHARACTERISTICS OF HEALTHY NEWBORNS AND NEWBORNS WITH PATHOLOGIES

Yasmina Kheddache¹, Chakib Tadj¹

¹Department of Electrical Engineering, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, H3C 1K3 Canada
Article publié dans « J. Biomedical Science and Engineering »
vol.6, Aout 2013, pp.796-804

Résumé

Dans de nombreuses études, plusieurs hypothèses ont été formulées à la base d'observation des spectrogrammes des signaux de cris des nouveau-nés. Notre étude est basée sur quelques-unes de ces hypothèses. Le but de cet article est de différencier les cris pathologiques de ceux des bébés en bonne santé à travers l'analyse acoustique des cris en utilisant les paramètres neurophysiologiques des nouveau-nés. L'estimation automatique des caractéristiques des cris exprimé en pourcentages, tels que la phonation, l'hyperphonation, la dysphonation, aussi bien les sons non voisés et le changement de mode, nous ont aidé à distinguer les pathologies étudiées.

Les résultats obtenus nous ont permis d'établir des associations quantitatives entre les caractéristiques de cris et les différentes pathologies affectant les nouveau-nés.

Mots clés: cris des nouveau-nés, cri dysphonique, estimation du bruit.

Abstract

Several hypotheses have been formulated as a result of observing spectrograms of the audio signals of the newborn infant cry in numerous studies. Our study is based on a few of these hypotheses. The purpose of this article is to differentiate pathological crying from healthy crying through acoustic cry analysis based on neurophysiological parameters of newborns. The automatic estimation of the characteristics of relevant cry signals, such as phonation, hyperphonation, and dysphonation, expressed as percentages, as well as unvoiced sound and mode change percentages, have enabled us to distinguish among the pathologies selected for this study. The results obtained have helped us to make quantitative associations between cry characteristics and pathological conditions affecting newborns.

Keywords: Newborn Cry; Dysphonic Cry; Noise Estimation

3.1 Introduction

Crying provides information on the level of functioning of the central nervous system (CNS), because the cry signal is the result of coordination among several brain regions that control respiration and the vocal fold vibration from which the cry sounds are produced (LaGasse, Neal et Lester, 2005; Michelson Katarina 2007; Singer et Zeskind, 2001). Deficits in brain functioning may affect the vagal control of the cry (Michelson Katarina 2007; Singer et Zeskind, 2001).

Newborn crying is characterized by voiced/unvoiced sounds and a higher fundamental frequency, F_0 , than those of adults, and by abrupt changes in frequency that does not occur in adults. Unvoiced sounds, or turbulences, are produced by forcing air through a constriction in the vocal tract, and voiced sounds are generated by forcing air through the glottis (LaGasse, Neal et Lester, 2005; Manfredi, 1998).

We can distinguish three cry modes of vocal fold vibration (LaGasse, Neal et Lester, 2005): phonation, hyperphonation, and dysphonation, as defined in Section 3.4. By observing

spectrograms and spectra of audio signals of cries, associations have been made with pathological conditions (LaGasse, Neal et Lester, 2005; Wasz-Höckert, Michelsson et Lind, 1985).

Our main goal in this study is to use cry characteristics to improve infant monitoring in the first days of life. Automatic measurement of these characteristics allows us to associate the most relevant characteristics with pathologies of interest, and also to extend the measurement to other pathologies.

In this paper, we concentrate on the measurement of certain acoustic parameters of pathological and healthy newborn cries, particularly the percentage of dysphonic cries characteristic of the pathologies studied. The percentages of other cry characteristics, such as phonated and hyperphonated cries, which were investigated in (Kheddache et Tadj, 2013c), and also of unvoiced sounds and mode changes, are estimated.

The principal objective of this work is to determine the relationship between this acoustic parameter and the various pathologies studied. In order to evaluate dysphonia in healthy and pathological cries, the noise component in newborn cry signals must be estimated.

Several approaches have been proposed for the estimation of noise power levels. In the context of degraded speech due to additive noise, the MMSE (minimum mean-squared error)-based noise estimator and the ACF (adaptive comb filtering) method have been used (Frazier et al., 1976; Gerkmann et Hendriks, 2011; 2012). However, only the ACF method has been used with NNE (normalized noise energy) measurement for determining the noise component of pathological voice signals (Manfredi, 1998; Kasuya, Ogawa et Kikuchi, 1986).

In the case of the newborn, determination of the noise component in dysphonic cries using these two methods can produce an over estimation of noise. In the case of MMSE-based noise estimation, an initialization of 0.5 s with noise only is used. However, in the case of

ACF, the overload problem caused by rapid change in fundamental frequency produces undesirable results (Frazier, 1975).

In this paper, we present the performance of a new proposal based on the use of the MMSE and ACF methods to estimate the amount of noise present in the cries of newborns, and also to determine the percentage of dysphonia in healthy and pathological cries. We used the MMSE-based noise estimator combined with the ACF method to initialize noise estimation. This paper is organised as follows. We first present a broad definition of dysphonation in the newborn cry in Section 3.2, and the database used in this work in Section 3.3. We then explain our methodology for the automatic estimation of the percentage of dysphonation in healthy and pathological cries of newborns, and that for other relevant characteristics in Section 3.4. In Section 3.5, we address dysphonation estimation in newborn cries. We review NNE acoustic measurement in Section 3.5, and discuss the implementation of MMSE-based noise estimation adapted to newborn cries using the ACF-based noise estimator for initialization of the estimation. An analysis of our results is presented in Section 3.6, and our conclusions are provided in Section 3.7.

3.2 Dysphonic cry

In the literature, dysphonia is considered a functional voice disorder affecting the respiratory, cervical, costal, and abdominal muscles (Hirschberg et al., 1995). The level of dysphonia is measured by the amount of turbulent noise present, or by parameters (timbre, intensity, and pitch) that deviate from the normal (Hirschberg, 1999).

The dysphonic cry is the product of a portion of the expiratory phase of the respiratory tract. It is characterized by harmonics that are undefined and non equidistant (Cecchini, Lai et Langher, 2010), and caused by noisy or non harmonic vibration of the vocal folds due to unstable respiratory control (LaGasse, Neal et Lester, 2005; Lester et LaGasse, 2008). Dysphonation is a cry characteristic that yields negative cry ratings, such as increased variability of F0 and increased hyperphonation (Cecchini, Lai et Langher, 2010).

According to (LaGasse, Neal et Lester, 2005), dysphonation and frequent mode changes between phonation and dysphonation are associated with prenatal insult and perinatal risk. The presence of one of these markers indicates poor control of the respiratory system. The authors of (Cecchini, Lai et Langher, 2010) show that the percentage of dysphonic crying is more important than the presence or absence of that characteristic.

According to (Wasz-Höckert, Michelsson et Lind, 1985), noise concentration is extremely rare in normal infant crying. Without any alteration of the voice organs, a healthy infant can temporarily produce noisy vocalization. However, a healthy infant's cry is generally noiseless, compared to that of an infant with a pathological condition (Hirschberg, 1999). A dysphonic cry can also be a result of the immature innervation of the larynx. In this case, the cry spontaneously disappears (Hirschberg et al., 1995; Hirschberg, 1999). The infant that cannot produce a regular cry at all is certainly sick (Hirschberg, 1999). A higher noise concentration has been found in both the voiced and voiceless signals of the cries of infants with herpes simplex virus encephalitis (Michelson Katarina 2007; Manfredi, D'Aniello et M. and Brusciaglioni, 1998) and laryngitis, and in infants exposed to parental cocaine, marijuana, or alcohol use (LaGasse, Neal et Lester, 2005; Michelson Katarina 2007). The dysphonic cries are associated with higher scores for anguish, anger, and urgency (Cecchini, Lai et Langher, 2010).

3.3 Cry database

The database we used is the same as the one used in Kheddache et Tadj, 2013c). It contains 3396 cry samples of 1 s duration from 68 newborn babies aged 1 day to 4 weeks. Thirty-one healthy newborns produced 1774 of the samples (764 of the babies were premature) and thirty-seven newborns with a pathological condition produced 1622 of the samples (1079 of the babies were premature). Table 3.1 shows the pathologies studied according to the gestational age of the newborns (Premature infant with less than 37 weeks gestation and full-term infant between 37 and 42 weeks).

Table 3.1 The pathologies studied

Gestational age	Pathology	Sample size
Full-term newborn (t)	Healthy	1010
	Thrombosis in the vena cava	77
	Meningitis	115
	Peritonitis	20
	Asphyxia	190
	Lingual frenum	141
Preterm newborn (P)	Healthy	764
	IUGR- Microcephaly (Intra-Uterine Growth Retardation)	78
	Tetralogy of Fallot	53
	Hyperbilirubinemia	250
	Gastroschisis	280
	IUGR - asphyxia	148
	RDS (Respiratory Distress Syndrome)	270

These cries were collected with the aid of medical staff in the Neonatology department at Saint-Justine Hospital in Montreal. The conditions in which the cries were registered are: hunger, blood sampling, and diaper changing. The cries were registered using a small recording device placed 10 cm from the baby's mouth, at a sampling rate of 44.1 kHz. For each baby, three recordings of 2 to 3 minutes' duration were made, with at least a one hour interval after each recording session (over a period of 10 days at most). The time and date of

the session; the gender and date of birth of the baby; the pathology of the baby; and the reason for the crying were noted for each crying episode.

The spectrograms of different cry modes are shown in Figure 3.1. The phonated cry was extracted from a healthy, full-term newborn. The hyperphonated cry was extracted from a premature newborn suffering from IUGR-microcephaly, and the dysphonated cry was extracted from a full-term newborn suffering from asphyxia.

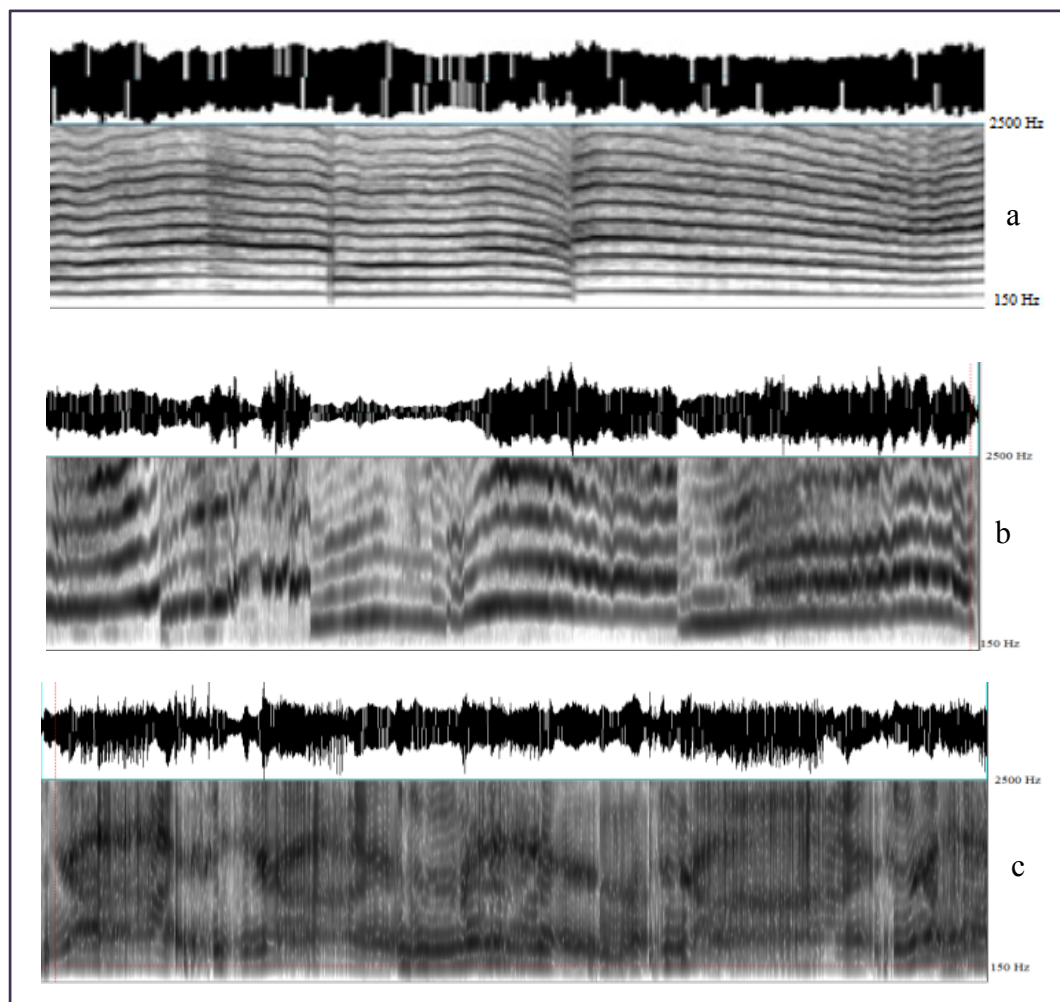


Figure 3.1 Spectrograms of different cry modes
a) phonation, b) hyperphonation, c) dysphonation

3.4 Methodology

The methodology adopted for automatic estimation of dysphonation and for the other characteristics presented in Table 3.2 for pathological and healthy newborn cries using Matlab is as follows:

- distribution of the cry samples by pathology and by gestational age;
- noise filtering and segmentation of the recordings into useful and non-useful portions using Praat;
- division of the signals into overlapping frames of 1024 samples with 512 recovering, and multiplication of each frame by the Hanning window;
- estimation of F0 using the SIFT algorithm (simple inverse filtering tracking) (Lederman, 2010; Markel, 1972);
- estimation of spectral noise power using the MMSE based noise estimator and the ACF method for initializing noise estimation;
- calculation of NNE (normalized noise energy);
- identification of phonic, dysphonic, and hyperphonic segments, as well as unvoiced segments;
- calculation of the average percentage of phonic (AP_{ph}), hyperphonic (AP_{hyp}), and dysphonic (AP_{dys}) segments, and the number of changes in cry mode ($AP_{ch.mo}$) and the number of unvoiced segments;

We followed this methodology, by pathology and by gestational age, using these formulas:

$$AP_{ph} = \frac{N_{ph}}{N_{total}} \quad (3.1)$$

$$AP_{hyp} = \frac{N_{hyp}}{N_{total}} \quad (3.2)$$

$$AP_{dys} = \frac{N_{dys}}{N_{total}} \quad (3.3)$$

$$AP_{ch.mo} = \frac{N_{ch.mo}}{N_{total}} \quad (3.4)$$

$$AP_{unv} = \frac{N_{unv}}{N_{total}} \quad (3.5)$$

where N_{total} , N_{ph} , N_{hyp} , N_{dys} , N_{unv} and $N_{ch.mo}$ are the total number of segments, the total number of phonic segments, the total number of hyperphonic segments, the total number of dysphonic segments, the total number of unvoiced segments, and the total number of changes in cry mode respectively.

- application of ANOVA variance analysis for healthy and pathological cries to analyze the estimated characteristics by gestational age (full-term, preterm).

The first three characteristics presented in Table 3.2 were investigated in (Kheddache et Tadj, 2013c).

Table 3.2 Measured characteristics of cries

Characteristic	Definition
Fundamental frequency F0	Vibration frequency (in Hz) of the vocal folds.
Phonation	Blocks of 1024 samples having an $F0 < 750$ HZ
Hyperphonation	Blocks of 1024 samples having an $F0 > 1000$ Hz.
Dysphonation	Blocks of 1024 samples containing noise or aperiodic sound.
Number of changes in cry mode	Number of blocks that the cry modes change between dysphonia and others modes.
Unvoiced sound	Blocks of 1024 samples having an $F0 = 0$ (no vocal fold vibration)

3.5 Dysphonation estimation in newborn cries

Many approaches have been proposed for evaluating the noise components of pathological voice signals, such as the harmonic-to-noise ratio (HNR) and FFT spectra methods, both of which are considered to be unhelpful for distinguishing a pathological voice from a normal voice (Kasuya et al., 1986).

In (Kasuya et al., 1986) and (Kasuya, Ogawa et Kikuchi, 1986), the authors indicate that NNE measurement is an effective method for doing so. If an NNE value is under a threshold (-10 dB), the voice is regarded as normal, and if it is over the threshold, or equal to it, the voice is regarded as pathological.

Our study is based on the hypothesis that a newborn's cries are considered to be dysphonic if the NNE value is larger than or equal to -10 dB (Manfredi, Tocchioni et Bocchi, 2006).

Two methods have been studied to estimate the noise component of healthy and pathological cry signals: 1) the ACF-based noise estimator; and 2) the MMSE-based noise estimator. However, both these methods either produce undesirable results, or overestimate noise in the case of newborn cries.

In the first method, according to (Frazier, 1975; Frazier et al., 1976; Lim, Oppenheim et Braida, 1978), overload problems due to rapid changes in F0 create segments on the cry waveform that are much shorter than neighboring segments. This is an issue because the filter coefficients are not in the same relative positions in their respective segments. The authors suggest turning off the coefficients involved in the short segments, but this proposal is at variance with the principle of the adaptive filtering method (Frazier, 1975).

In the second method, over estimation of the noise in the cry signals is caused by initialization of noise estimation in which noise is only considered in the initial 0.5 s of the signal, according to (Hendriks, Heusdens et Jensen, 2010).

To avoid this over estimation and improve noise tracking in the cry signal, we are using a new approach, based on combining the two methods: the MMSE-based noise estimator with an initial estimation of noise using the ACF-based noise estimator. This implementation is illustrated in the simplified block diagram in Figure 3.2.

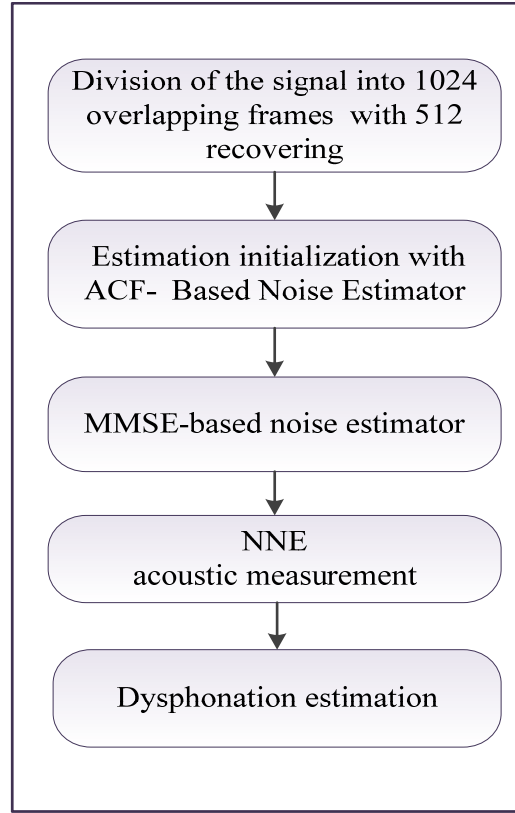


Figure 3.2 Dysphonation estimation in newborn cries

3.5.1 Normalized noise energy estimation (NNE)

The number of dysphonic components estimated in cry signals is quantified by means of NNE acoustic measurement, which, in this context, is defined as the ratio of the estimated noise component energy to the total cry energy (Kasuya, Ogawa et Kikuchi, 1986).

The cry signal $y_k(n)$ in the k^{th} frame is assumed to consist of periodic component $x_k(n)$ and the additive noise component $w_k(n)$, and is represented as follows:

$$y_k(n) = x_k(n) + w_k(n) \quad (3.6)$$

Let $Y(l)$, $X(l)$, and $W(l)$ be the l -point discrete Fourier transform (DFT) of $y(n)$, $x(n)$, and $w(n)$ respectively.

$$Y_k(l) = X_k(l) + W_k(l), l=1 \dots 1024 \quad (3.7)$$

$$NNE(k) = 10 \log \left[\frac{\sum_{l=l_L}^{l_H} \hat{W}_k(l)^2}{\sum_{l=l_L}^{l_H} Y_k(l)^2} \right], \quad k = 1, \dots, N \quad (3.8)$$

where: N is the number of frames over the whole analysis interval of the cry.

l_H and l_L are lowest and highest points of the frequency band where the noise energy is evaluated; and $\hat{W}$ is the estimate of W.

3.5.2 ACF-based noise estimator

A time variant digital ACF method was proposed by Frazier in 1975 for enhancing the intelligibility of speech signals degraded by the addition of competing speech signals or background noise (Frazier et al., 1976). This method is based on the observation that the waveforms of voiced sounds are periodic, with a period that corresponds to the fundamental frequency F0. We have adapted this method and applied it to estimating the noise components of a cry signal. The frequency spacing of the adaptive comb filter varies with the estimated F0 from a cry signal, and the spacing of the filter coefficients is adjusted to the individual pitch periods (Kasuya, Ogawa et Kikuchi, 1986). The estimation of F0 is achieved by means of the SIFT method, because it is so robust (Kheddache et Tadj, 2013c).

The ACF implementation is as follows:

The unit sample response of an adaptive comb filter $h(n)$ over one pitch period is:

$$h(n) = \sum_{k=-L}^L (a_k \cdot \delta(n - N_K)) \quad (3.9)$$

where $\delta(n)$ is a unit sample function.

According to (Lim, Oppenheim et Braida, 1978), for a given SNR (speech-to-noise ratio), ACF decreases in intelligibility as the filter length increases, but with a length of 3 pitch periods, the intelligibility score does not decrease. For this purpose, the length of the filter ($2L + 1$ pitch periods) is on the order of three times the length of one pitch period, which is equivalent to 3072 samples, or 0.069 s.

a_k : the filter coefficients that satisfy $\sum_{k=-L}^L a_k = 1$ are obtained from hamming windows shape (Gerkmann et Hendriks, 2012)

$$a_k = \frac{0.54 + 0.46 \cos(\frac{2\pi k}{L} + 1)}{\sum_{k=-L}^L 0.54 + 0.46 \cos(\frac{2\pi k}{L} + 1)} \quad (3.10)$$

N_k : updated once in every pitch period.

$$\begin{aligned} N_k &= - \sum_{l=k}^{-1} (T_l) & \text{for } k < 0 \\ N_k &= 0 & \text{for } k = 0 \\ N_k &= \sum_{l=0}^{k-1} (T_l) & \text{for } k > 0 \end{aligned} \quad (3.11)$$

T_l : pitch period which contains the points of cry waveform that is multiplied by the filter coefficient a_k .

The output of the filter is given by $x(n)$

$$x(n) = \sum_{k=-L}^L (a_k \cdot y(n - N_k)) \quad (3.12)$$

Then, an estimated noise component is given by:

$$\hat{w}(n) = y(n) - x(n) \quad (3.13)$$

3.5.3 MMSE-based noise estimator

The algorithm used to estimate the spectral density of the noise power is based on the minimum mean-squared error. The MMSE-based approaches have mainly been used in speech enhancement applications for reducing the additive non stationary noise without decreasing speech intelligibility (Gerkmann et Hendriks, 2012).

According to (Gerkmann et Hendriks, 2012), the MMSE- based spectral noise power estimator is much less demanding and at the same time robust to increasing noise levels. The authors state that it has been improved by using a soft speech probability presence (SPP) with fixed priors, instead of a voice activity detector (VAD). They show that the SPP approach results in less overestimation of the spectral noise power, and even less computational complexity.

The application of the MMSE-based noise estimator in cry signals is based on the assumption that the noise component of cries is more stationary than the periodic component signal, so the spectral noise power estimate of the previous frame is used for noise power (σ_w^2) estimation. In (Gerkmann et Hendriks, 2011; Hendriks, Heusdens et Jensen, 2010), a period of 0.5 s with noise only is used for initialization. To prevent initialization from being excluded from performance measurements, we use an ACF measure for initialization of the estimated noise in a period of 0.069 s, which corresponds to 3072 samples.

In our study, we replace the SPP with a periodic component probability presence (PCPP). The MMSE implementation using the PCPP is as follows:

The power spectrum of the noisy signal is estimated on a frame-by-frame basis and observed over a time span of about 1 s. This is assuming that the periodic component $x_k(n)$ and the additive noise component $w_k(n)$ have zero mean and are independent.

$|Y|^2$: cry signal,

$$E(|Y|^2) = E(|X|^2) + E(|W|^2) \quad (3.14)$$

$E(.)$: statistical expectation operator.

The spectral periodic component and noise power are defined respectively by:

$$E(|X|^2) = \sigma_x^2, \quad E(|W|^2) = \sigma_w^2 \quad (3.15)$$

Then, the a posteriori SNR is $\gamma = |Y|^2 / \sigma_w^2$ and the a priori SNR is $\zeta = \sigma_x^2 / \sigma_w^2$.

Assuming that both the periodic component and the noise DFT coefficients have a complex Gaussian distribution (Hendriks, Heusdens et Jensen, 2010), the PCPP is obtained with Bayes' theorem, assuming uniform priors $P(H_0) = P(H_1)$ (Gerkmann et Hendriks, 2011).

H_0 : absence of the periodic component; H_1 : presence of the periodic component.

The processing steps of the MMSE-based noise estimator are as follows:

- estimate the PCPP:

$$P(H_1|Y) = (1 + (1 + \zeta_{opt}) \exp\left(-\frac{|Y|^2}{\sigma_w^2} \frac{\zeta_{opt}}{1 + \zeta_{opt}}\right))^{-1} \quad (3.16)$$

σ_w^2 : Previous noise power estimate

$$\sigma_w^2 = \sigma_w^2(k-1) \quad (3.17)$$

$10\log_{10}(\zeta_{\text{opt}}) = 15\text{dB}$, optimal noise level fixed a priori.

Avoid the stagnation of the noise power update due to an underestimated noise power:

Recursive smoothing of the a posteriori PCPP:

$$P(k) = 0.9 P(k-1) + 0.1 P(H_1|Y(k)) \quad (3.18)$$

- forcing the a posteriori PCPP estimate to be lower than 0.99:

$$P(H_1|Y(k)) = \begin{cases} \text{Min}(0.99, P(H_1|Y(k))) & \text{if } P(k) > 0.99 \\ P(H_1|Y(k)) & \text{Else} \end{cases} \quad (3.19)$$

- estimate the noise periodgram:

$$\begin{aligned} E(|W|^2/Y) &= P(H_0|Y) |y|^2 + P(H_1|Y) \sigma_w^2 \\ P(H_0|Y) &= 1 - P(H_1|Y) \end{aligned} \quad (3.20)$$

- obtain the noise power spectral density via recursive smoothing with $\alpha=0.8$ (Gerkmann et Hendriks, 2012):

$$\sigma_w^2(k) = \alpha \sigma_w^2(k-1) + (1 - \alpha) E(|W|^2/Y) \quad (3.21)$$

3.6 Results and analysis

Cry signals were classified and analyzed according to the type of disease afflicting the newborn and the gestational age of the newborn (full term, preterm).

Figure 3.3 shows that there is a large difference in the amount of noise estimated in cry signals using the two methods studied when used separately. These results confirm that these methods (the MMSE-based noise estimator and the ACF-based noise estimator) produce an

over estimation of noise for both healthy and pathological cries. We also deduce from these results that noise estimation in cry signals using the MMSE-based noise estimator with ACF estimation initialization provides useful results for distinguishing between healthy and pathological cries. Consequently, we use this method for dysphonation estimation in healthy and pathologic cry signals.

The complete set of characteristics given in Table 3.2 was estimated for the entire database presented in Table 3.1, as explained in Section 3.4. The average percentage of dysphonic segments and unvoiced segments, and the percentage of mode changes by pathology are presented in Figure 3.4, and the average percentage of cry modes (dysphonia, hyperphonation, and phonation) by pathology and gestational age are presented in Figure 3.5.

Figure 3.4 shows that the estimated average percentages of dysphonic segments and unvoiced segments readily allow a distinction to be made between a healthy newborn and a sick newborn, especially for a full-term baby. These percentages are (17.2%, 1.84%) for meningitis, (25.6%, 5.13%) for vena cava thrombosis, (28.4%, 6.9%) for peritonitis, (36.3%, 3.1%) for lingual frenum, and (41.9%, 8.6%) for asphyxia respectively, compared to (16%, 1.3 %) for healthy cry samples.

The results shown in Figure 3.4 also indicate that the average percentages of dysphonic segments and un-voiced segments in the cry samples of healthy preterm newborns (19%, 2.8%) respectively are higher than those of healthy full-term newborns. This difference may be a result of the immature innervation of the larynx in preterm newborns. In this category of newborns, the average percentages of dysphonic and unvoiced segments contained in some pathological cry samples are a little higher, and are lower for other pathologies than those for healthy preterm newborns. These percentages are (23.5%, 2.6%) for gastroschisis, (25.1%, 4.9%) for RDS (respiratory distress syndrome), (23.9%, 3.5%) for hyperbilirubinemia, (26.5%, 5.7%) for IUGR-asphyxia, (11.4%, 0.9%) for the tetralogy of Fallot, and (14.5%, 3.6%) for IUGR-microcephaly respectively.

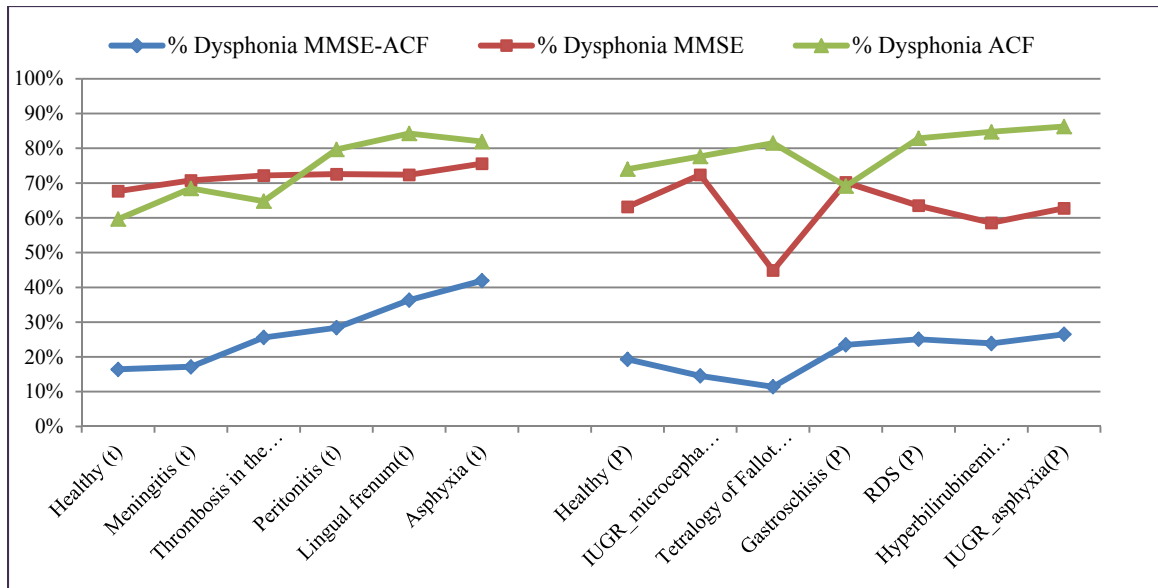


Figure 3.3 The average percentage of dysphonic segments by pathology and by measurement method

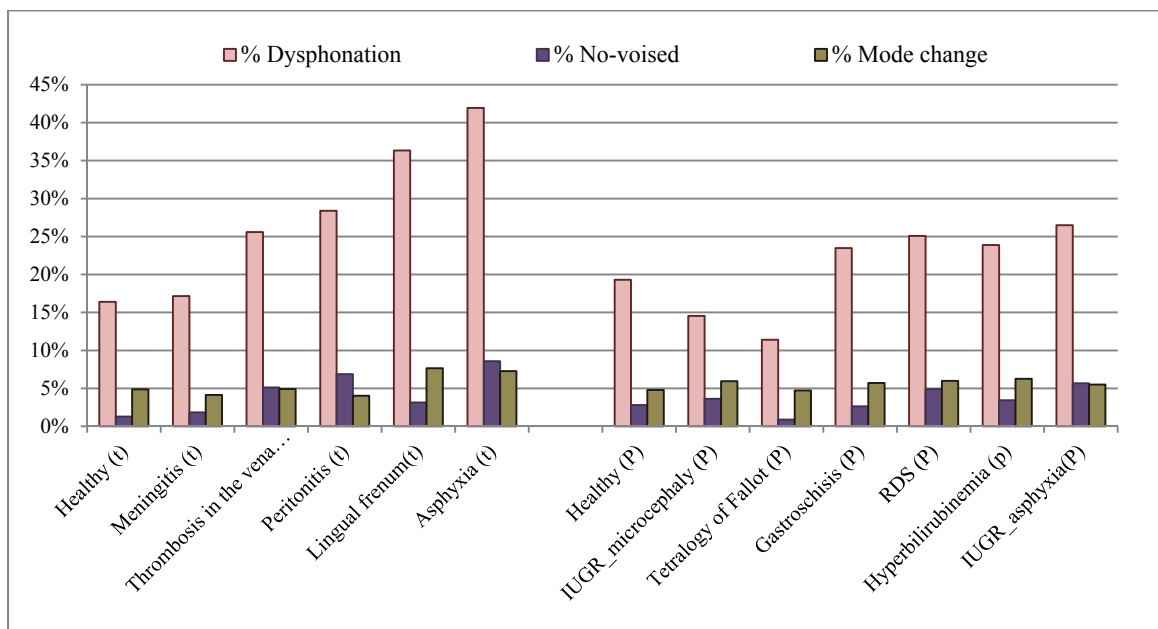


Figure 3.4 The average percentage of dysphonic and no-voice segments and the percentage of mode changes by pathology

Concerning the mode-change characteristic, Figure 3.4 shows that this characteristic varies from one pathological condition to another, and are dependent on gestational age. The percentages we found for full-term newborn cries are (4.1%) in meningitis, (4.9%) in vena cava thrombosis, (4%) in peritonitis, (7.7%) in lingual frenum, and (7.3%) in asphyxia, while it was (4.9%) for healthy cry samples. For preterm newborns, the average mode change percentages were as follows: (5.7%) for gastroschisis, (6%) for RDS, (6.3%) for hyperbilirubinemia, (5.5%) for IUGR-asphyxia, (4.7%) for the tetralogy of Fallot, and (6%) for IUGR-microcephaly, while it was (4.8%) for healthy cry samples.

We performed ANOVA variance analysis for healthy and pathological cries to compare the behavior of the estimated characteristics (percentage of dysphonic and no-voice segments, and of mode changes of cries) with respect to gestational age (full-term, preterm). We note from the results in Table 3.3 that there is significant statistical difference ($F = 9.710$, $p < 0.05$) between the average percentage of no-voice segments of healthy premature cries and that of full-term healthy cries, and also between the average percentage of mode changes in healthy pre-mature cries and that of full-term healthy cries ($F = 31.300$, $p < 0.05$). For all other cases, no significant difference was found by gestational age. We can conclude that the percentage of no-voice segments and mode changes is dependent on gestational age for healthy cries, and that these characteristics are not dependent on gestational age, but on the pathology itself for pathological cries. The percentage of dysphonia is not dependent on gestational age in either healthy or pathological cries.

Table 3.3 Analysis of variance (ANOVA) according to gestational age

Characteristics	Healthy		Pathologic	
	F	p	F	p
% Dysphonia	3,710	0,054	3,573	0,091
% No-voiced	9,710	0,002	1,399	0,267
% Mode change	31,300	0,000	0,015	0,905

The distribution of cry modes (dysphonation, hyperphonation, and phonation) by pathology and by gestational age that we show in Figure 3.5 is designed to provide a better characterization of pathological and healthy cries.

The results of this study show (Figures 3.4 and 3.5) that the cries of healthy full-term infants are phonic cries. Those of an infant with meningitis are slightly more hyperphonic and dysphonic. Those of an infant with vena cava thrombosis or peritonitis contain more hyperphonation, dysphonation, and no-voice segments than the cries of a healthy infant. Those of an infant with lingual frenum present with more hyperphonation and dysphonation, and more mode changes. The highest average percentages of all the estimated characteristics are found in the cries of an infant with asphyxia.

In the preterm newborn category, we found that the cries of an infant with IUGR-microcephaly are characterized by more hyperphonic and fewer dysphonic segments than those of a healthy preterm newborn. The cries of an infant with gastroschisis are more hyperphonic and dysphonic, and have more mode changes than the cries of a healthy infant. We also found that the percentages of all the estimated characteristics are higher in cry signals related to the following pathologies: RDS, hyperbilirubinemia, IUGR-asphyxia compared to healthy preterm cries. In contrast, the percentages of all the estimated characteristics are lower in the cries of an infant with the tetralogy of Fallot.

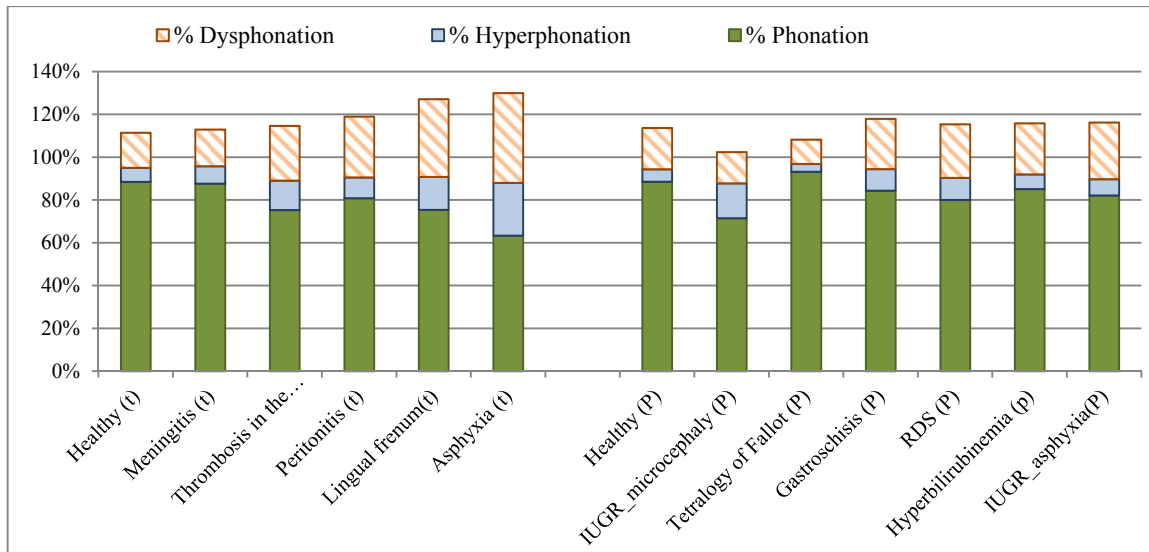


Figure 3.5 Distribution of cry modes (dysphonia, hyperphonation, phonation) by pathology

The calculation of the standard deviation for estimated characteristics by pathologies is shown in Figure 3.6. The results indicate a large dispersion for measured dysphonic segments in healthy preterm newborn, hyperbilirubinemia, asphyxia, Peritonitis, IUGR-asphyxia diseases. The best estimation of dysphonic segments is found in the case of healthy full-term, Vena cava thrombosis, Lingual frenum and meningitis diseases. For no-voice segments estimation, the large dispersion is found in the case of asphyxia, Peritonitis, IUGR-asphyxia, IUGR-microcephaly and RDS. The best estimation of no-voice segments is found in the case of Gastroschisis, tetralogy of Fallot and Lingual frenum diseases. For mode change segments estimation the large dispersion is found in the case of Gastroschisis and IUGR-asphyxia. The best estimation for this characteristic is found in healthy preterm newborn, meningitis, tetralogy of Fallot and hyperbilirubinemia diseases.

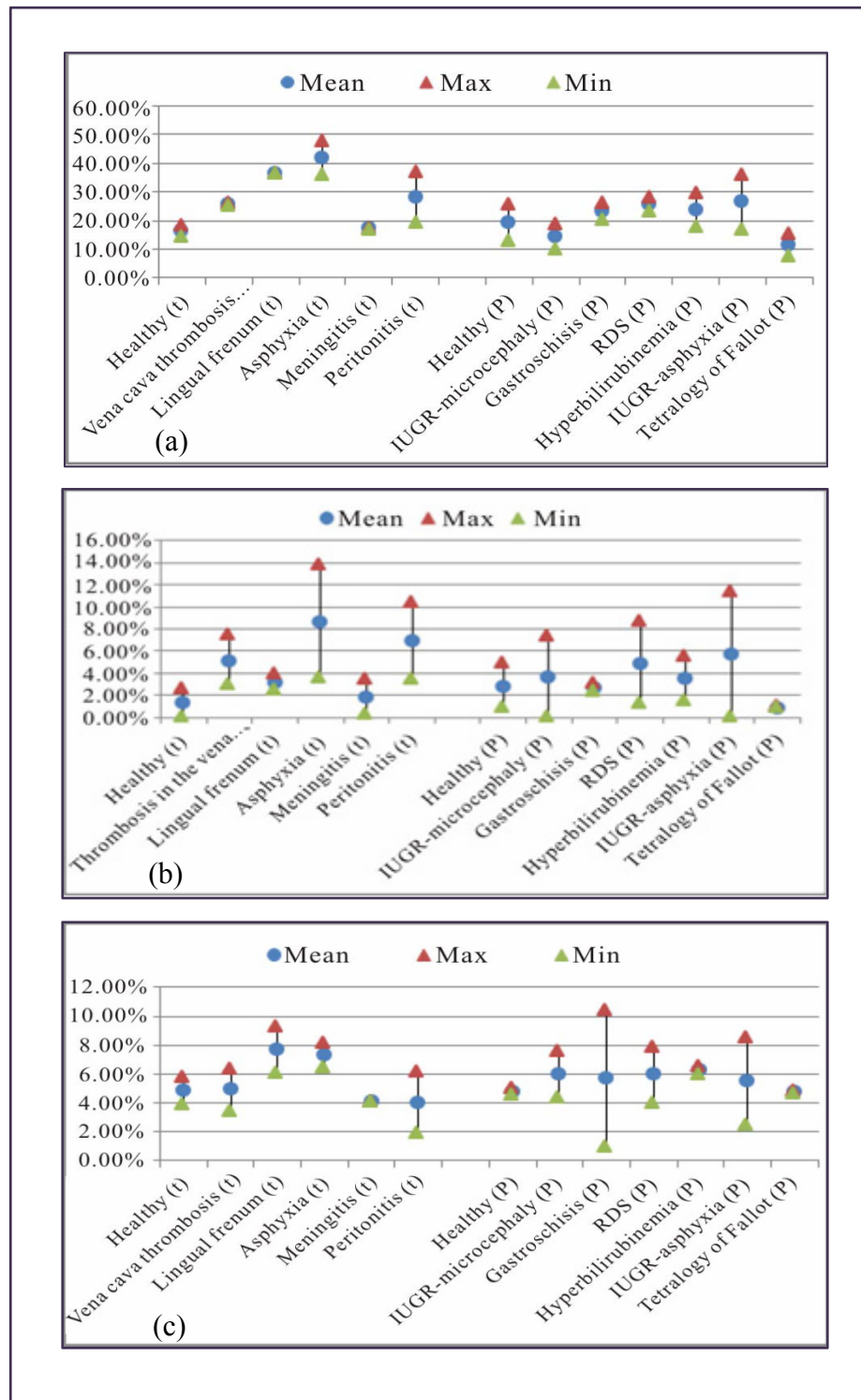


Figure 3.6 The mean percentage and standard deviation of a) dysphonic segments; b) no-voice and c) mode change by pathologies

3.7 Conclusions

In this paper, we have presented one of the most important steps of our research project, which is the quantitative characterization of cry signals. We have performed an automatic estimation of the most promising characteristics associated with pathological conditions that have been reported in the literature. As a result, we have established quantitative relationships between the pathologies studied and the characteristics of the cries of newborns.

We have also proved that some of the characteristics of healthy cries depend on gestational age, and other characteristics depend on the pathology itself.

The conclusion from the results we obtained is that there are clear differences between the characteristics of healthy cries and those of pathological cries in some pathologies, like asphyxia, gastroschisis, and vena cava thrombosis, RDS, hyperbilirubinemia, and IUGR-microcephaly. For each pathology studied, we have identified the most relevant characteristics for differentiating pathological cries from healthy cries. That differentiation can be used in diagnosing pediatric pathology.

However, it should be noted that validation of these results will require extension of the tests to a larger database containing a greater variety of pathologies and more subjects for each pathology.

Acknowledgements

We thank Dr. Barrington and members of the Neonatology group at Saint-Justine Hospital in Montreal (QC) for helping us to collect the Infant Cry database. This research has been funded by a grant from the Bill & Melinda Gates Foundation through the Grand Challenges Explorations Initiative.

CHAPITRE 4

RESONANCE FREQUENCIES BEHAVIOR IN PATHOLOGIC CRIES OF NEWBORNS

Yasmina Kheddache¹, Chakib Tadj¹

¹Electrical Engineering Department, École de technologie supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, H3C 1K3 Canada

Article publié dans « Journal of Voice », 2014

Résumé

Nous présentons dans cet article, une nouvelle approche de quantification automatique des paramètres acoustiques des cris pour les bébés en bonne santé et les bébés malades. Cette étude inclut F0, RF1, RF2, les durées et le pourcentage des tunings, les durées de transitions, la pente de RF2 et le ratio de RF1:RF2.

L'analyse statistique des résultats, révèle que la distribution de ces paramètres acoustiques dépend de la pathologie du nouveau-né. Dans ce travail, nous avons identifié avec succès la relation quantitative qui existe entre les caractéristiques acoustiques des cris que nous avons considérés et les pathologies étudiées.

Notre conclusion est que la variabilité de ces paramètres est utile pour différencier les cris d'un nouveau-né en bonne santé de ceux d'un nouveau-né malade. Ces données peuvent être utilisées pour le diagnostic précoce de certaines pathologies.

Mots clés : cri du nouveau-né, fréquence de résonance, tuning, durées des transitions, pente de RF2, ratio de RF1:RF2.

Abstract

A new approach to the automatic quantification of the acoustic parameters of the cries of healthy newborns and newborns with pathologies is presented. This study included F0, RF1, RF2, percentage and tuning duration, transition duration, RF2 slope, and RF1:RF2 ratio. Statistical analysis results reveal that the distribution of these acoustic cry parameters depend on the pathology of newborn. In this work, we successfully identify the quantitative relationship between the acoustic cry characteristics we examined and the diseases we studied. Our conclusion is that quantification of the variability of these parameters is useful for differentiating the cries of a healthy newborn from those of a newborn with a pathology, and that these data can be used for the early diagnosis of newborn diseases.

Keywords: Newborn cry; resonance frequencies; tuning, transition duration, RF2 slope, RF1:RF2 ratio.

4.1 Introduction

In this paper, the newborn cry is assessed acoustically, in order to correlate its acoustic properties to specific pathological conditions and to show how acoustic variations in the cry reflect the degree of distress of the newborn. In most crying research, the spectrogram has been used in cry analysis (K. Michelson, 2007; Wasz-Höckert, Michelsson et Lind, 1985; Michelsson et al., 2002). This method of analysis requires manual selection, which in turn requires some technical knowledge. In our work, we automate the evaluation of acoustic cry characteristics in order to improve infant monitoring in the first days of life.

To further investigate the relationship between infant crying and medical conditions from which infants may suffer, a set of the most important acoustic measures must be selected containing the most relevant information available based on physiological models of cry production. It is also important to understand the physioacoustic structure and levels of the nervous system responsible for muscle control and cry production.

The various acoustic characteristics are formed in accordance with changes in the status of the vocal tract, and may be affected by activity along the entire vocal production pathway, including respiration, vocal fold behavior, and vocal tract shape (LaGasse, Neal et Lester, 2005; Lester et LaGasse, 2008).

Greater control of vocalization in infants at around 1 to 2 months of age leads to greater differentiation in their cries, as a result of the physiological and anatomical changes occurring during this period. According to (Soltis, 2004), at around 3 months of age, crying is context-dependent, intentional, and communicative. At around 7 months, “babbling” occurs (Buder et al., 2008).

Consequently, because from 1 month of age, infants begin to acquire voluntary control of the vocal tracts (Cecchini, Lai et Langher, 2010), looking at the spontaneous cries of infants during the first weeks of life is crucial if we are to apply this information to the early diagnosis of various newborn pathologies. Our study concentrates on the characterization of the cries of newborns 1 day to 1 month old.

In this paper, the acoustic characteristics of healthy and pathological cries of preterm and full-term newborns are measured by automated means. The median and interquartile range of these characteristics are used to assess the degree of variation of:

- the average resonance frequencies (RF1, RF2) and the RF1:RF2 ratio;
- the fundamental frequency (F0);
- the percentage and duration of tuning periods for RF1 and RF2 with the first 10 harmonics of F0;
- the duration of the transition periods (TRP) between two consecutive TUPs;

- the RF2 slope.

These acoustic parameters were selected based on previous studies in which their usefulness, either for the development of phonatory and articulatory capabilities or for characterizing the production deficit in dysarthric speech, was reported. In (Kim et al., 2009; Kim, Kent et Weismer, 2011; Wermke et al., 2005) for example, the authors conclude that the irregular behavior of any of these parameters could provide suggest possible neurological dysfunction.

However, F0 and its variations over time are viewed as an essential component of reliable information on the health status of newborns (Kheddache et Tadj, 2013a; Kheddache et Tadj, 2013c) and resonance frequencies (RFs) reflect important acoustical characteristics of the vocal tract of the infant (Kim et al., 2009). Transition duration and tuning between RFs and the F0 harmonics have not been determined in infants younger than 4 months (Wermke et al., 2005). These characteristics are interpreted as early articulatory activity in the infant cry (Manfredi et al., 2003), and are related to the development of phonatory and articulatory capabilities (Wermke et al., 2005).

The RF1:RF2 ratio is an important marker of the relationship between the first and second formants, and helps us to understand the resonance frequencies (RFs) behavior in healthy and pathological newborn cries. This parameter is used to evaluate changes in voice and articulation in children with cochlear implants, for example (Poissant, Peters et Robb, 2006; Seifert et al., 2002).

Also, the RF2 slope has been reported in previous studies as one of the most sensitive indices of vocal tract function for speech production. It has been considered to be an indicator of the severity of dysarthria and of neurodegenerative diseases (Kim, Kent et Weismer, 2011). This parameter has not yet been measured in newborns or younger infants, however.

The method typically applied for RFs estimation is the parametric autoregressive approach. This method is particularly well suited to the assessment of the newborn cry, which is characterized by higher RFs than those of adults (Kasuya et al., 1986; Kim et al., 2009). For F0 estimation, we use the SIFT algorithm, described in (Kheddache et Tadj, 2013c; Kheddache et Tadj, 2013b).

This paper is organized as follows. We first present a physioacoustic model of the newborn cry in section 4.2, and characteristics of newborns cries in a medical environment in section 4.3. We then explain our methodology for measuring the characteristics of pathological and healthy cries in section 4.4. In sections 4.5 and 4.6, we provide a follow-up analysis of the results, and present our conclusions in section 4.7.

4.2 Physioacoustic Model and Neurological Basis of the Newborn Cry

The cry signal is the result of coordination among several areas of the brain which control respiration and the vocal fold vibration from which the cry sounds are produced (LaGasse, Neal et Lester, 2005; Michelson Katarina 2007). These sounds are created by the respiratory system (lungs and trachea), the vocal folds (larynx), and the vocal tract (pharynx, and oral and nasal cavities) (Soltis, 2004).

Cry sounds are generated in the larynx, which contains the vocal folds and glottis, by air being forced through a constricted tube, which causes vibration of the vocal folds at F0. This sound is then filtered as it proceeds through the vocal tract and the lips. The vocal tract acts as a resonance cavity, and its instantaneous shape determines the RFs. So, the cry sound is affected by formant frequency, the size and contours of upper vocal tract resulting in the audible cry (Cecchini, Lai et Langher, 2010; Manfredi et al., 2009).

The shape of the newborn's vocal tract is more like that of a chimpanzee than that of a human adult, where the position of the larynx is higher in the vocal tract. The vocal tract of the newborn is also shorter ($L \approx 8.5$ cm) (Manfredi et al., 2009) and has a different structure than that of an adult. This means that it is associated with higher resonances and a higher

fundamental frequency than it is in adults (Fort et Manfredi, 1998), because the formant frequencies decrease as the length of the resonance tract increases.

F0 varies from 250 Hz to 450 Hz, with an initial formant at a frequency FR1 of 1100 Hz and a second formant at a frequency FR2 of 3100 Hz (Michelson Katarina 2007). Vocal fold vibrations produce three identifiable modes of cry, defined as: a) a basic cry or phonation with F0 = 350-750 Hz; b) a cry with a high F0 (750-1000 Hz) or hyperphonation F0 (1000-1800 Hz); and c) noisy, turbulent, or dysphonic cries (Kheddache et Tadj, 2013a).

Interaction between laryngeal and pharyngeal activity is interpreted by tuning processes between the cry melody and the resonance frequencies (Manfredi et al., 2003). The tuning period (Michelsson, Tuppurainen et Aula) is defined as the time (>20 ms) during which an RF remains close to the harmonic distance (<100 Hz). Transition duration (TRP) is time between two consecutive TUPs (Kim et al., 2009; Manfredi et al., 2009). TRP depends on neurophysiological maturity and the integrity of the underlying control systems (Wermke et al., 2005). According to (Sundberg, Lã et Gill, 2013), RF tuning at round F0 increases the sound level of the vocal output.

The nervous system innervates the muscles that control the respiratory system, the vocal folds, and the vocal tract (Soltis, 2004). Consequently, the characteristics of cries are influenced by the cranial nerves that innervate the larynx, pharynx, and chest (Soltis, 2004) (Cecchini, Lai et Langher, 2010).

4.3 Newborn Cry Characteristics in a Medical Environment

Deficits in brain functioning can affect the vagal control of the cry (Michelson Katarina 2007). Many of the early researchers examined healthy and pathological cries by observing spectrograms and the spectra of the audio signals of cries. They essentially assert that infant cries are dependent on physical and psychological status, as well as both internal and external stimuli (Sundberg, Lã et Gill, 2013; Verduzco-Mendoza et al., 2009; Várallyay Jr et al., 2005). In spite of differences in measurement procedures, all cry studies have shown that

a high F0 is an indicator of a neurological problem (LaGasse, Neal et Lester, 2005). Hyperphonic cries and very high pitched cries are associated with neurological problems. Other markers associated with neonatal disease include noisy or dysphonic cries, as well as changes in phonation mode, and variability in F0 and F1 (Kasuya et al., 1986; Kheddache et Tadj, 2013a). When a CNS (central nervous system) disorder is involved, the cry exhibits auditory abnormalities with a high F0 and an irregular melodic contour (Cecchini, Lai et Langher, 2010).

The work conducted in (Kheddache et Tadj, 2013c) shows that the average percentages of hyperphonic segments in cry samples are similar for both healthy premature infants and healthy full-term newborns. It also shows that the average percentages of F0 irregularity are slightly higher for premature infants than for healthy full-term newborns. In addition, according to (Kheddache et Tadj, 2013a), the average percentages of dysphonic segments in the cry samples of healthy preterm newborns are higher than those of healthy full-term newborns. This difference may be a result of the immature innervation of the larynx in preterm newborns.

RF1 and RF2 have been used to identify differences between children who are profoundly deaf and those with normal hearing (Poissant, Peters et Robb, 2006). According to (Cecchini, Lai et Langher, 2010; LaGasse, Neal et Lester, 2005), high variability in F1 and F2 (dysregulation) indicates poor or unstable neural control of vocal track and respiration. This characteristic has been associated with hyperbilirubinemia, as well as prenatal tobacco and cocaine exposure. A higher than normal F1 has been considered to be representative of an excitable neurobehavioral syndrome resulting from the direct effects of cocaine (Robb et Cacace, 1995).

4.4 Method

4.4.1 Recording sessions

The recordings were made in the Pediatrics department at Sainte-Justine Hospital under the supervision of medical personnel. The cries of newborns 1 day to 1 month old were recorded using a small recording device placed 10 cm from the baby's mouth, at a sampling rate of 44.1 kHz. These include full-term and preterm newborn cries (premature infants at under 37 weeks' gestation and full-term infants at 37 to 42 weeks' gestation), healthy infants, and infants suffering from pathological conditions.

The cries of each infant were recorded two or three times at intervals of at least one hour over a period of not more than ten days. The conditions during which the cries were recorded were: hunger, blood sampling, and diaper changing. The date and time of recording, and the infant's ID, birth date, diagnosis, ethnicity, and so on, were noted for each crying episode. In order to confirm the health status of the infant or to determine that an undiagnosed health problem was, in fact, present during recording, we noted the results of a follow-up examination conducted six months after the cries had been recorded. We further categorized the cry recordings by pathology and by gestational age. A sample from the newborn cry database (CDB) that we created is presented in Table 4.1. The CDB contains 3000 cry samples of 1 s duration from 65 newborn babies.

Table 4.1 Pathologies Studied

Gestational age	Pathology	Age at recording (days)	Sample size
Full-term newborn (t)	Healthy	1 to 3	1010
	Vena cava thrombosis	2 to 3	77
	Meningitis	14 to 15	115
	Peritonitis	28	20
	Asphyxia	4 to 28	190
	Lingual frenum	1 to 2	141
Preterm newborn (P)	Healthy	3 to 28	764
	IUGR – Microcephaly (in utero growth retardation)	1	78
	Tetralogy of Fallot	17	53
	Gastroschisis	6 to 20	134
	IUGR – Asphyxia (intra-uterine growth retardation)	30	148
	RDS (respiratory distress syndrome)	28	270

4.4.2 Procedure

Noise filtering and segmentation of the recordings into useful and non-useful categories were performed in the first step of the assessment procedure using Praat. For the acoustic analysis of newborn cries, an automated assessment procedure was performed using Matlab to measure the characteristics defined in Table 4.2. The main steps of the assessment procedure are presented in Figure 4.1.

Table 4.2 Cry Characteristics Measured

Characteristic	Definition
Fundamental frequency (F_0)	Vibration frequency (in Hz) of the vocal folds.
Harmonics	Multiples of fundamental frequency.
RF1, RF2 (Hz)	First and second vocal tract resonance frequencies.
RF1:RF2 ratio	Derived using RF1 and RF2.
Tuning	Blocks of 1024 samples during which an RF remains close to the harmonics of F_0 (distance <100 Hz).
Tuning duration	The time between the start and end of tuning.
Transition duration (TRP)	Duration between two subsequent TUPs, or the time between the start and end of transition.
Transition extent (Hz)	Frequency change between the transition start and end.
RF2 slope (Hz/ms)	Derived using transition extent and duration measures.

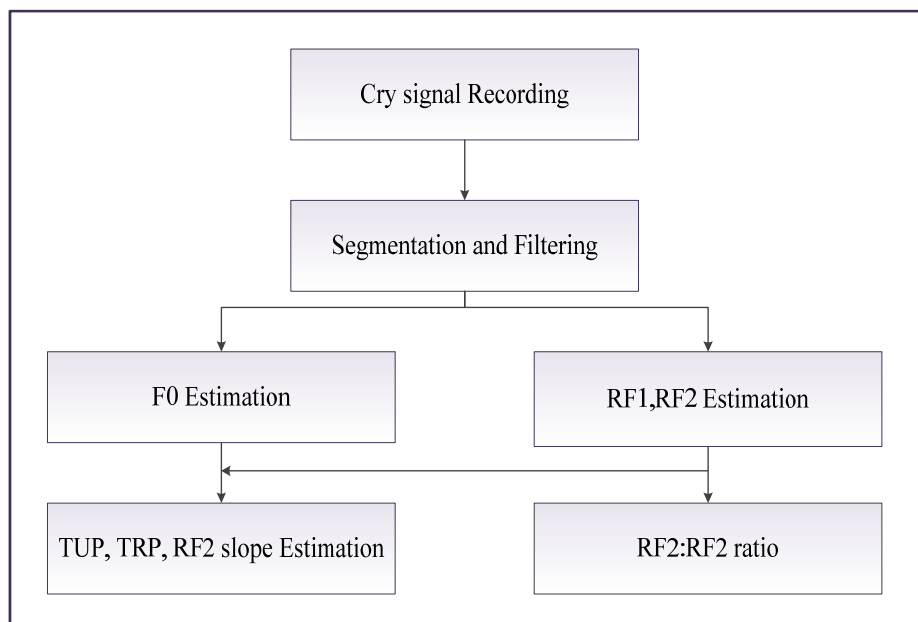


Figure 4.1 Methodology. (“Signal”; “Filtering”, “RF1, RF2”, “Slope”, “Ratio”)

We applied the following procedure for acoustic characteristic measurement:

- Division of the cry samples of 1 s in duration into overlapping frames of $N=1024$ samples, resulting in 512 covered frames, and the multiplication of each frame by the Hamming window;
- For each data frame, the first five acoustic measures defined in Table 2 were estimated as follows:
 - Estimation of F_0 and its 10 harmonics using the SIFT algorithm (simple inverse filtering tracking) (Markel, 1972);
 - estimation of RF_1 , RF_2 using the modified covariance method based on autoregressive power spectral density (AR-PSD) (Fort et Manfredi, 1998; Manfredi et al., 2009);
 - calculation of the $RF_1:RF_2$ ratio (V_{ratio});
 - determination of the tuning between the RFs and the first 10 harmonics.
- For each cry sample of 1 s duration, we estimated:
 - the average F_0 , RF_1 , RF_2 (A_{F_0} , A_{RF_1} , A_{RF_2});
 - the average $RF_1:RF_2$ ratio (A_{ratio});
 - the tuning percentage of (P_{TUP});
 - the tuning duration (D_{TUP}) for RF_1 , RF_2 and its average (A_{TUP});

- the transition duration (D_{TRP}) for RF1, RF2 and its average (A_{TRP});
- the extent of the transition;
- the RF2 slope (V_{SLP}) and its average (A_{SLP}). (absolute values were taken for the slope and the extent of the transition;

The following formulas were used:

$$A_{F0} = \frac{\sum F0}{N_{Total}} \quad (4.1)$$

$$A_{RF1} = \frac{\sum RF1}{N_{Total}} \quad (4.2)$$

$$A_{RF2} = \frac{\sum RF2}{N_{Total}} \quad (4.3)$$

$$A_{ratio} = \frac{\sum V_{ratio}}{N_{Total}} \quad (4.4)$$

$$P_{TUP} = \frac{N_{TUP}}{N_{Total}} \quad (4.5)$$

$$A_{TUP} = \frac{\sum D_{TUP}}{N_{TUP}} \quad (4.6)$$

$$A_{TRP} = \frac{\sum D_{TRP}}{N_{TRP}} \quad (4.7)$$

$$A_{SLP} = \frac{\sum V_{SLP}}{N_{TRP}} \quad (4.8)$$

N_{Total} : Total number of segments in cry samples of 1 s duration.

N_{TUP} : Number of frames where tuning occurred between RF1, RF2 and the F0 harmonics.

N_{TRP} : Number of transition durations in cry samples of 1 s duration.

- calculation of the median and interquartile range of the evaluated characteristics of pathological cries and healthy cries.

4.4.3 Resonance frequencies estimation

The cry signal as a speech signal contains both vocal tract and pitch information. In this paper, we use the source filter linear model of speech production in which the glottis and the vocal tract are considered independently. So, passing the glottal pulse train through the filter makes it possible to remove the correlation between the vocal tract information and the pitch information (Fort et Manfredi, 1998).

Assuming that the vocal tract acts as the filter, then, according to the all-pole autoregressive (AR) model, the cry signal is the output of an all-pole filter $1/A(z)$, $A(z)$ being called the inverse filter.

Consequently, the cry sample is a weighted linear combination of n previous samples described by the equation:

$$y(k) = \sum_{i=1}^n a_i y(k-i) + w(k) \quad (4.9)$$

where: N is the number of samples in a frame; $k=[1 \dots N]$; $y(k)$ is the cry signal.

$w(k)$ is the prediction error or (LP) residual, which contains the excitation contribution to the cry signal; a_i are the $A(z)$ polynomial coefficients; and n is the model order.

However, the formant structures of the model are defined by the zero of the filter $A(z)$, which corresponds to the power spectral density peaks of the cry signal.

The filter equation is:

$$A(z) = 1 + \sum_{i=1}^n a_i z^{-i} \quad (4.10)$$

The power spectral density is given by:

$$PSD(f) = \frac{1}{\left| 1 + \sum_{i=1}^n a_i e^{-2j\pi f T} \right|^2} \quad (4.11)$$

The RFs are therefore obtained by peak picking in the autoregressive power spectral density (AR-PSD), which is estimated using the modified covariance method. This method reduces spectral line splitting and the bias of frequency estimation (Manfredi, Tocchi, et Bocchi, 2006). It is based on simultaneously minimizing the forward and backward prediction errors $w(k)$ in the least squares sense.

The model order n is chosen by means of the following formula: $n = \sim 0.5 F_s$ (kHz). For $F_s = 44.1$ kHz, $n = 22$. This relation comes from the physical constraint $n = 2 L F_s / c$, where L is the length of the vocal tract = 8.5cm, and $c = 34$ cm/ms, which is the speed of sound (Manfredi et al., 2009). Formant splitting may occur if n is overestimated, while an underestimation of n smoothes the spectrum and causes the misallocation of spectral peaks (Manfredi et al., 2009).

4.4.4 Fundamental frequency estimation

In this paper, the modified SIFT algorithm is used to determine F0. It was demonstrated that this algorithm includes the autocorrelation properties and the cepstral pitch analysis technique (Kasuya et al., 1986). In addition, the performance of this algorithm has been tested on the database of a real newborn cries (Kasuya et al., 1986; Manfredi et al., 2009).

This method is based on determining the autocorrelation sequence after performing glottal inverse filtering to attenuate the influence of the vocal tract.

The coefficients $\{a_i\}$ $i = [1, \dots, 4]$ of the inverse filter $A(z)$ are determined by a forward linear predictor which minimizes the prediction error in the least squares sense (Markel, 1972).

Knowing $\{a_i\}$, the inverse filter output $\{s_k\}$ can be calculated using the equation:

$$s_k = y_k + \sum_{i=1}^4 a_i y_{k-i} \quad (4.12)$$

Then, the output autocorrelation $\{r_i\}$ sequence is calculated as the autocorrelation sequence of $\{s_k\}$.

$$r_j = \sum_{k=0}^{N-1-j} s_k s_{k+j}, j = [1, \dots, N-1] \quad (4.13)$$

Then, peak picking is performed and a decision algorithm in which the peak value is compared to the voice threshold is applied (Markel, 1972). F0 is estimated using the following equation:

$$T0 = 1 / F0 = \arg \max \{r_j\} \quad (4.14)$$

The contour of the F0 result is then smoothed by the median filter.

4.5 Results and Analysis

Plots of typical examples of the estimated F0 and its harmonics with resonance frequencies RF1, RF2 superimposed are presented on a color scale in Figure 4.2. This figure shows that early harmonics-resonance interaction exists in both the healthy and pathological cries of newborns. Figures 4.2(a) and 4.2(b) show that healthy cries contain relatively long periods of adjusted RF1 tuning with a fast transition duration time.

In Figure 4.2(c), RF1 is low with a long tuning period. Figure 4.2(d) shows low RF1, RF2 with fast tuning and long transition duration for RF1 and unstable transition duration for RF2. Figure 4.2(e) shows high RF1, RF2 with long transition duration and relatively fast tuning for RF2. In Figure 4.2(f), F0 is more irregular and the RFs are moving to the nearest harmonic to form a tuning event. Sudden pitch changes are noted in Figures 4.2(d) and 4.2(f), which cause changes in the harmonic and RFs structures.

The various results of our investigation into the distribution of the acoustic characteristics of healthy and pathological cries that we studied are presented by pathology and by gestational age using a box-and-whiskers plot (25th, 50th, and 75th percentile) in Figures 4.3 and 4.4. The recordings used for this assessment are those of premature newborn cries (p) and those of full-term newborn cries (t), and include healthy infants and infants suffering from the pathologies listed in Table 4.2.

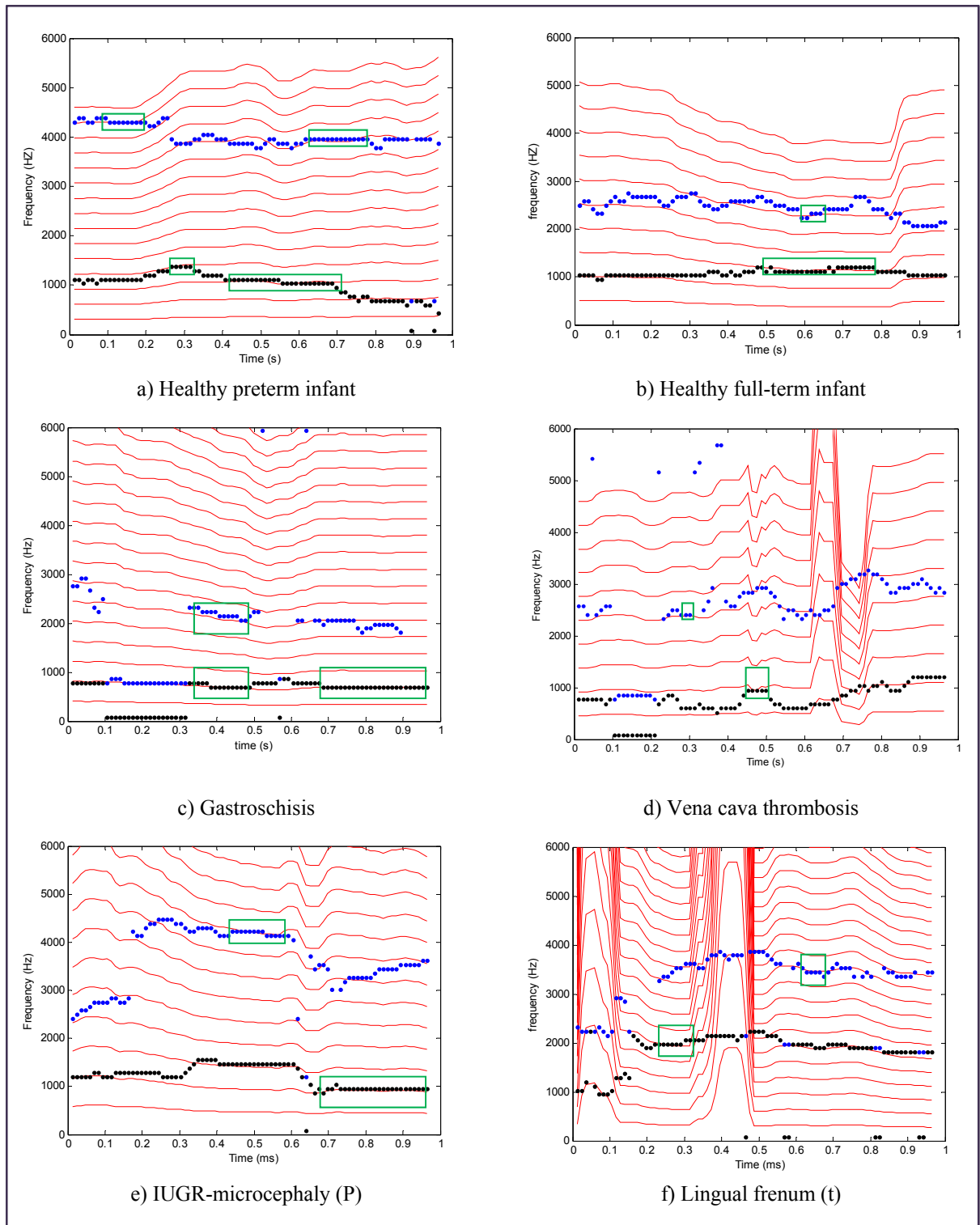


Figure 4.2 Fundamental frequencies and their harmonics (red lines), resonance frequencies: RF1 (black dots) and RF2 (blue dots), tuning between the RFs and F0 harmonics (green rectangles)

4.5.1 F0, RF1, and RF2 estimation results

Table 4.3 shows the median and interquartile range of the average F0, RF1, RF2 for all the pathologies studied. According to these results and those shown in Figure 4.3(a), the median and interquartile range of the average F0 in the case of Asphyxia (t), Vena cava thrombosis (t), and IUGR-microcephaly (P) are higher than those of healthy full-term newborns. The median and interquartile range of the average F0 for healthy preterm newborns is lower than that of healthy full-term newborns. This difference may be attributable to the difference in age at recording. The smallest inter-quartile range of the average F0 is found in the Tetralogy of Fallot (P) and the largest inter-quartile range of the average F0 is found in Lingual frenum (t) following Asphyxia (t).

As for the estimated RF1 average, its highest median and interquartile range are found in Lingual frenum (t), as shown in Figure 4.3(b) and Table 4.3. In cases of IUGR_asphyxia(P) and the Tetralogy of Fallot (P), the estimated median is high compared to those of healthy full-term and preterm newborns. But the interquartile range of the average RF1 in the case of the Tetralogy of Fallot (P) is the smallest. The lowest medians for the estimated RF1 average are found in Meningitis (t) and Peritonitis (t).

The estimation results in Table 4.3 and Figure 4.3(c) on the average RF2 indicate that the median of the average RF2 for IUGR_asphyxia (P) is higher and the interquartile range larger than those of the other pathologies studied or of healthy newborns. We also found that the median is high and the interquartile range large for Lingual frenum (t), and that the median of the average RF1, RF2 in healthy full-term newborns is lower and the interquartile range smaller than those of healthy preterm newborns. The lower medians are evident in both Vena cava thrombosis (t) and Peritonitis (t).

4.5.2 RF1:RF2 ratio and RF2 slope evaluation results

The medians and interquartile ranges of the A_{ratio} are shown in Table 4.3 and Figure 4.3(d). The results indicate a median value that is relatively the same for both healthy full-term and

preterm infants and also for newborns suffering from IUGR-microcephaly (P) and RDS (P), but with different interquartile ranges. The highest median, A_{ratio} , is associated with the cries of newborns with Lingual frenum(t), and the lowest with those with Vena cava thrombosis (t). As for Asphyxia (t), IUGR_asphyxia(P), the Tetralogy of Fallot (P), and Gastroschisis P, the A_{ratio} cry signal medians are higher than those in healthy cry signals, and for Meningitis (t) and Peritonitis (t) they are lower.

The A_{SLP} median and interquartile range results for the for both healthy and pathological newborn cries are given in Table 4.3 and illustrated in Figure 4.3(e). These results indicate that the median of A_{SLP} in healthy full-term newborns cries is slightly higher than in those of healthy preterm newborns. The newborn cries of infants with IUGR-microcephaly (P) or Vena cava thrombosis (t) are characterized by a high A_{SLP} median and a large interquartile range, unlike the newborn cries of newborns with Lingual frenum(t), which are characterized by a smaller A_{SLP} median compared with either healthy or pathological newborn cries. Also, not surprisingly, the A_{slp} median is lower than those of healthy newborns for Asphyxia (t), Meningitis (t), RDS (P), IUGR_asphyxia(P), the Tetralogy of Fallot (P), and Gastroschisis P).

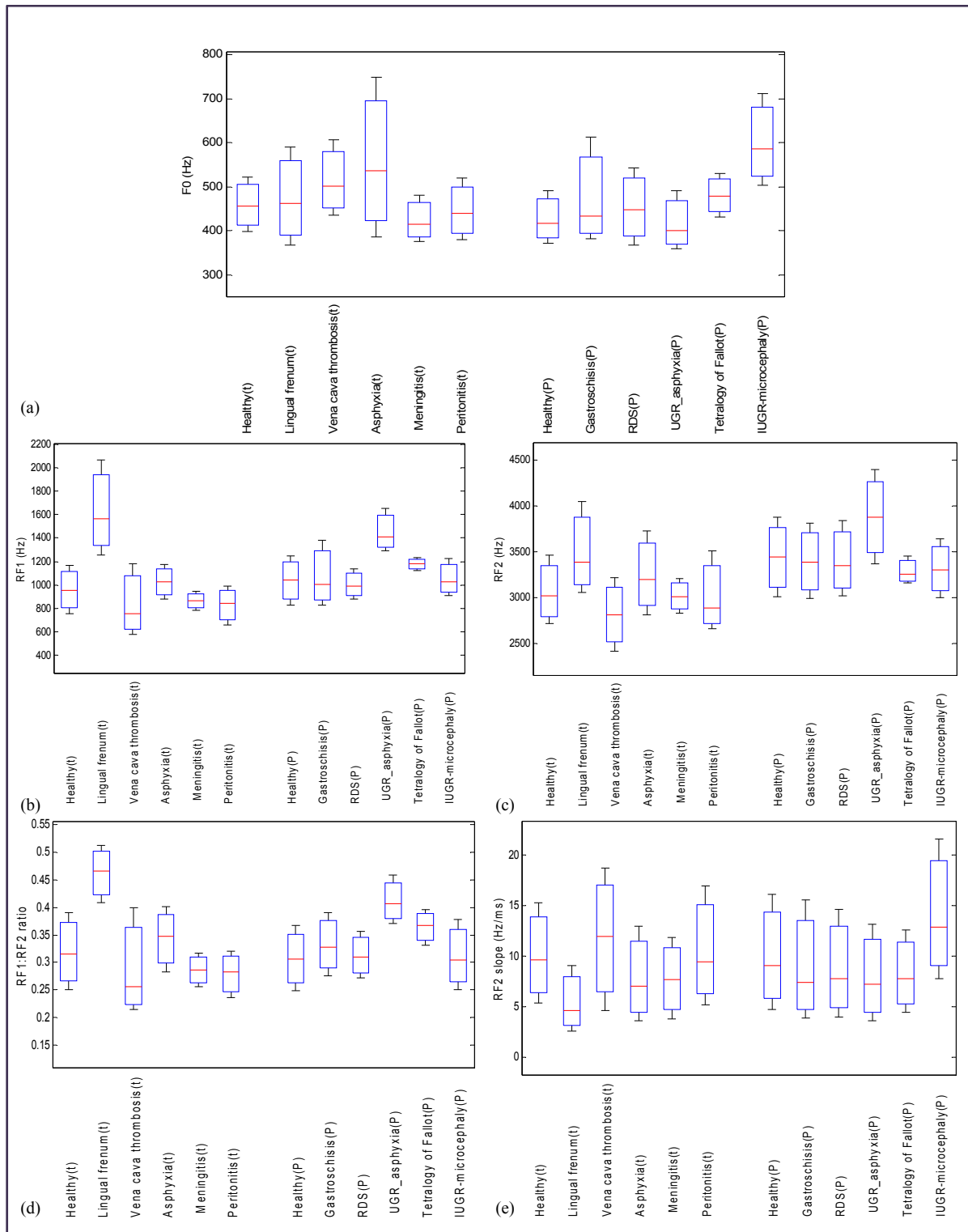


Figure 4.3 Box-and-whiskers plots for the averages of: a) F0, b) RF1, c) RF2, d) the RF1:RF2 ratio, e) the RF2 slope, by pathology and gestational age

Table 4.3 Measured characteristics of cries

Pathology	A_{F0} (Hz)		A_{RF1} (Hz)		A_{RF2} (Hz)		A_{ratio}		A_{slp} (Hz/ms)	
	Median	Inter-quartile range	Median	Inter-quartile range	Median	Inter-quartile range	Median	Inter-quartile range	Median	Inter-quartile range
Healthy (t)	456	94	955	311	3024	559	0.31	0.11	9.5	7.43
Lingual frenum (t)	462	168	1569	609	3393	738	0.46	0.08	4.63	4.82
Vena cava thromb (t)	501	128	754	456	2817	601	0.25	0.14	11.88	10.58
Asphyxia (t)	536	272	997	255	3002	635	0.35	0.1	6.98	7.06
Meningitis (t)	415	79	867	118	3009	281	0.28	0.05	7.67	6.06
Peritonitis (t)	440	105	846	248	2885	632	0.28	0.06	9.41	8.8
Healthy (P)	417	89	1045	314	3447	641	0.3	0.09	9.05	8.52
Gastroschisis (P)	433	174	1009	414	3388	620	0.33	0.08	7.35	8.79
RDS (P)	449	131	991	192	3347	616	0.31	0.06	7.7	8.01
IUGR_asphyxia (P)	400	98	1411	269	3878	773	0.4	0.06	7.14	7.15
Tetralogy of Fallot (P)	479	74	1180	79	3256	222	0.37	0.05	7.37	6.07
IUGR-microcephaly (P)	587	156	1028	241	3300	482	0.3	0.1	12.85	10.32

4.5.3 TUP and TRP evaluation results for RF1 and RF2

The various TUP and TRP measurements of the RFs are shown in Table 4.4 and Figure 4.4. The results shown in Figures 4.4(a), 4.4(b), and 4.4(c) indicate that the median and interquartile range of P_{TUP} between RF1 and the harmonics of F0 are relatively the same in healthy full-term and preterm newborns, and A_{TRP} is the same in both cases. But the A_{TUP} median is the highest in the healthy preterm cries studied. The P_{TUP} median between RF1 and the F0 harmonics is the highest in Lingual frenum(t), Meningitis (t), Gastroschisis (P), the Tetralogy of Fallot (P), and IUGR-microcephaly (P). The lowest P_{TUP} medians are found in IUGR_asphyxia (P) and in Vena cava thrombosis (t), but with a large interquartile range. Higher A_{TUP} medians for RF1 in Lingual frenum (t), Meningitis (t), Peritonitis (t), Gastroschisis (P), and RDS (P) are noted compared with those of the other pathologies studied and those of healthy newborns. The lowest A_{TUP} median and the highest A_{TRP} median for RF1 with a large interquartile range are noted in Vena cava thrombosis (t). The lowest A_{TRP} medians for RF1 are found in Lingual frenum (t), Meningitis (t), IUGR_asphyxia (P), and the Tetralogy of Fallot (P).

As for P_{TUP} , A_{TUP} , and A_{TRP} between RF2 and the F0 harmonics, Table 4.4 and Figures 4.4 (d), 4.4(e), and 4.4(f) indicate that the estimated median of P_{TUP} and A_{TUP} in healthy full-term newborns are higher than those of preterm newborns. Unlike P_{TUP} and A_{TUP} , the A_{TRP} median is the lowest in healthy full-term cries. We note the same lower P_{TUP} median in Lingual frenum (t), Vena cava thrombosis (t), and RDS (P). In Meningitis (t) cries, we found a higher P_{TUP} median and a lower A_{TRP} . The highest A_{TUP} median is found in Peritonitis (t) cries, and the highest median and largest A_{TRP} interquartile range are found in the cries of newborns suffering from Lingual frenum(t) and IUGR_asphyxia(P).

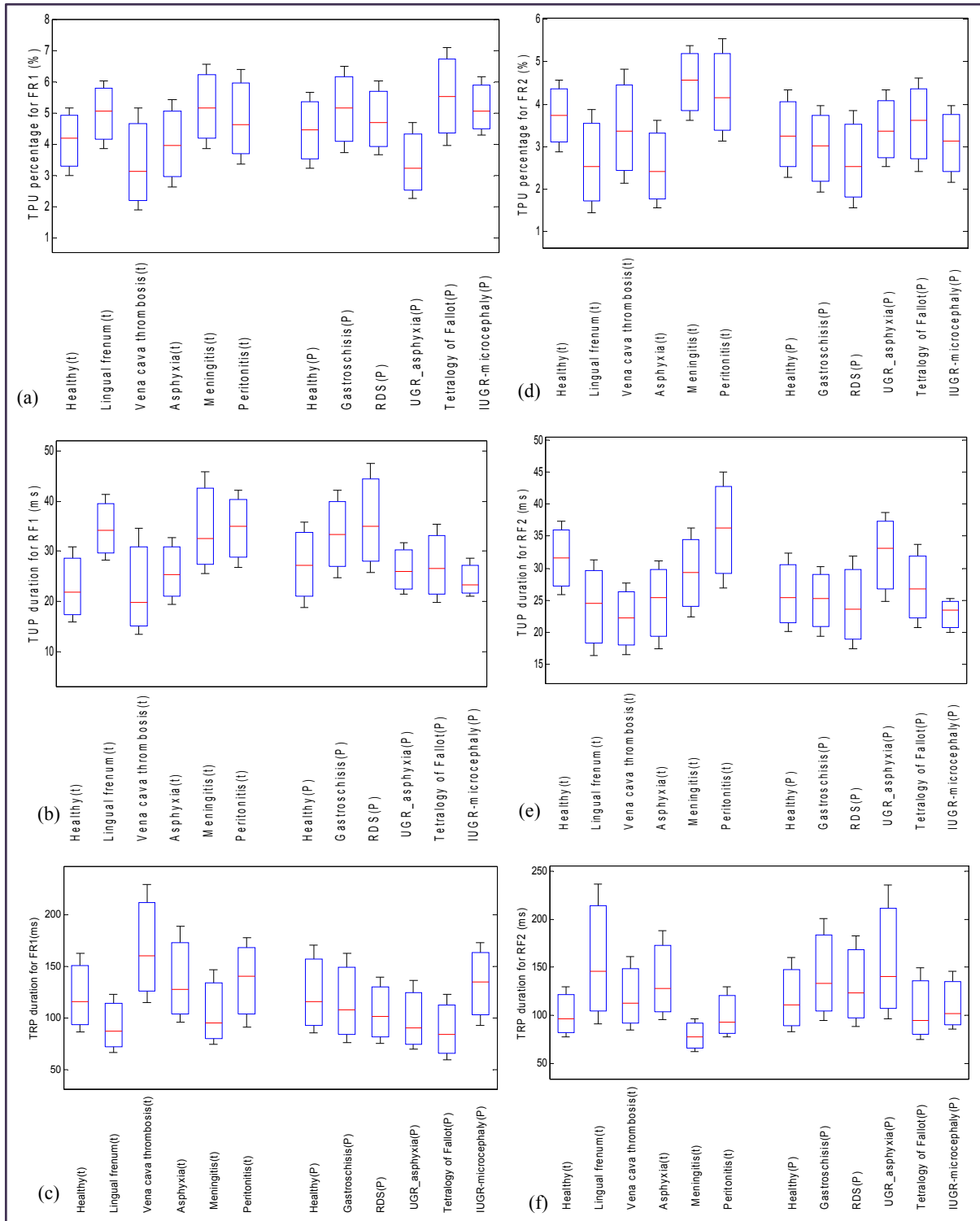


Figure 4.4 Box-and-whiskers plots for the average of : a) P_{TUP} for RF1, b) A_{TRP} for RF1, c) A_{TRP} for RF1, d) P_{TUP} for RF2, e) A_{TRP} for RF2, f) A_{TRP} for RF2, by pathology and gestational age

Table 4.4 Measured Characteristics of Cries

Pathology	P _{TUP} for RF1 (%)		A _{TUP} for RF1		A _{TRP} for RF1		P _{TUP} for RF2 (%)		A _{TUP} for RF2		A _{TRP} for RF2	
	Median	Inter-quartile range	Median	Inter-quartile range	Median	Inter-quartile range	Median	Inter-quartile range	Median	Inter-quartile range	Median	Inter-quartile range
Healthy (t)	4.2	1.6	22	11.2	116.1	57	3.7	1.3	31.6	8.6	96.5	39.1
Lingual frenum (t)	5	1.6	34.2	9.8	87.5	42.3	2.5	1.8	24.4	11.3	146.3	109
Vena cava thrombosis(t)	3.1	2.5	19.8	15.9	160	85.1	3.4	2	22.3	9.3	112.2	56.9
Asphyxia (t)	4	2	25.4	9.9	127.7	69.5	2.5	1.5	25.4	10.3	127.7	69.4
Meningitis (t)	5.2	2	32.6	15.3	95.5	53.8	4.6	1.3	29.3	10.4	78	25
Peritonitis (t)	4.6	2.3	35	11.7	140.5	64.3	4.2	1.8	36.3	13.6	92.8	39.5
Healthy (P)	4.5	1.8	27.2	12.7	116.1	64	3.3	1.5	25.4	9	110.7	58.7
Gastroschisis (P)	5.2	2.1	33.3	13	108	64.7	3	1.5	25.2	8.2	133.8	79
RDS (P)	4.7	2.3	35.1	16.4	101.5	47.9	2.5	1.7	23.6	10.8	123.2	70.6
IUGR _{asphyxia} (P)	3.3	1.8	25.9	7.6	90.5	50.1	3.4	1.4	33	10.5	140.4	104.3
Tetralogy of Fallot (P)	5.5	2.3	26.5	11.7	84	47.3	3.6	1.7	26.7	9.7	94.6	55.9
IUGR-microcephaly (P)	5	1.4	23.2	5.64	135	60	3.1	1.4	23.5	4	101.9	44.9

4.6 Discussion

An average value is greatly influenced by extreme values, which is not the case for a median. For this reason, we chose to represent our evaluation results in the form of medians and interquartile ranges in this work. This method of display is useful, because it makes possible both global and local dispersion of the data results.

In this study, we first focus our attention on the resonance frequencies (RF1, RF2), the RF1:RF2 ratio, and the RF2 slope, which notify us of changes in the configuration of the vocal tract. To understand the behavior of F0 and (RF1, RF2) in healthy and pathological newborn cries, we concentrated our investigation on the percentage and duration of tuning (P_{TUP} , A_{TUP}), as well as on the transition duration (A_{TRP}) for both RF1 and RF2. These two acoustic phenomena can tell us whether or not a true relationship exists between F0 and the RFs at an early age in the newborn, and also whether or not the various medical conditions we studied influence the extent of these acoustic phenomena.

The association we have established between the characteristics of newborn cries and the pathologies studied is provided in Table 4.5. This association is based on our comparison of the cries of pathological and healthy newborns.

According to (Seifert et al., 2002), in a normal group of children aged 3.8 to 10.2 years, A_{ratio} varies between 0.5 and 0.63. In our study, we found a significant difference in this parameter between healthy newborns aged 1 day to 1 month and pathological newborns of the same age. This is presumably because successful word production requires greater and more rapid changes in vocal tract geometry than cries. In spite of the increase in RF1 and RF2 in healthy preterm newborn cries compared to healthy full-term newborn cries, we found no significant difference in either case for the estimated A_{ratio} . In the case of pathological cries, the values were higher than they were for healthy newborns for some pathologies and lower for others. However, lower values of A_{ratio} are attributable to the decrease in RF1 values, and higher values of A_{ratio} are the result of increased RF1 values. A_{ratio} is related to the position of the tongue in the oral cavity. RF1 is associated with tongue height, and RF2 refers to tongue

advancement (Kim, Kent et Weismer, 2011). These relationships are confirmed by the results of our study. Among the congenital malformations, Lingual frenum(t) cries are characterized by the highest RF1 and A_{ratio} of all the healthy and pathological cries studied. RF2 was also the highest among full-term newborn cries, and its A_{SLP} was the lowest among the newborn cries studied. We can infer from this that a reduced A_{SLP} can be explained by the relative slowness of tongue movement, which is induced by the shortness of the transition and the long RF2 transition duration (A_{TRP}) (Table 4.4). However, in this case, we noted a decrease in A_{TRP} , along with an increase in (P_{TUP}) and (A_{TUP}) for RF1, compared to the values of these parameters in healthy full-term newborns.

In another congenital malformation, Gastroschisis (P), the average RF2 slopes (A_{SLP}) are lower for newborn cries, which contain a high percentage of RF tuning (P_{TUP}) and a long duration (A_{TUP}) of RF1 tuning. In the case of the neurological disease IUGR-microcephaly (P), highest medians of A_{F0} and A_{SLP} were observed and also a high P_{TUP} median and a low A_{TUP} median for RF1. A long transition duration (A_{TRP}) was also noted for RF1. This behavior is possibly the result of very rapid changes in the configuration of the vocal tract (Kim et al., 2009). In the other neurological diseases, in IUGR_asphyxia (P), for example, high A_{RF1} , A_{RF2} , and A_{ratio} values and a slow A_{SLP} and long A_{TRP} were observed. Asphyxia cries are characterized by a high A_{F0} and A_{ratio} , a low A_{SLP} and high A_{TRP} for RF2, and a high A_{TRP} for RF1. In the heart defect diseases, the cries of newborns with Vena cava thrombosis (t) are characterized by a high A_{F0} and the lowest A_{RF1} and A_{RF2} , and an A_{ratio} with a high (A_{SLP}) and a high (A_{TRP}) for RF1 and RF2. Tetralogy of Fallot (P) cries contain the highest percentage of RF1 tuning ($A_{P_{TUP}}$) and a shorter RF1 and RF2 transition duration (A_{TRP}). In this case, A_{ratio} is high with a low A_{SLP} .

However, in the infectious diseases category, Meningitis (t) and Peritonitis (t), A_{RF1} and A_{ratio} are lower. In Meningitis (t) cries, the A_{SLP} and A_{TRP} for both RF2 and RF1 are lower with a high P_{TUP} and A_{TUP} for RF1. Unlike Meningitis (t), Peritonitis (t) cries are characterized by a low A_{RF2} and a high A_{TRP} for RF1. In the respiratory disease category, A_{SLP} is lower in RDS cries, and P_{TUP} for RF1 and A_{TRP} for RF2 are high.

We can surmise from the results of this study that there are three tuning patterns between the RFs and F0 in newborn cries. The first pattern is made up of a high percentage of long tuning duration times with short transition duration times, as in the case of Lingual frenum (t) and Meningitis (t). In the second pattern, we find a high percentage of short tuning duration times with long transition duration times, as in IUGR-microcephaly. In the third pattern, the tuning percentage is low, but with a long duration time and long transition duration time, as for RF2 in healthy full-term newborns.

We note from our results that a decrease in RF1 and RF2 becomes more apparent in healthy newborns with increasing gestational age, which seems more relevant than the age at recording in this case.

To further investigate the influence of gestational age (full-term, preterm) on the characteristics measured, we performed an ANOVA variance analysis for healthy and pathological cries for each characteristic, to determine whether or not they change with gestational age. Our results, shown in Table 4.6, indicate significant statistical differences in all the assessed characteristics between healthy premature cries and healthy full-term cries. Only for the average transition duration of RF2 (A_{TRP}) was no significant difference found based on gestational age. In the case of pathological cries, significant statistical differences ($F=11.304$, $P<0.05$) between the average RF2 (A_{RF2}) of premature pathological cries and full-term pathological cries was noted. For the remaining characteristics, no significant statistical differences were found based on gestational age. We have determined, therefore, that all the characteristics assessed, except for the transition duration of RF2, depend on gestational age in healthy cries. In pathological cries, with the exception of the average RF2, the assessed characteristics do not depend on gestational age, but rather on the pathology itself.

Table 4.5 Characteristics Associated with the Pathologies Studied

Pathologies	Cry characteristics
Lingual frenum(t)	↑RF1, ↑RF1 tuning , ↓RF1 transition, ↑RF2, ↓RF2 tuning , ↑ RF2 transition, ↑ RF1:RF2 ratio, ↓slope RF2
Vena cava thrombosis (t)	↑F0, ↓RF1, ↓RF1 tuning, ↑RF1 transition, ↓RF2, ↓RF2 tuning, ↑RF2 transition, ↓RF1:RF2 ratio, ↑slope RF2
Asphyxia (t)	↑F0, ↑RF1 transition, ↑RF2 transition, ↑RF1:RF2 ratio, ↓slope RF2
Meningitis (t)	↓RF1, ↑RF1 tuning, ↓RF1 transition, ↑RF2 tuning, ↓RF2 transition, ↓RF1:RF2 ratio, ↓slope RF2
Peritonitis (t)	↓RF1, ↓RF2, ↑RF1 tuning, ↑RF1 transition, ↑RF2 tuning, ↓RF2 transition, ↓RF1:RF2 ratio
Healthy (P)	↑RF1, ↑RF2, ↓RF2 tuning, ↑RF2 transition
Gastroschisis (P)	↓RF1, ↑RF1 tuning, ↓RF1:RF2 ratio, ↑RF2 transition, ↓slope RF2
RDS (P)	↓RF1, ↑RF2, ↑RF1 tuning, ↓RF2 tuning, ↑RF2 transition, ↓slope RF2
IUGR_asphyxia(P)	↑RF1, ↓RF1 tuning, ↓RF1 transition, ↑RF2, ↑RF2 tuning, ↑RF2 transition, ↑RF1:RF2 ratio, ↓slope RF2
Tetralogy of Fallot (P)	↑RF1 tuning, ↓RF1 transition, ↑RF2 tuning, ↓RF2 transition, ↑ RF1:RF2 ratio, ↓slope RF2
IUGR-microcephaly (P)	↑F0, ↑RF2, ↑RF1 tuning, ↑RF1 transition, ↓RF2 tuning, ↑ slope RF2

Table 4.6 Analysis of Variance (ANOVA) based on Gestational Age

Characteristic	Healthy		Pathological	
	F	<i>p</i>	F	<i>p</i>
A_{F0}	218.064	0.000	0.052	0.824
A_{RF1}	311.061	0.000	0.683	0.428
A_{RF2}	63.762	0.000	11.304	0.007
A_{ratio}	225.770	0.000	0.143	0.713
A_{SLP}	43.570	0.000	0.036	0.853
$P_{TUP \text{ for RF1}}$	58.332	0.000	0.623	0.448
$A_{TUP \text{ for RF1}}$	25.119	0.000	0.012	0.913
$A_{TRP \text{ for RF1}}$	44.890	0.000	1.304	0.280
$P_{TUP \text{ for RF2}}$	322.175	0.000	0.743	0.409
$A_{TUP \text{ for RF2}}$	536.565	0.000	1.304	0.280
$A_{TRP \text{ for RF2}}$	0.780	0.377	0.458	0.514

4.7 Conclusion

In this paper, a new automated method for newborn cry analysis is presented. This measurement method allows us to associate the most relevant characteristics with pathologies of interest, and also to extend the method to other pathologies.

The goal of this study has been to examine the possibility of classifying pathological and healthy cries on the basis of their acoustic attributes. We focused our investigation on the distribution of RFs and F0 in the cries of newborns 1 day to 1 month old. To identify the early relationship between these two frequencies, the tuning between the RFs and F0, and

the transition duration were evaluated. Variation patterns of the RF1:RF2 ratio and the RF2 slope were also examined, based on the healthy and pathological cries studied.

The calculation of percentile values for each assessed characteristic provides details about its evolution in healthy and pathological cries. The distribution of these acoustic variables could be used to develop a quantitative metric to differentiate pathological cries from healthy ones. We conclude from our results that, even though laryngeal phonation and vocal tract-based articulation are independently controlled systems in newborns, these two systems interact according to the status of the newborn. Also, we infer that the characteristics of cries in the newborns we studied who were suffering from pathologies do not vary with gestational age, but rather mostly according to the pathology itself, which alters vocal folds and vocal tract resonance.

The results of this study suggest that the acoustic measures that we studied (A_{F0} , A_{RF1} , P_{TUP} , A_{TUP} , A_{TRP} , A_{SLP} , A_{ratio}) all show significant differences among diseases. The automated approach presented in this paper for assessing these characteristics can be successfully used to distinguish pathological cries from healthy cries.

Acknowledgments

This work is supported by the Bill & Melinda Gates Foundation through the Grand Challenges Explorations Initiative. Special thanks to Dr. Barrington and members of the Neonatology group at Saint-Justine Hospital in Montreal (QC) for helping us to collect samples for the Infant Cry Database.

CHAPITRE 5

SYNTHÈSE ET CONCLUSION

5.1 Synthèse des articles

L'analyse des signaux des cris des nouveau-nés est importante comme appui au diagnostic clinique. Les travaux de recherches dans ce domaine sont divisés en trois principales catégories : les travaux de recherches sur les raisons des cris, les travaux de recherches sur le développement des cris, et les travaux de recherches sur les relations qui existent entre les maladies et les caractéristiques des cris.

Notre projet de recherche entre dans le cadre de la troisième catégorie. Donc notre premier objectif a été de définir les relations qualitatives et quantitatives qui existent entre les caractéristiques des cris des nouveau-nés et les pathologies étudiées. Le deuxième objectif a été de développer un outil de mesure automatique de ces caractéristiques pour les utiliser en appui du diagnostic médical. Cet outil permettra aussi d'analyser et de caractériser les signaux de cris de nouvelles pathologies non étudiées.

La plus grande partie de notre travail dans cette thèse a été consacrée à la caractérisation des cris qui a été présentée dans les trois articles. Dans notre étude, nous avons utilisé les cris des bébés en bonne santé et des bébés malades, des bébés nés à terme et des bébés prématurés. L'âge des bébés lors de l'enregistrement ne dépasse pas 1 mois, car durant le premier mois les cris sont spontanés sans aucun contrôle volontaire des nouveau-nés.

Les résultats d'estimation des caractéristiques sélectionnés ont été associés à une étude d'analyse statistique dans le but de quantifier, modéliser et optimiser l'influence de certains de ces paramètres sur les performances du système de diagnostic.

La première partie de ce travail (article 1, Annexe I, Annexe II), a été consacrée à la mise au point d'un plan de caractérisation des cris pathologiques. L'étude spectrographique qui a été

réalisée dans ce contexte en utilisant Praat, nous a permis d'identifier plusieurs caractéristiques importantes dans la différenciation des cris pathologiques.

Nous avons commencé par étudier le comportement de la fréquence fondamentale, car elle permet de distinguer les différents types de cris. Ensuite la mesure de la plupart des caractéristiques acoustiques dépend de cette fréquence fondamentale. Des mesures précises de cette fréquence et ses variations dans le temps sont importantes pour obtenir des informations fiables sur l'état de santé des nouveau-nés.

Notre choix pour la mesure de la fréquence fondamentale s'est porté sur l'algorithme SIFT modifié qui permet de dépasser la surestimation et la sous-estimation de cette fréquence.

Les caractéristiques étudiées dans cette partie du travail sont : l'hyperphonation ; l'irrégularité de F_0 ; les hautes fréquences fondamentales. La quantification de ces caractéristiques est faite pour chaque pathologie étudiée en utilisant la moyenne des pourcentages de segments de 20ms qualifiés par ces caractéristiques dans des signaux de cris d'une durée d'une seconde. Nous avons utilisé le STD pour l'analyse des résultats.

L'estimation de ces caractéristiques par pathologie et par âge gestationnel, indique bien que :

- la variation des caractéristiques étudiées dépend de l'état de santé des bébés;
- la moyenne de l'hyperphonation dans les cris des bébés prématurés en bonne santé est la même que dans les cris des bébés sains nés à terme;
- la moyenne de l'irrégularité de F_0 dans les cris des bébés prématurés en bonne santé est légèrement supérieure à l'irrégularité dans les cris des bébés sains nés à terme;
- l'hyperphonation et l'irrégularité de F_0 dans les cris des bébés malades ne dépendent pas de l'âge gestationnel;

- certains cris pathologiques sont caractérisés par une haute fréquence fondamentale.

Dans le contexte de la modélisation et de la reconnaissance des cris des nouveau-nés, des tests de faisabilité et de validation du système de diagnostic en utilisant deux techniques de paramétrisation (MFCC, WPT) des signaux de cris ainsi que les deux caractéristiques étudiées (hyperphonation, l'irrégularité de F0) ont été réalisés : ils sont détaillé en Annexe I et dans l'article 1.

Nous avons évalué les performances du système de diagnostic dans les différents tests en utilisant le taux d'identification correcte. Les résultats trouvés indiquent bien une amélioration de la reconnaissance des cris par pathologie en utilisant les caractéristiques estimées en complément des paramètres MFCC. Ces résultats nous ont encouragés à focaliser nos investigations, dans les articles II et III, sur la caractérisation des cris pathologiques en utilisant d'autres caractéristiques.

Donc, la deuxième partie de ce travail de thèse (article II), a concerné l'estimation de nouvelles caractéristiques comme : la dysphonation, le changement de mode entre la dysphonie et les autres modes ainsi que les sons non voisés.

Là aussi, la quantification des caractéristiques estimées a été effectuée par le calcul du pourcentage moyen des segments (bloc de 1024 échantillons) identifiés par les caractéristiques étudiées dans des signaux de cris d'une seconde.

L'évaluation du bruit durant la phonation (dysphonation) est faite par la technique d'estimation nommée « Normalised Noise Energy ». Deux méthodes d'estimation du bruit ont été testées pour la mesure de la dysphonie :

- le filtrage en peigne adaptatif (ACF);

- la méthode qui se base sur le calcul de l'erreur quadratique moyenne minimale (MMSE) en utilisant la probabilité de présence de la composante périodique dans le signal du cri (PCPP).

Ces deux méthodes ont révélé une surestimation du bruit dans le cas des cris des bébés. Pour remédier à ce problème, nous avons implémenté une nouvelle méthode de mesure du bruit qui s'adapte mieux aux cris des bébés malades.

La nouvelle méthode proposée et utilisée pour l'estimation du bruit et la dysphonation dans les sons de cris se base donc sur la combinaison des deux méthodes précédemment citées. L'estimation du bruit a été initialisée par la méthode ACF. Par la suite, nous avons utilisé MMSE et PCPP pour l'estimation principale du bruit tout le long d'un signal de cri. Les résultats de cette estimation montrent que dans certaines pathologies étudiées, les cris sont nettement dysphoniques par rapport à ceux des bébés en bonne santé.

L'analyse statistique ANOVA qui a été menée sur les caractéristiques estimées indique que le changement de mode et les cris non voisés dépendent de l'âge gestationnel chez les bébés sains alors que chez les bébés malades ils dépendent de la pathologie elle-même. Les résultats font aussi apparaître que la dysphonation que ce soit chez les bébés en bonne santé ou malades, ne dépend pas de l'âge gestationnel.

La troisième partie de cette thèse (article III), est centrée sur la caractérisation des cris des nouveau-nés en utilisant de nouvelles caractéristiques qui n'ont pas été identifiées dans les recherches spectrographiques antérieures sur les cris pathologiques. Ces caractéristiques qualifient principalement le comportement des fréquences de résonances (RF1, RF2) ainsi que l'interaction de ces fréquences avec la fréquence fondamentale F0.

Dans cette partie du travail, nous avons utilisé une autre approche pour analyser les caractéristiques quantifiées. Cette approche se base sur le calcul de l'écart interquartile et la médiane des caractéristiques estimées pour des signaux d'une seconde. Pour l'estimation de

F0, nous avons utilisé le SIFT modifié et pour l'estimation de RF1, RF2, nous avons retenu la méthode de récolte des sommets dans la densité spectrale de puissance autorégressive (AR-PSD).

Cette étude inclut principalement : la moyenne de F0, RF1, RF2; le pourcentage et la durée des interactions (tuning) entre RF1, RF2 et les 10 harmoniques de F0; la durée des transitions entre les tuning; le « slope » de RF2 qui représente la pente de RF2 pendant la durée de transition et enfin le ratio RF1 :RF2.

Les résultats trouvés nous ont permis de comprendre le comportement des cris chez les nouveau-nés. Une interaction précoce entre les fréquences de résonances et la fréquence fondamentale existe déjà chez les nouveau-nés. La durée et le pourcentage de cette interaction « tuning » ainsi que la durée des transitions dépendent de la pathologie du bébé. Le ratio RF1 :RF2 et le « slope » de RF2 eux aussi en dépendent.

L'analyse ANOVA a révélé que l'âge gestationnel chez les bébés en bonne santé, influe sur toutes les caractéristiques étudiées sauf sur les durées de transition pour RF2. Par contre, chez les bébés malades l'âge gestationnel influe seulement sur la moyenne de RF2.

Nous considérons que les résultats obtenus au terme de ces travaux de caractérisation sont satisfaisants. Car les estimations et les mesures effectuées sur les signaux de cris, indiquent bien que les caractéristiques sélectionnées varient selon l'état de santé des nouveau-nés. L'approche proposée pour la quantification de ces caractéristiques permet leur utilisation dans le système de diagnostic médical et facilite l'analyse des signaux de cris.

5.2 Contribution

Cette thèse s'inscrit dans une volonté de développer un système de diagnostic robuste qui permet une détection précoce de l'état de santé du nouveau-né en l'absence de symptômes apparents. L'approche proposée s'appuie sur l'amélioration des performances de reconnaissance des cris pathologiques en utilisant de nouveaux traits acoustiques qui n'ont

pas encore été utilisés dans les recherches précédentes sur la classification des cris. Donc, notre travail a été principalement axé sur la caractérisation des cris pathologiques par rapport aux cris des bébés en bonne santé. Cette caractérisation permet la sélection et l'extraction des paramètres les plus pertinents dans la différenciation des cris pathologiques.

Le travail effectué dans cette thèse a été réalisé dans un grand souci de flexibilité avec possibilité d'utilisation des outils de caractérisation développés dans le cadre de cette recherche, pour d'autres pathologies non étudiées et pour d'autres recherches d'analyses sur les cris.

Notre contribution principale consiste en :

- proposition d'utilisation de nouveaux paramètres acoustiques dans la classification des cris;
- implémentation de méthodes robustes pour l'estimation de ces paramètres;
- développement d'une nouvelle approche pour la quantification de ces paramètres;
- caractérisation quantitative automatique des pathologies étudiées avec possibilité d'extension pour d'autres pathologies non étudiées;
- proposition d'une méthode de fusion de ces paramètres avec d'autres traits acoustiques pour leurs utilisations dans la classification des cris;
- validation des approches proposées.

5.3 Recommandations

En ce qui concerne l'utilisation des caractéristiques acoustiques proposées dans le développement d'un système de diagnostic, les perspectives et les recommandations sont

nombreuses. À l'heure actuelle, de nombreux classificateurs ont vu le jour pour la reconnaissance des cris pathologiques. Ces classificateurs ne tiennent pas compte de l'influence de beaucoup de phénomènes acoustiques qui qualifient la configuration du conduit vocal et des cordes vocales sur la qualité du cri. C'est pour cette raison que nous nous sommes focalisés sur la caractérisation fréquentielle des cris.

Pour une meilleure caractérisation des cris, il est utile :

- d'étendre l'étude sur une plus grande base de données de cris;
- d'étudier et quantifier d'autres caractéristiques temporelles comme la durée du cri et les durées des poses, etc.

Pour une meilleure utilisation de ces caractéristiques dans le système de diagnostic, il est utile de :

- procéder à des tests de classification de cris par âge gestationnel. Car la majorité des caractéristiques étudiées dépendent de l'âge gestationnel des bébés qui sont en bonne santé;
- procéder à des tests de classification en utilisant des méthodes de modélisation des cris autres que le PNN qui a été testé dans cette thèse;
- tester de nouvelles méthodes de fusion, soit des méthodes autres que la concaténation qui a été appliquée dans cette thèse pour combiner les caractéristiques étudiées avec les traits acoustiques des cris.

5.4 Conclusion

Le principal avantage du diagnostic basé sur l'analyse des cris est de permettre la détection et le traitement précoce d'un grand nombre de pathologies sévères. De nombreux nourrissons ne se voient toujours pas offrir des modalités de diagnostic précoce systématique des troubles congénitaux les plus courants. Les méthodes de diagnostic les plus connues sont l'examen clinique, la radiographie (Rx), l'électrocardiogramme (ECG) et l'échocardiographie. L'implantation systématique universelle de ces méthodes présente des coûts élevés que beaucoup de pays en voie de développement ne peuvent supporter.

Dans ce cadre, notre projet de recherche contribue à l'avancement de la recherche dans le développement d'un système de diagnostic robuste et efficace. L'implémentation aisée de ce système à moindre coût permettra la prise en charge précoce d'un grand nombre d'enfants atteints de maladies sévères dans le monde, plus particulièrement dans les pays pauvres.

Le système de diagnostic se base principalement sur la modélisation des traits acoustiques des cris pathologiques. Dans les travaux que nous avons présentés, nous avons réussi à définir de nouveaux traits acoustiques qui permettent d'améliorer la performance de ce système. L'approche de quantification que nous avons adoptée facilite leurs utilisations dans le système de diagnostic médical.

ANNEXE I

NEWBORN'S PATHOLOGICAL CRY IDENTIFICATION SYSTEM

Yasmina Kheddache¹, Chakib Tadj¹

¹Electrical Engineering Department, École de technologie supérieure,

1100 Notre-Dame Ouest, Montréal, Québec, H3C 1K3 Canada

Article publié dans « The 11 th international conference, on information sciences,

Signal Processing and their application», 2-5 July 2012, pp. 1024-1029

Abstract

In this paper we compare the performance of an identification system of the pathological and normal cries of the newborn, using various methods of characterisation of cries. This system is similar to a speaker identification system. It contains two main parts namely a cry signal characterisation and modeling. We used Mel-Frequency Cestrum Coefficients and Mel Frequency Discret Wavelet Coefficients to characterize the newborn cry signals. We also applied Best Structure Abstract Tree algorithm and the Principal Component Analysis to reduce the number of Wavelet packet transform WPT coefficients. In this study a Probabilistic Neural Network classifier is used. The best result obtained is 96.99 % of correct identification using Best Structure Abstract Tree algorithm.

Keywords: Classification, pathologic cry, WPT, PCA, Best abstract Tree.

1. Introduction

The cry of babies is the first sign of life to the birth. It is an early behavior which plays a fundamental part in survival, health and the development of child. It is also less explicit mean of communication to indicate the distress and the pain. The majority of the babies reached of diseases seem in good health at birth (Lester et LaGasse, 2008). Our interest in this research project is to analyze the sounds of the newborn cries and to develop a diagnostic system which allows paediatricians and neonatologists detect pathologies affecting newborns in an early stage. For this purpose, we present in this paper the performance of an identification system of a newborn's pathological and normal cry using various techniques of characterisation of cry signals. Pathologic cries concern different pathologies. In this context, we study four types of analysis to characterize acoustic features of these signals, namely, a) MFCC (Mel Frequency Cepstral Coefficients) is based on Fourier analysis and filter bank on a Mel scale. b) MFDWC (Mel Frequency Discrete Wavelet Coefficients) based on wavelet analysis on a Mel scale. c) MFDWC and the application of BSAT (Best Structure Abstract Tree) algorithm presented in (Tadj et Benlahouar) to modify the structure of the admissible tree and selected nodes. Finally d) application of PCA (Principal Component Analysis) on the MFDWC.

This paper is organised as follows: section 2 presents a review of previous works. Section 3 presents a brief definition of the basic techniques used to design a newborn's pathological cry identification system (PCIS). The next sections present all the experiments and results obtained. We conclude this paper with a discussion and a conclusion.

2. Previous works

Analysis of newborn's cry signals is an important help to clinical diagnosis, because the characteristics of these signals reflect the central nervous system integrity (LaGasse, Neal et Lester, 2005). The researches in this area are divided into three main categories: research on

the reasons of cry, research on the development of the cry and research on the relationship between the disease and the characteristics of the cry.

The classification of the cries falls within the automation of a natural perception of cries and its modelling using advanced signal processing techniques. Cohen et al (Cohen A et Zmora E, 1984) suggested an algorithm based on a simple minimum distance classifier. This algorithm has shown to be successful in classification of hunger and pain cries of healthy full-term infants. Fausto et al. (Galaviz et García, 2005) developed an automatic recognition system for three types of newborn cries (normal, deaf, neonatal asphyxiate). They used the linear predictive coding (LPC) coefficients and MFCC and generic characteristics selection system combined with “feed forward input delay neural network”. They obtained around 96% of cries recognition. Another approach was also presented by Lederman et al. (Lederman et al., 2008); they used LPC and MFCC coefficients and Gaussian mixture models (GMM). It was applied on cry signals of infants with cleft lip and cleft palate. They obtained 91% of recognition in subject- and age-dependent. Hariharan et al. (Hariharan, Yaacob et Awang, 2011) presented another system based on wavelet packet transform and a probabilistic neural network (PNN) to identify neonatal asphyxiate / healthy babies cry. They obtained 99% of recognition.

3. Techniques of parameterization used

The objective of the acoustic analysis is to extract the most representative coefficients of the cry signal. Several methods are used for this purpose.

3.1 MFCC Coefficients

Represent the most used features in the most recent speaker recognition systems (Sturim, Campbell et Reynolds, 2007). They have the advantage to be slightly correlated to each other. They can also represent properly various models of cries (Lederman et al., 2008), because they allow decoupling between the characteristics of the vocal tract and the

characteristics generated by the source of excitation. The cepstrum of the signal is defined as the inverse Fourier transform of the logarithm of the spectral power density.

3.2 Wavelet Transform

The wavelet transform allows the characterisation of a signal in the temporal and frequency domain. We find several types namely: the continuous wavelet transform (CWT), the discrete wavelet transform (DWT), and the wavelet packet transform (WPT). Each transformation has its advantages and its inconveniences. The WPT has the advantage of providing level by level complete decomposition of the studied signal.

3.3 Principal component analysis

The PCA is a method used generally for dimensionality reduction of data. It is based on linear transformation and decomposition of correlated observations into new uncorrelated variables of maximum variance and which are linear combinations of the variables of origins (Jackson, 2005). These new variables are called principal component. In our work, the PCA is used to find a better wavelet basic structure.

3.4 The algorithm of the best structure abstract tree (BSAT)

It is an algorithm which allows building the best structure of abstract tree according to a criterion from entropy or energy (Lester et LaGasse, 2008). We adapted this algorithm and applied for a PCIS as follow:

- pre-processing of the signal;
- decomposition of the signal in wavelet packets;

- calculation for all the nodes of the tree the values of the parameters using the selected criterion (entropy or energies);
- concatenation of all values to obtain a corpus and a tree for each class of cries (healthy babies cries and sick babies cries) (Lester et LaGasse, 2008);
- application of the PCA for a given rate of variance, to seek the nodes (wavelet coefficients) which allow a better structure of tree. This structure is specific to each class of cries.

3.5 PCA-based algorithm to reduce WPT coefficients dimension

This algorithm allows reduction of the WPT coefficients size for two classes of cries using PCA. The main steps of this algorithm are:

- pre-processing of the signal;
- decomposition of the signal in wavelet packets;
- calculation for all the nodes of the tree the values of the parameters using the entropy;
- convert each matrix of WPT (entropy) coefficients specific to a cry sample into only one vector;
- the vectors of WPT (entropy) coefficients of all cries samples are disposed in one matrix $N \times M$ (N : number of WPT coefficients ($63 \times \text{Nb of segments of 20 ms in sample cry}$), M : number of vectors agree to the number of cries sample);
- application of the PCA for matrix of WPT (entropy) coefficient to reduce row dimension of this matrix.

4. Experimental implementation

The adopted methodology for the design of a PCIS is represented on the simplified diagram blocks illustrated in Figure-A I-1. Matlab is used for the development of all blocks of the system. We used also “Wavelet Toolbox” and “Neural Network Toolbox”. Praat is used for the segmentation of the cry signals. This is a freeware program for the analysis and reconstruction of acoustic speech signals.

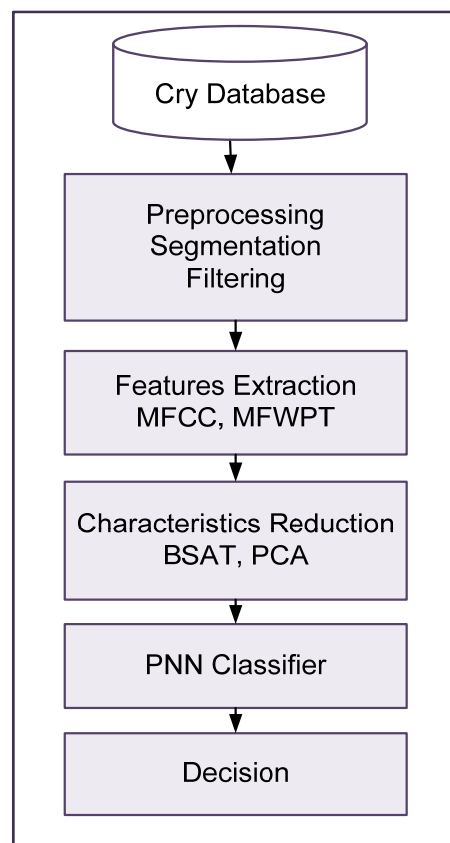


Figure-A I-1 Synoptic of the system

5. Database

Database used contains cry signals of 29 newborn babies. These cries were collected in neonatology department at Saint-Justine hospital in Montreal. They concern full-term and

preterm newborn, healthy newborn and newborn who present some diseases as shown in Table-A I-1. The spectrograms of different pathologic cries are show in Figure-A I-2.

Table-A I-1 Summary of available pathologies

Categories	Pathology
Full-Term	Healthy
	Tetralogy of Fallot
	Thrombosis in the vena cava
	Bovine proteins allergy
Premature	Healthy
	Tetralogy of Fallot
	Anomaly chromosomes X

The conditions in which the cries are registered are: hunger, sampling blood, change of diapers. The recording of cries is done using a small recorder, at a distance of 10 to 30 cm of babies' mouth with a sampling rate of 44.1 kHz. For each baby, three recordings of duration 2 to 3 minutes are made with at least one hour interval after each recording session (over a period of ten days at most). The time, date and gender, date of birth, diagnosis, and reason of cries are noted for each episode of cry.

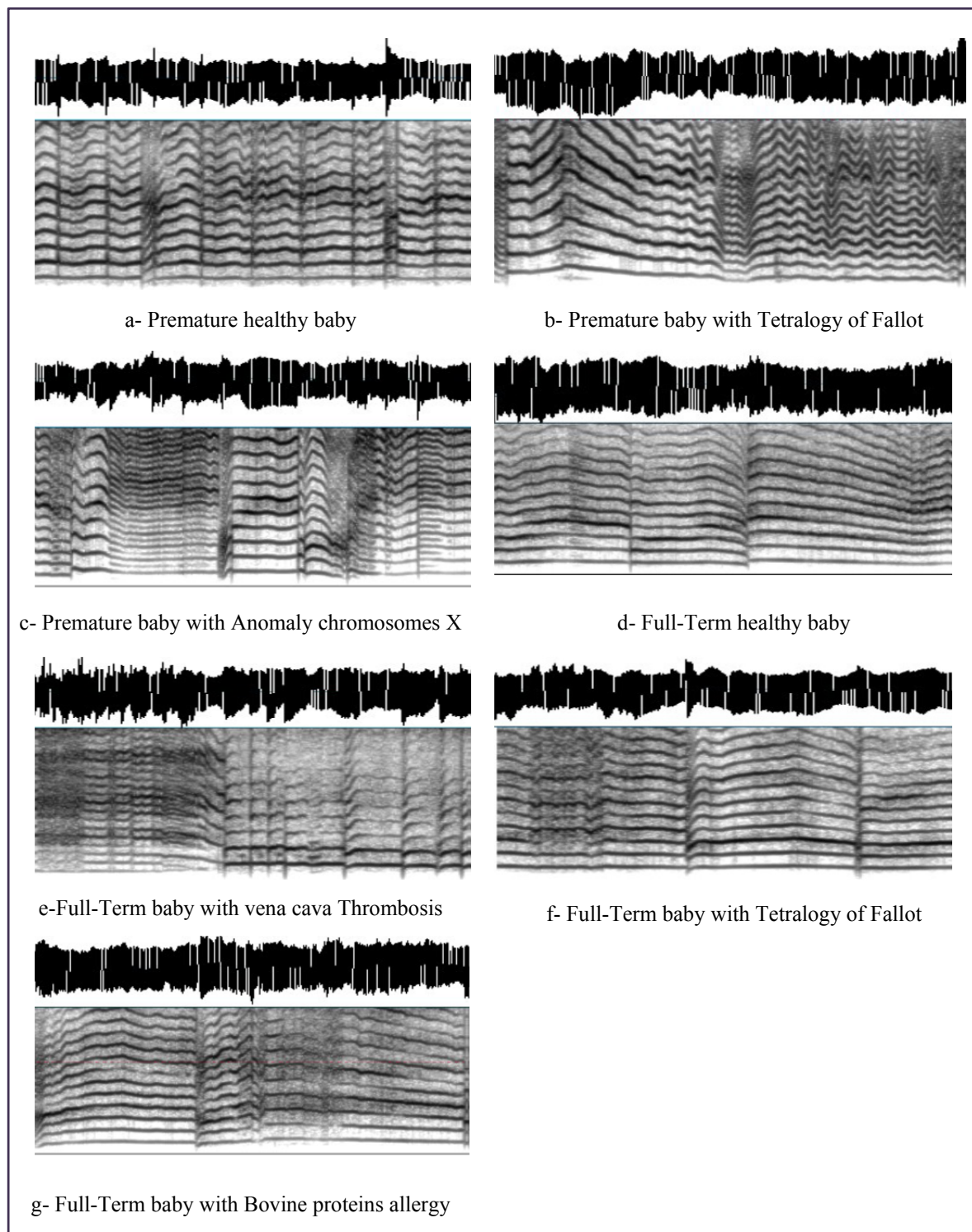


Figure-A I-2 Spectrograms of different pathologic cries

The samples of cries are distributed in two classes, healthy cries and pathological cries. To test the influence of the length's cry samples on the performance of the PCIS, two subsets are used: In the first one, the cry samples were divided into one second segments to create a database of 664 samples of cries (332 healthy and 332 pathological). In the second one, the cry samples were divided into segment of 3 seconds to create 188 samples (94 healthy and 94 pathological). For the test of the identification system $\frac{1}{4}$ of each subset are used for the tests and $\frac{3}{4}$ for training. The choice of 1s and 3s samples is to test the influence of the length's cry samples on the performance of the PCIS. It is important to note that the subset with 1s (664) sample is larger than 3 s subset (188*3).

6. Preprocessing and features characteristic extraction

We used a high-pass filter with a cut-off frequency of 150 Hz to reduce an important part of the ambient noise. We also used a pre-emphasis filter to eliminate the spectral combination from the lips and larynx.

MFCC and WTP coefficients are extracted using 20ms interlaced frame with 10ms recovering, providing 99 windows for one second cry samples and 299 windows for 3 seconds cry samples. Thereafter each frame is multiplied by a Hamming window to keep the continuity of the frame.

6.1 MFCC extraction

We extracted 12 MFCC parameters for each 20ms of a cry signal. We obtained an MFCC matrix of 12 lines $\times$ 99 columns for each sample of 1second and a MFCC matrix of 12 lines $\times$ 299 columns for each sample of 3secondes. The two sets of characteristics once obtained are used as an input of the classifier.

6.2 MFDWC Extraction

The features characteristics are obtained from the wavelet packet decomposition of cry samples. We used Daubechies wavelet of order 8. The overall decomposition in wavelet packet is made for 5 levels. The general structure tree obtained is represented on Figure-A I-3.

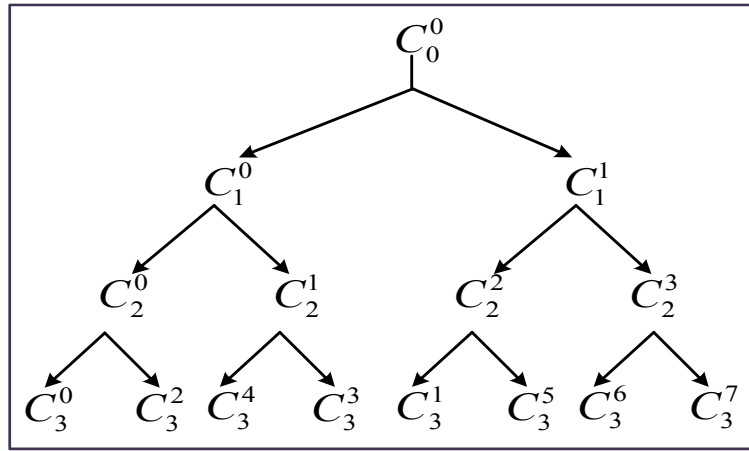


Figure-A I-3 General structure of wavelet packet decomposition on 3 levels

Hence we obtain 63 WPT coefficients for each 20ms segment for all interlaced segments of a cry sample. We obtained a WPT matrix of 99 lines×63 columns for each sample of one second and a WPT matrix of 299 lines× 63 columns for each sample of 3seconds. Thereafter we calculate for each node the value of the entropy and energy. The matrix of entropy and energy obtained for all the cry samples are structured in ways to build two various admissible trees: a) Mel scale admissible tree by cry's samples and b) Mel scale admissible tree by cry's classes.

6.2.1 Mel Scale Admissible Tree by Cry's Samples (TbCS)

The first structure consists of converting each matrix of WPT (entropy) coefficients specific to a cry sample into only one vector of 6237=63×99 dimension for the one second samples and 18837=63×299 for the 3 second samples. This disposition of data is used as input of the

classifier (Lederman et al., 2008). We also apply the PCA on this disposition of data before using them as an input of the classifier.

6.2.2 Mel Scale Admissible Tree by Cry's Classes (TbCC)

The second structure consists of concatenating all energy matrices obtained for the samples of the same class to obtain a tree by class of cries (Tadj et Benlahouar). These data are used to find a better structure of tree adapted to each class according to algorithm BSAT (Tadj et Benlahouar).

7. Application of PCA

To obtain the best base of decomposition, and to search out the best adapted approximation coefficients to represent the signal, we apply the PCA for the two structures of admissible trees found at the preceding step. In the TbCS case, we use the PCA to reduce the dimension of the training and test data separately by using the singular value decomposition (SVD) of the matrix of covariance.

In the TbCC case, we apply the PCA for the (training and test) data. For a given rate of variance, we obtain a specific structure of abstract tree to each class. Once the admissible tree is obtained, we apply the discrete cosine transform to keep 12 coefficients DCT (Discrete Cosine Transform).

8. Probabilistic neural network (PNN)

PNN is generally used for classification problems in the medical domain (Othman et Basri, 2011; Sweeney, Musavi et Guidi, 1994). The use of the PNN classifier is motivated by its speed and simplicity of the training process. There is no need for weight adaptation. PNN is able to learn new incoming training data without having to repeat the whole training process. These characteristics are ideal for real time applications (Sweeney, Musavi et Guidi, 1994).

The PNN has three layers: the Input layer, Radial Basis Layer and the Competitive Layer. The second layer which is a radial network gives information on the resemblance between the input data and the data file used at the training phase. The third layer produced output vector of probability. The assigned class is the one with the highest probability (Demuth, Beale et Hagan, 1992).

Classifier PNN is obtained using function `newpnn ()` in Matlab. The inputs of classifier PNN in our study are the vectors of the characteristics obtained at the preceding step, according to various cases presented previously.

9. Experimental results

To evaluate the efficiency of the studied system, four experiments were performed. The first experiment consists in using MFCC coefficients as vectors of features characteristic for the training and the test of the two classes of cries (pathologic and normal).

The second experiment consists in using WPT (entropy) coefficients as vectors of features characteristic for the training and the test of the two classes of cries.

The third experiment consists in applying the PCA to reduce the dimension of the WPT (entropy) coefficients of training and test data. The WPT (entropy) coefficients obtained are used as vectors of features characteristic for the training and the test of the two classes of cries.

The fourth experiment consists in applying BSAT algorithm to reduce the dimension of WPT (energy) coefficients of training and test for the two classes of cries.

The test of studied system was performed with cross-validation method. The results of the correct identification rate (Overall accuracy) for all experiments are shown in Table-A I-3 for the 1second samples and Table-A I-4 for the 3second samples. To evaluate the

performance of this system, other measures such as specificity, sensitivity are also calculated from true positive (TP), true negative (TN), false positive (FP), and false negative (FN) as presented in Table-A I-2 (Hariharan, Yaacob et Awang, 2011).

$$\text{Sensitivity} = \frac{TP}{(TP + FN)}$$

$$\text{Specificity} = \frac{TN}{(TN + FP)}$$

$$\text{Overall accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Table-A I-2 Confusion matrix

Actual classification	Predicted classification	
	Pathological	Normal
Pathological	TP	FN
Normal	FP	TN

Table-A I-3 Results for various experiments (1second samples)

	Sensitivity	Specificity	Overall accuracy
MFCC	80.72	86.75	83.73
MFDWC	56.10	89.02	72.56
MFDWC + PCA	68.57	84.29	76.43
MFDWC + BSAT	93.97	100	96.98

Table-A I-4 Results for various experiments (3second samples)

	Sensitivity	Specificity	Overall accuracy
MFCC	78.31	69.88	74.10
MFDWC	81.82	90.91	86.36
MFDWC + PCA	81.82	81.82	81.82
MFDWC + BSAT	100	83.91	91.95

10. Discussion

The results obtained concern mainly the use of PCA and BSAT algorithm to reduce the dimension of WPT coefficients. MFCC parameters are also used to compare the performance of the PCIS. The best performance with MFCCs is obtained for one second samples. The use of WPT (entropy and energy) coefficients as features characteristics is tested in several ways. By using WPT (entropy) coefficients directly without any reduction, then best performance is 86.36% for the 3second samples. The second test corresponds to application of PCA to reduce WPT (entropy) coefficients. This test gives a better identification rate for the two classes of cries for 3second samples. The BSAT Algorithm is also tested. The test of BSAT algorithm corresponds to its application for the training and test data for each class of cries. The correct identification rate found is 96.99% for the 1second samples.

11. Conclusion

In this experimental work, we study the performance of a PCIS using a PNN classifier and several methods of characterisation of newborn cries. The various methods tested are MFCC, MFDWC as well as the application of the PCA and BSAT algorithm to reduce the dimension of WPT (entropy and energy) coefficients.

The results from this study showed that the performance of PCIS depends on length of cry sample and on the technique of parameterization used. We also deduce from the experimental results that the use of PCA does not necessarily improve the performance of a PCIS. In fact, it sometimes performs better and sometimes worse according to the length of cry sample. Application of PCA-based BSAT algorithm generated better results.

Result obtained using MFDWC and PNN classifier is different compared with the result presented in the reference (Hariharan, Yaacob et Awang, 2011). It is important to note that in this work, we are dealing with full-term and premature babies, both with and without pathologies. This increases significantly the complexity of a PCIS. However in (Hariharan, Yaacob et Awang, 2011), cry signals studied (neonatal asphyxiate / healthy babies cry) contain significant differences in the acoustic signals (Michelsson, 1971) and that make the identification process easier.

The use of the wavelet coefficients and the PCA are very effective for the development of a pathological cry detection system. This work is still in progress. The results obtained by using WPT coefficients as features characteristics of newborn cries can be improved using other tree structures. We also expect to improve the results by using other classifier and other acoustic characteristics in addition to WPT coefficients and PCA.

In order to achieve a reliable evaluation of the classification performances we are still recording a larger database with a greater variety of pathologies and more subjects for each pathology.

Acknowledgements

This research work has been made possible through the funding provided by Bill & Melinda Gates Foundation under the Grand Challenges Explorations program.

ANNEXE II

CHARACTERISATION OF PATHOLOGIC CRIES OF NEWBORNS BASED ON FUNDAMENTAL FREQUENCY ESTIMATION

Yasmina Kheddache¹, Chakib Tadj¹

¹Electrical Engineering Department, École de technologie supérieure,

1100 Notre-Dame Ouest, Montréal, Québec, H3C 1K3 Canada

Article publié dans le journal « Engineering », vol.5, octobre 2013, pp. 272-276

ABSTRACT

The majority of the sick babies seem in good health at birth and early diagnosis of the various pathologies that can afflict newborn is crucial. For this reason we introduce in this paper the use of the fundamental frequency and the different modes of cries as hyperphonation, phonation and high pitched cries to differentiate the pathological cries from the cries of the healthy babies. The automatic estimation of these characteristics enabled us to establish a quantitative characterization of healthy newborn cries and pathologic newborn cries. The results obtained agree with the spectrographic analysis carried out on the healthy and pathologic cries of the newborns.

Keywords: Newborns' Cry; Fundamental Frequency; Cry Modes

1. Introduction

For many researchers, the fundamental frequency F_0 which represents the speed of opening and closure of the vocal folds (Lester et LaGasse, 2008) is considered as the most important in the perception of the infant's cries by adults. Generally the increase of F_0 is associated with a negative perception of the cries. The cries with a high F_0 are valued as aversive, sign of emergency, distress and disease (LaGasse, Neal et Lester, 2005).

Cry analysis can be an additional tool in the clinical diagnosis of newborn. The fundamental frequency of newborn cry has been found to be a discriminative characteristic in several medical problems (Lester et LaGasse, 2008) .

Cry analysis presented in this paper falls within the research on the relationship between the disease and the characteristics of the cry. Newborns do not have phonatory control due to neurological immaturity at early age (Verduzco-Mendoza et al., 2009). The main objective of this work is to analyze cries of healthy and newborn with different categories of diseases in order to evaluate a fundamental frequency of these cries. We also establish quantitative relationships between the different modes of cries and studied pathology. The SIFT algorithm 'Simple Inverse Filtering Tracking' is used for estimation of the fundamental frequency because the performance of this algorithm has been tested on a real database of cries by (Lederman, 2010; Manfredi et al., 2009).

This paper is organized as follows: Section 2 presents definitions of fundamental frequency of cries. Section 3 presents previous works. Section 4 presents our adopted methodology for automatic estimation of F_0 , phonic, hyperphonic and high pitched segments for healthy and pathological cries. Section 5 presents the database used in this study. The next sections present the results obtained and we conclude this paper with a conclusion.

2. Fundamental Frequency of Cries

Fundamental frequency (F0) is the average vibratory frequency (in Hz) of the vocal folds. The sound spectrogram displays the voiced tone as black lines. The fundamental frequency is the lowest line and its harmonics appear above it as parallel lines. The maximal fundamental frequency corresponds to the highest measurable point of the fundamental frequency and the minimal fundamental frequency corresponds to the lowest measurable point of the fundamental frequency on the spectrogram (Wasz-Höckert, Michelsson et Lind, 1985).

Lester et al. (LaGasse, Neal et Lester, 2005) defines three identifiable modes of cries due to vocal cord vibrations: 1) Basic cry or phonation with F0 350 - 750 Hz; 2) Cry with high fundamental frequency F0 (750 - 1000 Hz) or Hyperphonation F0 (1000 - 2000 Hz) and 3) Noisy, turbulent or dysphonic cries.

3. Previous Works

Several authors focused their investigations on the healthy children since birthday to one year age. Their work is based on observation and spectrographic analysis of healthy and pathological cries signals of newborns. They found that there is no significant difference between the preterm newborn cries and the full-term newborn cries. They also found that there is not variation of F0 compared to the sex and the gestational age (Verduzco-Mendoza et al., 2009). In the case of the newborn with various pathologies, it was shown that the level of the cries is high and presents a weak punctuation, high irritability and low physiological stability (Lester et LaGasse, 2008), (Verduzco-Mendoza et al., 2009). The authors in this research area show that the cries of healthy full-term newborn are characterized by an F0 varying between 400 and 600 Hz and averaging 450 Hz (Manfredi et al., 2009), (Michelsson et al., 2002). The cries of healthy premature newborns can be more intense or similar to the cries of full-term newborns (Lester et LaGasse, 2008). The cries of newborns with neurological disturbances exhibit auditory abnormalities Wasz-Höckert, Michelsson et Lind,

1985). Hyper- phonic cries and very high-pitched cries are associated with neurological problems (LaGasse, Neal et Lester, 2005) and neonatal risk (Manfredi, Tocchioni et Bocchi, 2006).

4. Methodology

Our methodology for the characterization of healthy and pathologic cries based on F0 estimation is represented on the simplified diagram blocks illustrated in Figure-A II-1. The following approach was taken when estimating fundamental frequency and the different frequency modes of cries:

- recording of healthy and pathologic cries of newborns;
- noise filtering and segmenting recordings into useful and non-useful portions;
- estimation of F0 in short intervals typically of 20 ms interlaced frame with 10 ms recovering using SIFT algorithm (Simple Inverse Filtering Tracking) (Lederman, 2010), (Markel, 1972). The main steps of this algorithm are presented in Figure-A II-2;
- estimating of phonic and hyperphonic segments, as well as segments with high F0 as defined in Table-A II-1 for all cries studied;
- identification of relationships between cries characteristics and studied pathologies.

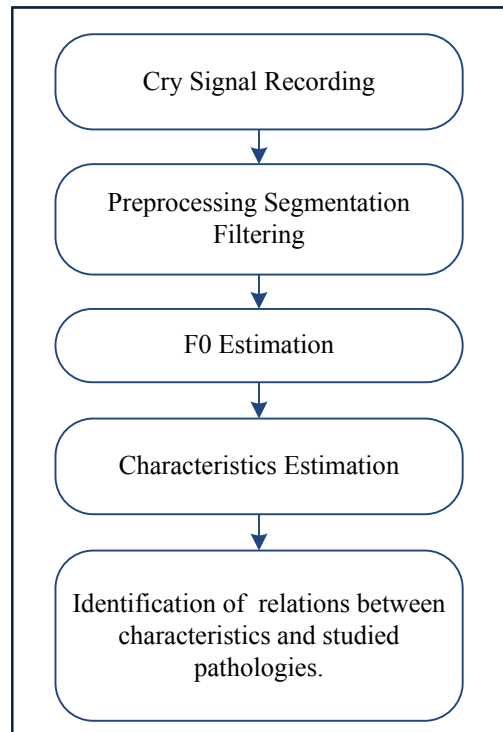


Figure-A II-1 Newborn cries characterization based on F0 estimation

Table-A II-1 Measured characteristics of cries

Characteristics	Definition
Hyperphonation	The average percentage of 20ms blocks having a F0 >1000 Hz.
Phonation	The average percentage of 20ms blocks having a F0 <750Hz
High-pitched	The average percentage of 20ms blocks having a F0 in the 750-1000Hz range

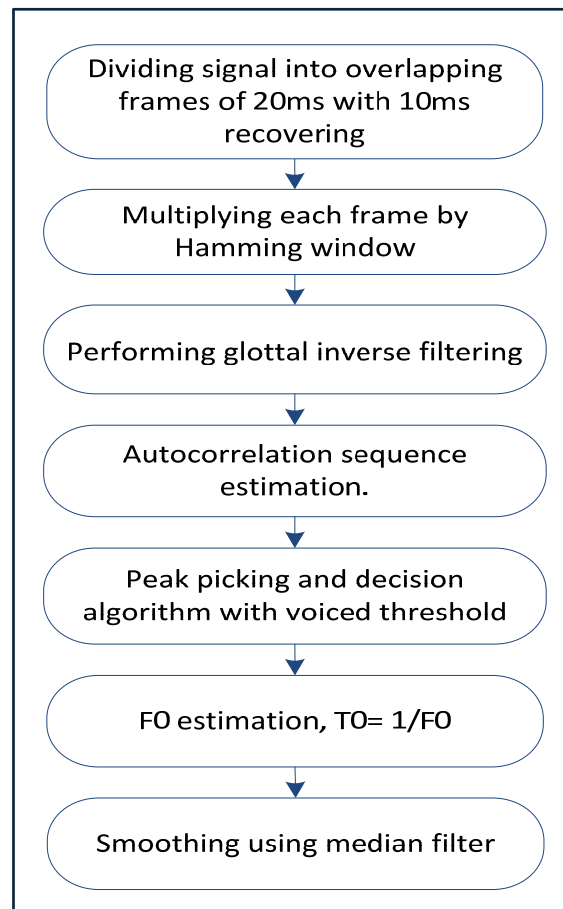


Figure-A II-2 The main steps of SIFT algorithm

5. Database: Cry recording

The database used contains 2800 cry samples of 1s duration from 48 of newborn babies. 1774 from healthy newborn (among them 764 are premature) and 1010 from newborn who present some diseases (among them 628 are premature) as shown in Table A II.2.

These cries were collected with the aid of medical collaborators of neonatology department at Saint-Justine Hospital in Montreal. They concern full-term and preterm newborn aged 1 to 30 days. The conditions in which the cries are registered are: hunger, sampling blood, change of diapers. The recording of cries is done using a small recorder, at a distance of 10 cm of babies' mouth with a sampling rate of 44.1 kHz.

For each baby, three recordings of duration 2 to 3 minutes are made with at least one hour interval after each recording session (over a period of ten days at most). The time, date and gender, date of birth, diagnosis, and reason of cries are noted for each episode of cry.

Table-A II-2 Pathologies studied

Category	Pathology	Sample size
Full-term newborn (t)	Healthy	1010
	Thrombosis in the vena cava	77
	Meningitis	115
	Asphyxia	190
Preterm newborn (P)	Healthy	764
	IUGR- Microcephaly (Intra-Uterine Growth Retardation)	78
	Gastroschisis	280
	RDS (respiratory distress Syndrome)	270

6. Results and analysis

We developed our own measuring tool using Matlab. This tool estimates F0 as well as the average percentages of phonic, hyperphonic and high pitched segments. Examples of spectrograms and F0 estimation are reported in Figure-A II-3 for two pathologic cries (asphyxia and gastroschisis) and also for a healthy newborn. Notice high F0 and more irregular spectrograms and F0 for pathologic cries compared to healthy cry. The range of F0 for pathologic cries is 150 Hz - 1600 Hz and for healthy cry around 450 Hz. The estimated values of frequency characteristics listed in Table 1 for healthy and pathologic newborn's cry are presented in Figures-A II-4 and A II-5 according to pathologies studied and gestational age of babies.

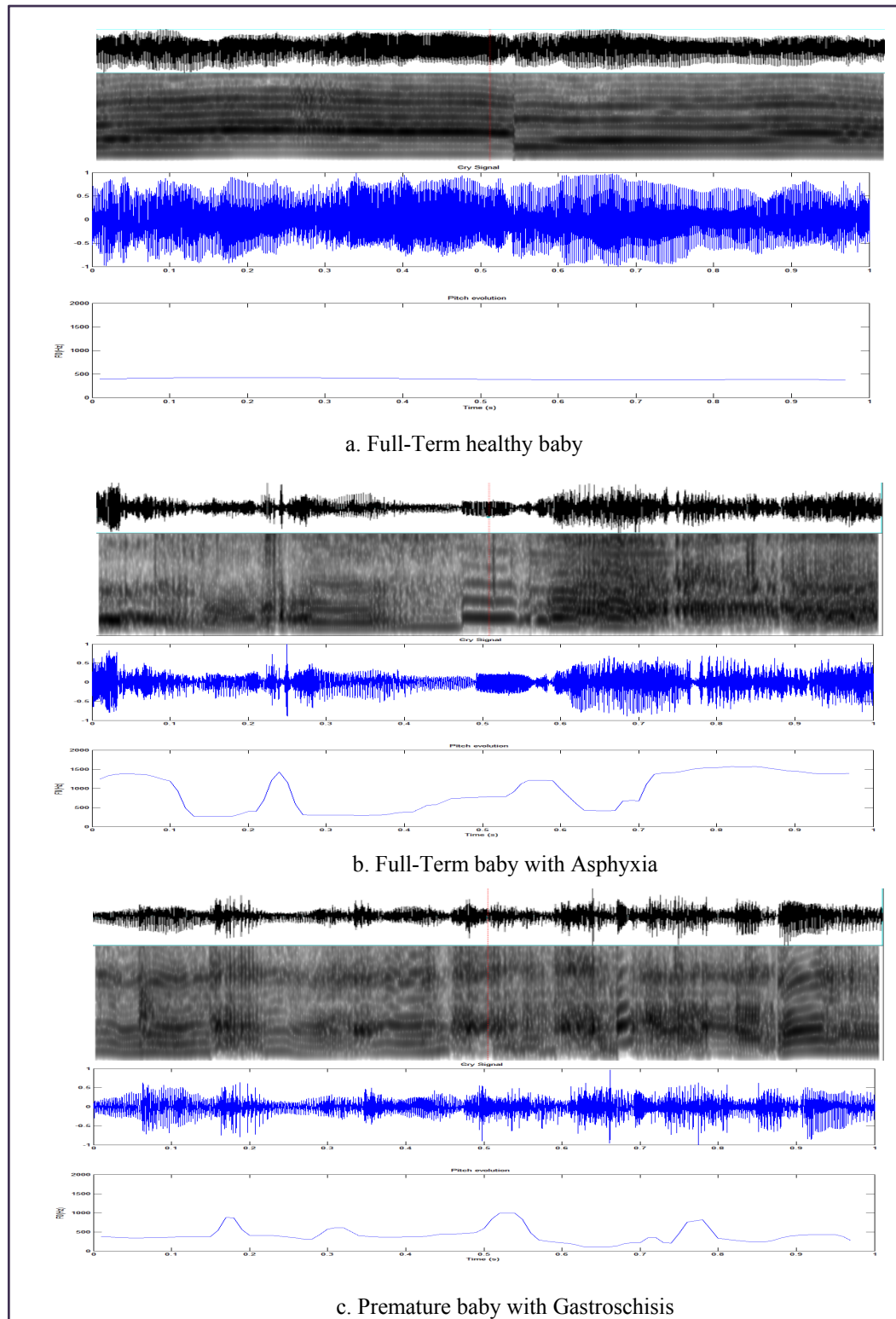


Figure-A II-3 Spectrograms and F0 estimation for healthy and pathologic cries

The results from this study show in Figure A II.4 that the cries of healthy full-term or preterm newborns are rather phonics with a presence of 89% of segments with a $F0 < 750$ Hz and round 9% of segments with high $F0 > 750$ Hz. Figure A II.5 shows that the cries of healthy full-term or preterm newborns contain round 6% of Hyperphonic segment. On the other hand, the cries of the sick newborns contain more pitched and hyperphonic segment compared to the healthy newborns. According to the pathologies studied, the percentage varied from 11% to 28% of segments with a $F0 > 750$ Hz. These percentages vary from one pathological condition to another, and are dependant to level of prematurity.

We notice in Figures A II.4 and A II.5, which the cries of newborns with asphyxia contain 28% of high pitched segments ($F0 > 750$ Hz) among them 25% are hyperphonic. In the case of meningitis, the cries contain respectively 11% and 8% of high pitched segments ($F0 > 750$ Hz) and hyperphonic segments.

The cries of newborns with microcephaly associated with IUGR present 25% of high pitched segments ($F0 > 750$ Hz). These cries are more hyperphonic with 17% of segment with $F0 > 1000$ Hz. The cries of newborns with congenital malformation gastroschisis contain 13% of high pitched segments with $F0 > 750$ Hz. The percentage of hyperphonation is 10% of segment with $F0 > 1000$ Hz. The percentage of hyperphonation is 10% of segment with $F0 > 1000$ Hz. The cries of newborns with heart defect (Thrombosis 750 Hz, among them 14% are hyper-phonic. The cries of newborn with RDS (Respiratory Distress Syndrome) include 15% of segments with $F0 > 750$ Hz among them 11% are hyper-phonic.

The highest average percentage of hyperphonic cries is found in the category of neurological problems (asphyxia, microcephaly-IUGR) and also the highest average percentage of high pitched cries with $750 < F0 < 1000$ Hz is found in the category of neurological problems (microcephaly-IUGR).

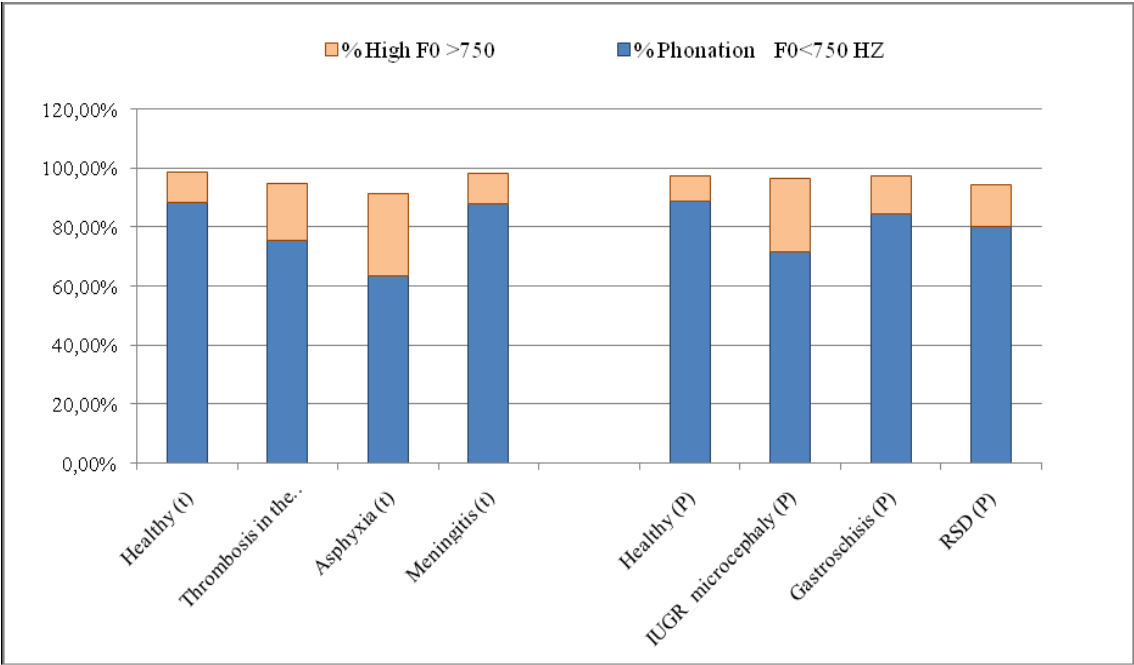


Figure-A II-4 The average percentage of phonic and high pitched segments (F0 > 750 Hz) by pathologies

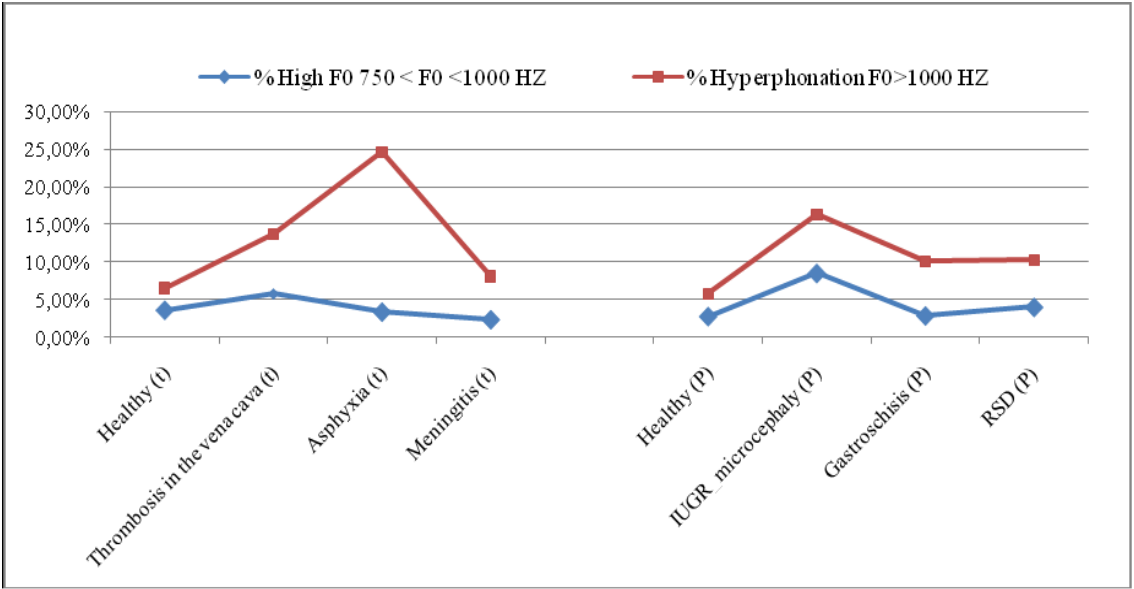


Figure-A II-5 The average percentage of hyperphonic (F0 > 1000 Hz) and high pitched segments (750 < F0 < 1000 Hz) by pathologies

7. Conclusions

This study mostly deals with characterization of healthy and pathologic newborn's cry according to the modes of cry as hyperphonation, phonation, and high-pitched cry. The conclusion from this work is that there are clear differences in frequency characteristics of healthy and pathologic cries. This difference can be used in pediatric diagnosis. The results obtained are very encouraging because they are consistent with spectrographic studies of crying newborns (Lester et LaGasse, 2008), (LaGasse, Neal et Lester, 2005; Michelson Katarina 2007).

This work is still in progress. The results obtained by using different modes of cry as discriminative characteristics for healthy and pathologic newborn's cry can be improved using other acoustic characteristics cited in the literature in association with severe medical conditions. More significant results can be expected using the developed tool on a larger database with a greater variety of pathologies and more subjects for each pathology.

Acknowledgements

We would like to thank Dr. Barrington and members of neonatology group of Saint-Justine Hospital in Montreal (QC) for their dedication of the collection of the Infant's cry database. This research work has been funded by a grant from the Bill & Melinda Gates Foundation through the Grand Challenges Explorations Initiative.

BIBLIOGRAPHIE

- 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.
- Alaie, Hesam Farsaie, et Chakib Tadj. 2012. « Cry-based classification of healthy and sick infants using adapted boosting mixture learning method for gaussian mixture models ». *Modelling and Simulation in Engineering*, vol. 2012, p. 55.
- Amaro-Camargo, Erika, et CarlosA Reyes-García. 2007. « Applying Statistical Vectors of Acoustic Characteristics for the Automatic Classification of Infant Cry ». In *Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues*, sous la dir. de Huang, De-Shuang, Laurent Heutte et Marco Loog. Vol. 4681, p. 1078-1085. Coll. « Lecture Notes in Computer Science »: Springer Berlin Heidelberg. < http://dx.doi.org/10.1007/978-3-540-74171-8_109 >.
- Atal, B. S. 1974. « Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification ». *The Journal of the Acoustical Society of America*, vol. 55, n° 6, p. 1304-1312.
- Bogert B.P, M.J.R. Healy, J.W. Tukey. 1963. « The quefrency alanalysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross cepstrum, and saphe cracking ». In *Time Series Analysis*, sous la dir. de M. Rosenblatt, Ed. p. 209–243.
- Buder, Eugene H., Lesya B. Chorna, D. Kimbrough Oller et Rebecca B. Robinson. 2008. « Vibratory Regime Classification of Infant Phonation ». *Journal of Voice*, vol. 22, n° 5, p. 553-564.
- Cano Ortiz, SergioD, Daniell Escobedo Beceiro et Taco Ekkel. 2004. « A Radial Basis Function Network Oriented for Infant Cry Classification ». In *Progress in Pattern Recognition, Image Analysis and Applications*, sous la dir. de Sanfeliu, Alberto, JoséFrancisco Martínez Trinidad et JesúsAriel Carrasco Ochoa. Vol. 3287, p. 374-380. Coll. « Lecture Notes in Computer Science » : Springer Berlin Heidelberg. < http://dx.doi.org/10.1007/978-3-540-30463-0_46 >.
- Cano, Sergio, Israel Suaste, Daniel Escobedo, CarlosA Reyes-García et Taco Ekkel. 2006. « A Combined Classifier of Cry Units with New Acoustic Attributes ». In *Progress in Pattern Recognition, Image Analysis and Applications*, sous la dir. de Martínez-Trinidad, JoséFrancisco, JesúsAriel Carrasco Ochoa et Josef Kittler. Vol. 4225, p. 416-425. Coll. « Lecture Notes in Computer Science »: Springer Berlin Heidelberg. < http://dx.doi.org/10.1007/11892755_43 >.

- Cecchini, Marco, Carlo Lai et Viviana Langher. 2010. « Dysphonic newborn cries allow prediction of their perceived meaning ». *Infant Behavior and Development*, vol. 33, n° 3, p. 314-320.
- Cohen A, et Zmora E. 1984. « Automatic classification of infants' hunger and pain cry ». In *Proceedings of the international conference on digital signal processing*. (1984). Amsterdam.
- Corwin, Michael J, Herbert Kayne, Barry M Lester, Carol Sepkoski, Sarah McLaughlin et Howard L Golub. 1992. « Effects of in utero cocaine exposure on newborn acoustical cry characteristics ». *Pediatrics*, vol. 89, n° 6, p. 1199-1203.
- Corwin, Michael J, Barry M Lester, Carol Sepkoski, Mark Peucker, Herbert Kayne et Howard L Golub. 1995. « Newborn acoustic cry characteristics of infants subsequently dying of sudden infant death syndrome ». *Pediatrics*, vol. 96, n° 1, p. 73-77.
- Delahaie, Marc (1-90). 2009. *L'évolution du langage de l'enfant: De la difficulté au trouble* Jeanne Herr.
- Demuth, Howard, Mark Beale et Martin Hagan. 1992. « Neural Network Toolbox™ 6 ». *User Guide, COPYRIGHT*, vol. 2008.
- Feng-Ling, Yuan, et Huang Xiang-Nian. 2010. « Infant cry recognition based on feature extraction ». In *Information Networking and Automation (ICINA), 2010 International Conference on*. (18-19 Oct. 2010) Vol. 2, p. V2-166-V2-169.
- Fisichelli, Vincent R., et Samuel Karelitz. 1966. « FREQUENCY SPECTRA OF THE CRIES OF NORMAL INFANTS AND THOSE WITH DOWN'S SYNDROME ». *Psychonomic Science*, vol. 6, n° 4, p. 195-196.
- Fort, A., A. Ismaelli, C. Manfredi et P. Bruscaglioni. 1996. « Parametric and non-parametric estimation of speech formants: application to infant cry ». *Medical Engineering & Physics*, vol. 18, n° 8, p. 677-691.
- Fort, Ada, et Claudia Manfredi. 1998. « Acoustic analysis of newborn infant cry signals ». *Medical engineering & physics*, vol. 20, n° 6, p. 432-442.
- Veillez sélectionner un type de document autre que « Generic » afin de faire afficher la référence bibliographique.
- Frazier, Ronald H, S Samsam, LD Braidia et AV Oppenheim. 1976. « Enhancement of speech by adaptive filtering ». In *Proc. ICASSP*. p. 251-253.

- Galaviz, OrionFaustoReyes, et CarlosAlbertoReyes García. 2005. « Infant Cry Classification to Identify Hypo Acoustics and Asphyxia Comparing an Evolutionary-Neural System with a Neural Network System ». In *MICAI 2005: Advances in Artificial Intelligence*, sous la dir. de Gelbukh, Alexander, Álvaro Albornoz et Hugo Terashima-Marín. Vol. 3789, p. 949-958. Coll. « Lecture Notes in Computer Science »: Springer Berlin Heidelberg. < http://dx.doi.org/10.1007/11579427_97 >.
- Gerkmann, T., et R. C. Hendriks. 2011. « Noise power estimation based on the probability of speech presence ». In *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*. (16-19 Oct. 2011), p. 145-148.
- Gerkmann, T., et R. C. Hendriks. 2012. « Unbiased MMSE-Based Noise Power Estimation With Low Complexity and Low Tracking Delay ». *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, n° 4, p. 1383-1393.
- Golub, Howard L, et Michael J Corwin. 1982. « Infant cry: a clue to diagnosis ». *Pediatrics*, vol. 69, n° 2, p. 197-201.
- Hariharan, M., J. Saraswathy, R. Sindhu, Wan Khairunizam et Sazali Yaacob. 2012. « Infant cry classification to identify asphyxia using time-frequency analysis and radial basis neural networks ». *Expert Systems with Applications*, vol. 39, n° 10, p. 9515-9523.
- Hariharan, M., Sazali Yaacob et Saidatul Ardeenaawatie Awang. 2011. « Pathological infant cry analysis using wavelet packet transform and probabilistic neural network ». *Expert Systems with Applications*, vol. 38, n° 12, p. 15377-15382.
- Hendriks, R. C., R. Heusdens et J. Jensen. 2010. « MMSE based noise PSD tracking with low complexity ». In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. (14-19 March 2010), p. 4266-4269.
- Hirschberg, Jenö. 1999. « Dysphonia in infants ». *International Journal of Pediatric Otorhinolaryngology*, vol. 49, Supplement 1, n° 0, p. S293-S296.
- Hirschberg, Jenö, P. H. Dejonckere, Minoru Hirano, Kazunori Mori, H. J. Schultz-Coulon et Karel Vrtička. 1995. « Voice disorders in children ». *International Journal of Pediatric Otorhinolaryngology*, vol. 32, Supplement, n° 0, p. S109-S125.
- Irwin, Julia R. 2003. « Parent and Nonparent Perception of the Multimodal Infant Cry ». *Infancy*, vol. 4, n° 4, p. 503-516.
- J. D. Newman. 1994. *Crying in Infants*, Second Edition. Philadelphia: Lippincott Company,, 761-763 p.
- Jackson, J Edward. 2005. *A user's guide to principal components*, 587. Wiley. com.

- Jam, M. M., et H. Sadjedi. 2009. « A System for Detecting of Infants with Pain from Normal Infants Based on Multi-band Spectral Entropy by Infant's Cry Analysis ». In *Computer and Electrical Engineering, 2009. ICCEE '09. Second International Conference on.* (28-30 Dec. 2009) Vol. 2, p. 72-76.
- Jolliffe, Ian. 2005. *Principal component analysis*. Wiley Online Library.
- K. Michelson, H. Todd de Barra, O. Michelson. 2007. « Sound spectrographic cry analysis and mothers perception of their infant's crying ». In *Focus on Nonverbal communication research*, sous la dir. de R.Lewis, Finley. p. 31-64. Nova Science.
- Kasuya, H., S. Ogawa et Y. Kikuchi. 1986. « An adaptive comb filtering method as applied to acoustic analyses of pathological voice ». In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '86.* (Apr 1986) Vol. 11, p. 669-672.
- Kasuya, Hideki, Shigeki Ogawa, Kazuhiko Mashima et Satoshi Ebihara. 1986. « Normalized noise energy as an acoustic measure to evaluate pathologic voice ». *The Journal of the Acoustical Society of America*, vol. 80, n° 5, p. 1329-1334.
- Kheddache, Y. , et C. Tadj. 2013a. « Acoustic measures of the cry characteristics of healthy newborns and newborns with pathologies ». *Journal of Biomedical Science & Engineering*, vol. 6, n° 8, p. 798-804.
- Kheddache, Y. , et C. Tadj. 2013b. « Characterization of Pathologic Cries of Newborns Based on Fundamental Frequency Estimation ». *Engineering*, vol. 5, p. 272.
- Kheddache, Y. , et C. Tadj. 2013c. « Frequential characterization of healthy and pathologic newborn cries. ». *American Journal of Biomedical Engineering*, vol. 3, p. 182-193.
- Kheddache, Y., et C. Tadj. 2012. « Newborn's pathological cry identification system ». In *Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on.* (2-5 July 2012), p. 1024-1029.
- Kim, Y., G. Weismer, R. D. Kent et J. R. Duffy. 2009. « Statistical Models of F2 Slope in Relation to Severity of Dysarthria ». *Folia Phoniatrica et Logopaedica*, vol. 61, n° 6, p. 329-335.
- Kim, Yunjung, Raymond D. Kent et Gary Weismer. 2011. « An Acoustic Study of the Relationships Among Neurologic Disease, Dysarthria Type, and Severity of Dysarthria ». *Journal of Speech, Language, and Hearing Research*, vol. 54, n° 2, p. 417-429.

- LaGasse, L. L., A. R. Neal et B. M. Lester. 2005. « Assessment of infant cry: acoustic cry analysis and parental perception ». *Ment Retard Dev Disabil Res Rev*, vol. 11, n° 1, p. 83-93.
- Laitman, Jeffrey T., Raymond C. Heimbuch et Edmund S. Crelin. 1978. « Developmental change in a basicranial line and its relationship to the upper respiratory system in living primates ». *American Journal of Anatomy*, vol. 152, n° 4, p. 467-482.
- Laitman JT, Crelin ES. 1980. « Developmental change in the upper respiratory system of human infants ». *Perinatol Neonatol*, vol. 4, p. 16 –22.
- Lederman, Dror. 2010. « Estimation of Infants' Cry Fundamental Frequency using a Modified SIFT algorithm ». *arXiv preprint arXiv:1009.2796*.
- Lederman, Dror, Ehud Zmora, Stephanie Hauschildt, Angelika Stellzig-Eisenhauer et Kathleen Wermke. 2008. « Classification of cries of infants with cleft-palate using parallel hidden Markov models ». *Medical & Biological Engineering & Computing*, vol. 46, n° 10, p. 965-975.
- Leger, Daniel W, Ross A Thompson, Jacquelyn A Merritt et Joseph J Benz. 1996. « Adult perception of emotion intensity in human infant cries: Effects of infant age and cry acoustics ». *Child development*, vol. 67, n° 6, p. 3238-3249.
- Lester, B. M., et L. L. LaGasse. 2008. « Crying ». In *Encyclopedia of Infant and Early Childhood Development*, sous la dir. de Haith, Marshall M., et Janette B. Benson. p. 332-343. San Diego: Academic Press.
< <http://www.sciencedirect.com/science/article/pii/B9780123708779000451> >.
- Lester, Barry M, et CF Zachariah Boukydis. 1992. « No language but a cry ». *Nonverbal vocal communication: Comparative and developmental approaches*, p. 145.
- Lester, Barry M, CF Zachariah Boukydis, Cynthia T Garcia-Coll, William Hole et Mark Peucker. 1992. « Infantile colic: Acoustic cry characteristics, maternal perception of cry, and temperament ». *Infant Behavior and Development*, vol. 15, n° 1, p. 15-26.
- Lester, Barry M. 1987. « Developmental Outcome Prediction From Acoustic Cry Analysis in Term and Preterm Infants ». *Pediatrics*, vol. 80, n° 4, p. 529-534.
- Lim, J. S., A. V. Oppenheim et L. Braid. 1978. « Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition ». *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 26, n° 4, p. 354-358.
- Lind, J., V. Vuorenkoski, G. Rosberg, T. J. Partanen et O. Wasz-Höckert. 1970. « Spectrographic Analysis of Vocal Response to Pain Stimuli in Infants with Down's Syndrome ». *Developmental Medicine & Child Neurology*, vol. 12, n° 4, p. 478-486.

- Mallat, S. G. 1989. « A theory for multiresolution signal decomposition: the wavelet representation ». *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, n° 7, p. 674-693.
- Manfredi, C., L. Bocchi, S. Orlandi, L. Spaccaterra et G. P. Donzelli. 2009. « High-resolution cry analysis in preterm newborn infants ». *Medical Engineering & Physics*, vol. 31, n° 5, p. 528-532.
- Manfredi, C., M. D'Aniello, P. Bruscaiglioni 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, C., D'Aniello et P. M. and Bruscaiglioni. 1998. « Acoustic measure of noise energy in patients having undergone a vocal fold operation ». In *Proceedings of 9th European Signal Processing Conference EUSIPCO 98*. Vol. 2, p. 1141-1144.
- Manfredi C., D'Aniello M. 1998. « Acoustic measure of noise energy in patients having undergone a vocal fold operation ». In *EUSIPCO 98*. Vol. 2, p. 1141-1144.
- Manfredi, Claudia, Werner Mende, Pierro Bruscaiglioni et Kathleen Wermke. 2003. « Resonance development and formant tuning phenomena in infant's crying ». In *Proc. 3rd Int. Workshop MAVEBA*. p. 35-38.
- 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.
- Markel, J. 1972. « The SIFT algorithm for fundamental frequency estimation ». *Audio and Electroacoustics, IEEE Transactions on*, vol. 20, n° 5, p. 367-377.
- Michelson Katarina , Helena Todd de Barra, Olivier Michelson. 2007. « Sound spectrographic cry analysis and mothers perception of their infant's crying ». In *Focus on Nonverbal communication Research*, sous la dir. de Lewis, Finley R., p. 31-64. Nova Science Publisher.
- Michelsson, K, N Tuppurainen et P Aula. 1980. « Cry analysis of infants with karyotype abnormality ». *Neuropediatrics*, vol. 11, n° 04, p. 365-376.
- Michelsson, K., A. L. Järvenpää et A. Rinne. 1983. « Sound spectrographic analysis of pain cry in preterm infants ». *Early Human Development*, vol. 8, n° 2, p. 141-149.

- Michelsson, Katarina. 1971. « CRY ANALYSES OF SYMPTOMLESS LOW BIRTH WEIGHT NEONATES AND OF ASPHYXIATED NEWBORN INFANTS ». *Acta Pædiatrica*, vol. 60, p. 9-45.
- Michelsson, Katarina, Kenneth Eklund, Paavo Leppänen et Heikki Lyytinen. 2002. « Cry characteristics of 172 healthy 1-to 7-day-old infants ». *Folia phoniatica et logopaedica*, vol. 54, n° 4, p. 190-200.
- Michelsson, Katarina, et Oliver Michelsson. 1999. « Phonation in the newborn, infant cry ». *International Journal of Pediatric Otorhinolaryngology*, vol. 49, Supplement 1, n° 0, p. S297-S301.
- Michelsson, Katarina, Pertti Sirviö et O. L. E. Wasz-Höckert. 1977. « PAIN CRY IN FULL-TERM ASPHYXIATED NEWBORN INFANTS CORRELATED WITH LATE FINDINGS ». *Acta Pædiatrica*, vol. 66, n° 5, p. 611-616.
- Michelsson, Katarina, Pertti Sirviö M.A et Ole Wasz-Höckert. 1977. « Sound Spectrographic Cry Analysis of Infants with Bacterial Meningitis ». *Developmental Medicine & Child Neurology*, vol. 19, n° 3, p. 309-315.
- Moreau, F. 1995. *Méthodes de traitement de données géophysiques par transformée en ondelettes*. Université de Rennes I.
- Oppenheim V. Alan , Ronald W. Schaffer. 1975. *Digital Signal Processing* Englewood Cliffs. New Jersey: Prentice-Hall, 585 p.
- Orozco, José, et Carlos A Reyes García. 2003. « Detecting pathologies from infant cry applying scaled conjugate gradient neural networks ». In *Proc. European Symposium on Artificial Neural Networks*. p. 349-354.
- Othman, M. F., et M. A. M. Basri. 2011. « Probabilistic Neural Network for Brain Tumor Classification ». In *Intelligent Systems, Modelling and Simulation (ISMS), 2011 Second International Conference on*. (25-27 Jan. 2011), p. 136-138.
- P. Boersma, D. Weenink. « Praat: doing phonetics by computer ». < <http://www.praat.org/> >.
- Poissant, Sarah F., Kimberly A. Peters et Michael P. Robb. 2006. « Acoustic and perceptual appraisal of speech production in pediatric cochlear implant users ». *International journal of pediatric otorhinolaryngology*, vol. 70, n° 7, p. 1195-1203.
- Rabiner, Lawrence R. 1989. « A tutorial on hidden Markov models and selected applications in speech recognition ». *Proceedings of the IEEE*, vol. 77, n° 2, p. 257-286.

- Robb, Michael P., et Anthony T. Cacace. 1995. « Estimation of formant frequencies in infant cry ». *International Journal of Pediatric Otorhinolaryngology*, vol. 32, n° 1, p. 57-67.
- Seifert, Eberhard, Monika Oswald, Ulrike Bruns, Mattheus Vischer, Martin Kompis et Rudolf Haeusler. 2002. « Changes of voice and articulation in children with cochlear implants ». *International Journal of Pediatric Otorhinolaryngology*, vol. 66, n° 2, p. 115-123.
- Singer, L.T. , et P.S. Zeskind. 2001. « Biobehavioral assessment of the infant ». In *Analysis of Infant Crying.*, sous la dir. de Zeskind, P.S. and Lester, B.M. New York, : The Guilford Press.
- Soltis, J. 2004. « The signal functions of early infant crying ». *Behav Brain Sci*, vol. 27, n° 4, p. 443-58; discussion 459-90.
- Stevens, Stanley S, et JOHN Volkmann. 1940. « The relation of pitch to frequency: A revised scale ». *The American Journal of Psychology*, vol. 53, n° 3, p. 329-353.
- Sturim, D. E., W. M. Campbell et D. A. Reynolds. 2007. « Classification Methods for Speaker Recognition ». In *Speaker Classification I*, sous la dir. de Müller, Christian. Vol. 4343, p. 278-297. Coll. « Lecture Notes in Computer Science »: Springer Berlin Heidelberg. < http://dx.doi.org/10.1007/978-3-540-74200-5_16 >.
- Sundberg, Johan, Filipa M. B. Lã et Brian P. Gill. 2013. « Formant Tuning Strategies in Professional Male Opera Singers ». *Journal of voice : official journal of the Voice Foundation*, vol. 27, n° 3, p. 278-288.
- Sweeney, Walter P., Mohamad T. Musavi et John N. Guidi. 1994. « Classification of chromosomes using a probabilistic neural network ». *Cytometry*, vol. 16, n° 1, p. 17-24.
- Tadj, Chakib, et A Benlahouar. « Speaker characterization using principal component analysis and wavelet transform for speaker verification ». In *Euro speech*. (Geneva, Switzerland, September 1-4, 2003).
- Thodén, Carl-Johan, et Katarina Michelsson. 1979. « Sound Spectrographs Cry Analysis in Krabbe's Disease ». *Developmental Medicine & Child Neurology*, vol. 21, n° 3, p. 400-402.
- Varallyay, G., et Z. Benyo. 2007. « Melody Shape - A Suggested Novel Attribute for the Biomedical Analysis of the Infant Cry ». In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. (22-26 Aug. 2007), p. 4119-4122.

- Várallyay, György. 2006. « Future Prospects of the Application of the Infant Cry in the Medicine ». *Electrical Engineering and Computer Science*, vol. 50, n° 1-2, p. 47-62.
- Várallyay, György, Jr. 2007. « SSM – A Novel Method to Recognize the Fundamental Frequency in Voice Signals ». In *Progress in Pattern Recognition, Image Analysis and Applications*, sous la dir. de Rueda, Luis, Domingo Mery et Josef Kittler. Vol. 4756, p. 88-95. Coll. « Lecture Notes in Computer Science »: Springer Berlin Heidelberg. < http://dx.doi.org/10.1007/978-3-540-76725-1_10 >.
- 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)*.
- Varallyay Jr György, A. Illényi, Z. Benyo. 2008. « The automatic segmentation of the infant cry ». *BUDAMED*. p. 28-32.
- Várallyay Jr, György, András Illényi, Zoltán Benyó, Zsolt Farkas et Gábor Katona. 2005. « An attempt to detect hearing disorders by acoustic features of the infant cry ». In *Proceedings of the Forum Acusticum 2005 Congress, Budapest*. Vol. 29, p. 1-6.
- Verduzco-Mendoza, Antonio, Emilio Arch-Tirado, CarlosA Reyes García, Jaime Leybón Ibarra et Juan Licona Bonilla. 2009. « Qualitative and Quantitative Crying Analysis of New Born Babies Delivered Under High Risk Gestation ». In *Multimodal Signals: Cognitive and Algorithmic Issues*, sous la dir. de Esposito, Anna, Amir Hussain, Maria Marinaro et Raffaele Martone. Vol. 5398, p. 320-327. Coll. « Lecture Notes in Computer Science »: Springer Berlin Heidelberg. < http://dx.doi.org/10.1007/978-3-642-00525-1_32 >.
- Vohr, Betty R., Lester Barry, Gherardo Rapisardi, Catherine O'Dea, Lucy Brown, Mark Peucker, William Cashore et William Oh. 1989. « Abnormal brain-stem function (brain-stem auditory evoked response) correlates with acoustic cry features in term infants with hyperbilirubinemia ». *The Journal of Pediatrics*, vol. 115, n° 2, p. 303-308.
- Wasz-Höckert, Ole, Katarina Michelsson et John Lind. 1985. « Twenty-Five Years of Scandinavian Cry Research ». In *Infant Crying*, sous la dir. de Lester, BarryM, et C. F. Zachariah Boukydis. p. 83-104. Springer US. < http://dx.doi.org/10.1007/978-1-4613-2381-5_4 >.
- Wermke, Kathleen, Werner Mende, Anne Kempf, Claudia Manfredi, Piero Brusciaglioni et Angelika Stellzig-Eisenhauer. 2005. « Interaction patterns between melodies and resonance frequencies in infants' pre-speech utterances ». In *Proceedings of the 4th International Workshop. Models and Analysis of Vocal Emissions for Biomedical Applications*. p. 187-190.

- Wermke, Kathleen, et Michael P. Robb. 2010. « Fundamental Frequency of Neonatal Crying: Does Body Size Matter? ». *Journal of Voice*, vol. 24, n° 4, p. 388-394.
- Yu-Min, Zeng, Wu Zhen-Yang, Liu Hai-Bin et Zhou Lin. 2003. « Modified AMDF pitch detection algorithm ». In *Machine Learning and Cybernetics, 2003 International Conference on*. (2-5 Nov. 2003) Vol. 1, p. 470-473 Vol.1.
- Zeskind, PHILIP SANFORD, et BARRY M Lester. 2001. « Analysis of infant crying ». *Biobehavioral assessment of the infant*, p. 149-166.