Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

BIOINGEGNERIA

Ciclo XXVII

Settore Concorsuale di afferenza: 09/G2 - BIOINGEGNERIA

Settore Scientifico Disciplinare: ING-INF/06 - BIOINGEGNERIA ELETTRONICA E INFORMATICA

Non Invasive Tools for Early Detection of Autism Spectrum Disorders

Presentata da: Ing. SILVIA ORLANDI

Coordinatore Dottorato

Prof.ssa ELISA MAGOSSO

Relatore

Prof. ssa CLAUDIA MANFREDI

Correlatori
Dott.ssa MARIA LUISA SCATTONI
Prof. STEPHEN J SHEINKOPF

Controrelatore
Prof. MAURO URSINO

Esame finale anno 2015

To my Father,
Philosopher and mentor of my life.

To my Mother, my North Star.

I have always loved the desert.
One sits down on a desert sand dune, sees nothing, hears nothing. Yet through the silence something throbs, and gleams…
"What makes the desert beautiful," said the little prince, "is that somewhere it hides a well…"
I was astonished by a sudden understanding of that mysterious radiation of the sands. When I was a little boy I lived in an old house, and legend told us that a treasure was buried there. To be sure, no one had ever known how to find it; perhaps no one had ever even looked for it. But it cast an enchantment over that house. My home was hiding a secret in the depths of its heart…
"Yes," I said to the little prince. "The house, the stars, the desert what gives them their beauty is something that is invisible!"
[The Little Prince – Antoine De Saint-Exupéry]

Ringraziamenti

In ogni pagina di questa tesi sono racchiusi i volti e le parole dei miei meravigliosi Genitori, guide indiscusse del mio percorso, dei Professori, sempre pronti ad un confronto offrendomi ulteriori insegnamenti e consigli di vita, di quei pochi ma grandi Amici e dei Sopravvissuti della mia famiglia, sempre presenti anche solo con un gesto comprendendo i miei stati d'animo e facendomi di nuovo sentire a Casa. Un ringraziamento speciale va a tutti Voi che mi accompagnate anno dopo anno nella lunga Strada della Ricerca, perché la parola Grazie ha senso solo fino a quando le persone sono presenti dopo resta il Ricordo. I miei Ricordi sono certa che sarebbero fieri ed orgogliosi di questo lavoro e delle mie scelte.

Un sentito ringraziamento a tutti i bambini e alle loro famiglie che hanno permesso la realizzazione di questo progetto, augurandoci che sia un ulteriore passo in avanti per l'utilizzo della tecnologia nella pratica clinica e al porre attenzione ad ogni vagito, movimento e forma di comunicazione dei Piccoli, accompagnandoli il più a lungo possibile nella soluzione dell'Enigma più complesso e affascinante del mondo, cioè lo Sviluppo e la Vita.

Silvia

TABLE OF CONTENTS

Keywords

Autism Spectrum Disorders
Neurodevelopmental Disorders
Audio/Video Acquisition System
Contact-less Computer-Aided Diagnosis
Acoustical Analysis
Automatic Infant Cry Analysis
Voiced/Unvoiced Detection
Wavelet Transform
Parametric Spectral Estimation
Fundamental Frequency
Resonance Frequencies
Cry Melody
Newborn Cry Classification
Random Forest
Newborn General Movements Analysis
Background Subtraction
Skin colour model
Optical flow

Thesis Abstract

Background

Autism Spectrum Disorders (ASDs) describe a set of conditions classified as neurodevelopmental disorders in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). ASDs are characterised by social deficits and communication difficulties, stereotyped or repetitive behaviours and interests, sensory issues and in some cases cognitive delays.

The mechanisms behind ASD are not yet fully understood, but are believed to be related to abnormal neurotransmission in specific brain areas, in particular the limbic system, the amygdale region, the cerebellum and the sub-cortical areas that mediate motor control. ASD prevalence rates from 1/150 to 1/68 in the last 10 years suggesting that ASD represents a significant world public health problem.

The diagnostic evaluation of ASD children is based on the complex and difficult assessment of unusual social interaction, imitation, play, insistence on sameness and preference for fixed routines, verbal and non-verbal communication, thus it is not surprising that a reliable diagnosis is not possible before two years of age. Indeed currently, ASDs are not diagnosed before the 2nd year of life but an early identification of ASDs would be crucial in the light of findings indicating that early interventions are much more effective than specific therapies starting in later childhood.

The impairment of mental functions and difficulties in social interaction give reasons for developing non-invasive contact-less techniques that might aid ASDs early diagnosis. Two such approaches concern audio (cry) and video (movements) analysis in the newborn.

For this reason, the focus is on the analysis of age-specific motor and vocal repertoires, which are known to be later impaired in children with ASD and that have been found altered in other infant disorders. However, while the assessment of movement patterns, has already been performed in ASD children through home-videos, no studies exist concerning newborn infant cry and the correlation between cry and movements has never been reported before.

Crying is the infant's earliest form of communication and recent studies relate this primitive form of language to autism spectrum disorders. Cry involves the activation of the central nervous system and requires a coordinated effort of several brain regions, mainly brainstem and limbic system. For this reason, crying has recently aroused interest as an early sign of potential disturbances and pathologies involving the central nervous system. Thus, the acoustical analysis of newborn cry could contribute to the assessment of the newborn health state. Being completely non-invasive, it is an appealing approach for early ASD diagnosis. The main features for the acoustical analysis are the fundamental frequency (F_0), the resonance frequencies of the vocal tract (F_1, F_2 and F_3), the length of each cry unit (CU, defined as high-energy voiced part of the signal longer than 260 ms), the pauses between consecutive CUs and the crying melody (F_0 shape in time).

GMs can be perceptually detected in the first 6 months of life. At the end of pregnancy and during the first two months after childbirth, GMs are known as writhing movements and are characterised by an elliptic

shape of small to moderate amplitude and by slow to moderate speed. At the age of 6 to 8 weeks of life infants' movements undergo a change in form and a new pattern, called fidgety, emerges; these are elegant circular movements of small amplitude moderate speed and variable acceleration of neck, trunk and limbs in all directions. To date GMs are commonly assessed by expert clinicians (paediatrician/neurologist/child psychiatrist), scoring their variety and complexity according to strict protocols but solely through the perceptive examination. The automatic analysis of GMs concerns the following parameters: average speed, average acceleration, number of movement units (defined as an act of acceleration or deceleration), skewness of the velocity of feet and hands (related to the distribution of movement velocity) and kurtosis of acceleration (that quantifies the shape of the probability distribution of acceleration).

Aims

This PhD thesis is linked to the Italian Project "Young Researcher 2008 (GR3): Non-invasive tools for early detection of autism spectrum disorders" supported by the Italian Ministry of Health, that aims to detect early markers of ASDs through the study of infant crying and General Movements (GMs), a particular kind of spontaneous movements of the upper and lower limbs of the newborn. In fact, the GMs analysis in infancy revealed disturbances that could be clearly detected at age of 4-6 months of age whereas abnormal cry features occur in several neurological disorders. For this reason, the focus is on the analysis of age-specific motor and vocal repertoires in the first six months of life, which are known to be later impaired in children with ASD and that have been found altered in other infant disorders. The GR3 project addresses this issue by means of an innovative multidisciplinary approach, involving both early neurobehavioral characterization of murine ASD models and clinical observation of LR and HR infants during the early neonatal period. Animal models of ASD can provide translational tools to identify neurochemical markers and behavioural patterns that cannot be studied in the human infant. This kind of investigation has been carried out by neurobiologists involved in the project but it is not reported in this thesis.

The GR3 project concerns audio and video recordings performed at 5 time points: at 10 days, 6, 12, 18 and 24 weeks of age. The aim of the project is to set normative ranges for acoustical and motion parameters in a population of control infants, both male and female ("low-risk" newborns, LR) to be referred to for the analysis of "high-risk" (HR) infants, i.e. siblings of children already diagnosed with ASD. This comparison will allow finding early indicators of ASDs after a clinical validation carried out by clinical experts at the age of 36 months to diagnose newborns with Typical Development (TD) or those affected by anomalies in the neuro-development. The results of the acoustic analysis of the TD subgroup were then used to establish possible normative range in the 5 time points considered that might allow finding early indicators of ASDs.

This PhD thesis concerns the automatic analysis of infant crying and movements as an aid to the early diagnosis of ASDs. In addition to a thorough literature review that presents the state of the art on these topics, this three-year work entailed the implementation, development and use of tools for collecting and analysing data that will be described in detail in the following sections and that are summarized below .

Material and methods

According to the aims described above, during the three years of doctoral studies the following issues were addressed:

1. A systematic literature review was carried out to select the features for the acoustical analysis such as the frequency of vibration of the vocal folds (fundamental frequency F_0), the resonance frequencies of the glottic and vocal tract cavities (F_1 and F_2), the time duration of each cry episode (cry unit CU) and the pauses between consecutive CUs. These are the acoustic parameters of greatest clinical interest as related to the control of the central nervous system on the muscles of the phonatory apparatus. Moreover a systematic review was carried out to define the state of the art of the research on crying as an early indicator of ASDs. This research focused on papers concerning retrospective and prospective analyses of vocal development in children diagnosed with ASD as compared to their younger siblings. The same research was carried out for video-based techniques for GMs analysis. This will be described in Section I devoted to the state of art about autism, cry and GMs.

2. Definition and set up of the recording protocol that results in collection of data consistent across all the operators in charge of home recording (as required in the GR3 project). A specific database was setup for the management of patient data and clinical notes. Finally, ad hoc tables were developed for entering the personal data of infants, measurements of head circumference, length and weight, as well as a specific encoding of open fields (notes of the operators during recordings). The details are included in Section II devoted to Materials and Methods and to Experimental Results in Section III.

3. Analysis, design and implementation of a system for audio-video recording of newborn cry and spontaneous movements. The system, called Audio-Video Infant Recorder (AVIR), was developed starting from a previous one (Infant Recorder IR) implemented for the master thesis in Bioengineering. It is based on the recording criteria defined in the protocol and has been provided, along with the appropriate documentation, to the personnel engaged in the GR3 project that used it for audio-video recording, storage and management of data for both LR and HR newborns. Details are reported in Section II devoted to Material and Methods.

4. Analysis of the audio signals both with pre-existing software (BioVoice), whose automatic procedure for the extraction of acoustical features from infant cry recordings was revised and optimized, and with the development of innovative methods of analysis based on the wavelet transform. BioVoice was used for the detection of the vocalic parts of the signal (the so-called cry units, CU) on which the acoustical analysis was performed by estimating the parameters of clinical relevance listed above. The software BioVoice, has excellent performance in terms of frequency resolution thanks to the variable frame length that, for the neonatal cry, may be of very limited duration (typically 4-8ms). However, it suffers from the complexity of the algorithms implemented as regards the computation time required. In order to improve this aspect and taking into account the time-frequency high resolution properties of the wavelets, a new software named NeCA was developed. For the first time the wavelet approach is developed and applied to the estimation of acoustical parameters of newborn cry. Another innovative aspect, as never dealt with in the literature

concerning the analysis of the newborn cry, was the comparison of the two methods through the use synthetic signals.

In addition to the parameters estimated with BioVoice, with NeCA new F_0 shape descriptors are estimated (skewness, kurtosis, and percentiles) that might be useful for a statistical characterization of the melody of crying (F_0 shape in time). The comparison of the results obtained with the two methods showed the reliability and reduced computing time of the new method, in particular for the estimation of F_0 and F_1. Both methods were then applied to audio signals coming from TD newborns providing comparable results. All the results are collected in summary tables organized according to the needs of the GR3 project to allow statistical analysis. Methods and results are described in detail in Sect. II and Sect. III respectively.

5. Development of a classification method applied to the estimated acoustical parameters for the definition of normative ranges: this is in fact one of the main goals of the GR3 project. To this aim the following algorithms were applied under the open source software WEKA: logistic curve, multilayer perceptron, Support Vector Machines and Random Forest with k-fold cross-validation. The best results were obtained with the Random Forest method. To overcome the problem of possible overfitting, best attributes selection and WEKA split option were used. Two classifications were performed, into two classes and five classes respectively. The first one has the purpose to distinguish between TD and HR newborns. The second one concerns the classification of TD newborns according to classes corresponding to the 5 time points: 10 days 6, 12, 18 and 24 weeks of age. The results show that a properly chosen subset of these parameters allows a good differentiation of the characteristics of crying during the first six months of the infant's life, in particular between 10 days and 18 weeks of age (approximately 75% of right classification). Methods and results are reported in Sect. II and Sect. III respectively.

6. Concerning GMs analysis, the work has dealt with the study and development of methods for the automatic analysis of video clips capable to provide kinematic parameters of clinical interest. Two approaches were developed. The first one is based on background subtraction, skin colour model and optical flow for velocity estimation. The second one based on background subtraction, region of interest (ROI) and centroid identification. A binary mask was used to identify the infant’s silhouette using morphological operators in 5 ROIs: head, right and left arm, left and right leg. For each ROI the centroid was selected and motor parameters were extracted. After locating the baby's head in the upper half-part of the image, the image was divided into four sections corresponding to the four limbs. Parameter’s tracking is made searching the centroids in these quadrants. speed and acceleration of each centroid were computed through the first and second derivative of its coordinates and used to identify the indicators of motion (skewness, kurtosis, correlation …). A new method for representation (motiongram) allows to identify the presence or absence of motion and would allow a faster selection of the motion sequences of actual interest. Movement units (MU) are defined, corresponding to an act of acceleration or deceleration. A movement unit is recorded when increasing velocity exceeds the threshold value of 0.5 pixel/frame and when decreasing velocity falls below the same threshold value. This parameter tests whether the frequency of movements of individual limbs varies in infants. The skewness (Sk) of the velocity of the feet and hands was computed, that reflects the

distribution of movement velocity. In case of a normal distribution the skewness is zero. To evaluate the distribution of movement data, the kurtosis of acceleration (K) of each limb was estimated. The kurtosis quantifies the shape of the probability distribution of acceleration. A distribution with a large K, showing a more sharply peaked and heavy-tailed form, indicates intermittent occurrences of limb movements and deviates from a Gaussian distribution (K=3.0). Finally, based on the cramped synchronised movements observed by Prechtl et al. the cross-correlation (CC) of acceleration between hands and feet was computed. The aim is to determine whether the movement of selected points proceeds in the same direction at the same time. Thus the cross-correlation of acceleration between the left and right foot, related to the similarity of the spontaneous movements of both the feet, was computed, as well as that between the left and right hand, the left hand and right foot, the right hand and left foot, the right hand and right foot and the left hand and left foot. It was also necessary to build an active database (that is, a pull-down menu for each item) for the results of the clinical tests for GMs, to allow statistical analysis. Details are provided in Sect. II concerning materials and methods. Preliminary results are presented in Sect. III.

Results

During the three years of doctoral study the analysis of newborn cry signals was performed on a total of 72 LR and 20 HR infants. However only a subset of 39 LR and 10 HR children got the clinical evaluation before the end of the PhD studies. Specifically, infants were validated by clinicians at 18 and 24 months of life with the M-CHAT test and at 24 and 36 months with the ADOS test. All the LR cases were all classified as typically developed (TD) infants while in the HR group one child was diagnosed as affected by ASD, 2 children by language delay (LD) and one with Fragile X syndrome (genetic syndrome). Thus, the assessment was based on 39 TD and 10 HR subjects only, both with BioVoice and the wavelet approach. For each subject and for each time step the acoustical analysis was carried out on a specific sequence of crying selected within the whole recording according to the protocol.

Concerning TD subjects, an increasing trend of F_0 and a slightly decreasing trend for F_1-F_3 was found over the 5 time points. Statistical analysis (t-test) shows that, when the temporal distance between the recording sessions is greater than 6 weeks, there is a statistically significant difference between the average values of F_0. These differences tend to decrease and vanish when the age of the infant is greater than 18 weeks. This could indicate a faster development of the phonatory apparatus in the first weeks of life than in later weeks. The results obtained with the two implemented methods are comparable and in line with those reported in the literature that however are referred to the first 3 months and the 5/6 months of life only. On the basis of tests carried out on simulated data BioVoice was found more reliable with respect to the wavelet approach for F_2 and F_3 estimation. The results also show that the number of CU (calculated only with BioVoice) is not a significant parameter of differentiation among the five time points.

As regards the values of skewness and kurtosis in the 5 time points, about 60% of positive and 40% of negative skewness values were found, that correspond to falling and rising melody of F_0, respectively, with a strong prevalence (94%) of negative kurtosis values that indicate the peakedness of energy distribution. We

observe, however, that the detection of symmetrical curves (skewness = 0) was not taken into account for the lack of references in the literature for the definition of suitable thresholds around 0. An interpretation of this result is thus somewhat complex and will be the subject of future studies. The classification of TD and HR subjects was carried out with WEKA on the following 33 parameters estimated with BioVoice: mean length, standard deviation, minimum and maximum of each CU, mean, median, standard deviation, minimum and maximum of F_0, F_1, F_2 and F_3, signal duration, number, percentage and length of each CU, mean length, standard deviation, minimum and maximum of pauses. The results, ranging between 84% and 86% with split option, show that the acoustic parameters considered enable an excellent separation of the two classes of subjects in all the steps of recording. Significant differences between TD and HR newborns were found for F_0, F_1 and F_2 mean values (t –test, pvalue < 0.05). For HR subjects cry analysis has shown a different trend for F_0 as compared to the TD group especially in the last two step points (18^{th} and 24^{th} week). HR child diagnosed with ASD, two with developmental disorders (language disorders, LD) and one with Fragile X syndrome were compared separately with the LR group subsequently validated as TD. There is a clear difference of the values of F_0 at 18 weeks of life. It may also be observed a slight difference at 10 days, 6 and 24 weeks. Also in case of the two LD there is a greater difference at 18 weeks and a lower one at 12 weeks. The cases, however, are too few to draw further conclusions.

WEKA has also been applied to the classification of the 39 TD subjects in the 5 time step using the Random Forest algorithm and 10-fold cross-validation or split option. With this method the following 22 parameters were selected: length of the CU, mean, median, standard deviation, maximum and minimum of F_0, F_1, F_2 and F_3, number of points of F_0. The method has provided satisfactory results only for the steps: 10 days -18 weeks (75%) and 10 days- 24 weeks (76%). Once again we point out the fact that the case studies should be increased. The analysis of video clips was carried out on 5 TD children whose movements were clinically judged as smooth and fluent. The following parameters are computed with the automatic procedure described above: number/start/end of each sequence of motion (SM), average speed and acceleration of the centroid of each limb, number of MUs. Overall, the number of sequences of motion is variable, with an increasing trend in the last 2 time points. Instead, the number of movements decreases almost always from the first step onwards. These results could possibly support the analysis of the newborn's body symmetry in the first six month of life in an automatic way. Skewness of velocity and kurtosis of acceleration carry useful information to monitor how the newborn trains his/her cognitive processes. In particular, skewness of velocity of the feet reflects possible unequal distribution of movement velocity. Skewness values relative to feet were always negative, with an increasing trend observed from the 6th week on. Similar results were found in literature to quantify limb movements in epileptic seizures through colour-based video analysis. The same trend was found for the hands too. The number of MU/min from the 6th week on results comparable to that found in a study for neonatal clonic seizures from the 12th week of age on, while it reaches much higher values until the 6th week of age. Finally, kurtosis shows a slight increasing trend, with values comparable with the literature. In the future, this method should be compared and tested with perceptual methods used by clinicians. Preliminary results obtained with the proposed approach are encouraging, however, given the

complexity of the analysis and the limited number of case studies many problems are still to be solved. This research could thus be the object of future studies. Results of this analysis are reported in Sect. III devoted to the Experimental results.

Discussion

In literature, studies on the normative data of TD crying concern the trajectories of development up to 3 months of life and are based only on the fundamental frequency F_0. This PhD thesis presents for the first time a study on data collected in five time steps along the first 6 months of life where a large set of parameters is estimated.

From the methodological point of view, this PhD thesis presents new algorithms for the identification of voiced/unvoiced parts of the cry signal, acoustic parameters extraction with wavelet transform and a new approach for the analysis of the infant's movements.

A systematic literature review and a careful revision of existing software and the development of a new one was performed with features of high time-frequency resolution and robustness that make them more reliable than the software tools commonly used for this purpose. The estimated parameters are the fundamental frequency, the first three vocal tract resonance frequencies, shape descriptors of the melody and time domain parameters. Testing was performed on synthetic signals. To our knowledge the test of software tools for newborn cry analysis with synthetic data has never been applied before. A classification with soft computing techniques for the identification of crying patterns in the first 6 months was never carried out. The classification of 39 TD subjects (i.e. those clinically validated as TD within the end of the project) in the 5 time step, based on 22 optimal parameters, provided statistically significant differences for the steps: 10 days -18 weeks and 10 days- 24 weeks. This result points out that the acoustical characteristics of crying of TD newborns differ more with increasing temporal distance between the recording steps, i.e. with age. This could possibly be related to the development of the phonatory apparatus and the increase of the newborn's capabilities of the neurological control over it. However it is not possible to draw further conclusions for the comparison between ASD and TD as we have only one subject diagnosed as affected by ASD. Nevertheless the results for F_0 obtained for the ASD child and LR children are in agreement with those found in the literature at 6 months of age where a higher F_0 value is detected in ASD children with respect to TD newborns.

The overall results of cry analysis suggest that the selected acoustic parameters allow good differentiation of the characteristics of crying during the first six months of life, in particular between 10 days and 18 weeks of age. This result, although preliminary as obtained on a group of only 39 subjects, is nevertheless encouraging in view of a possible definition of normative ranges of the acoustic parameters of the newborn crying that to date does not exist, the main reason being the poor reliability of the estimation methods applied.

The automatic analysis of GMs is another innovative feature of this PhD work as at present a newborn body model for segmentation and 2D reconstruction does not exist. Moreover, there are no studies concerning the analysis of GMs in the first six months of a child's life to extract motor parameters in an automatic way.

Parameters used in marker-based studies for the analysis of general movements in infants with cerebral palsy were extracted. If a larger dataset would be available these first results could provide a first step towards the development of a new protocol after a comparison with the perceptive analysis of GMs performed by clinicians.

Finally, we notice that the data presented in this thesis are collected through an innovative experimental protocol based on a completely contact-less procedure and performed at the patient's home. A new analysis protocol of newborn crying was developed that would also allow setting up a standardized method for the selection of CUs and spontaneous crying (i.e. feeding cry) analysis. In fact, crying analysis commonly makes use of the first CU of a sequence of crying or of random sequences whose selection is often not clearly specified. In this study, the sequences of crying are selected with a robust automatic procedure applied to spontaneous hunger cries selected in a systematic way: after the pre-emphasis and moans only the acute phase of crying of varying duration was selected. In the final section devoted to Discussions the innovative aspects of this PhD work will be highlighted, as well as the difficulties, the limitations and the contribution of this research compared to the current state of the art.

Conclusions

The PhD project has provided first results concerning the automatic analysis of newborn crying through the development of new algorithms and the application of existing ones that will be presented in this thesis. A first approach for the automatic analysis of GMs was also developed as a support to the perceptual techniques used by clinicians. Both of these studies were carried out with the aim of showing that, as both GMs and cry features are related to the development and the integrity of the central nervous system, they could be successfully exploited through the analysis of audio and video recordings. Such contactless techniques are appealing for clinical diagnosis of several pathologies being easy to perform, cheap and completely non-invasive. This PhD thesis was relevant for several reasons. First, at present the analysis of newborn cry is performed through commercial or open source software tools that are not suitable or require the manual setting of some parameters and thus some technical skill to avoid wrong estimates of main acoustical features. One of innovative aspects of GMs analysis concerns the study of these kind of movements in the first 3 months of life with marker-less techniques.

Finally, this thesis proposes for the first time a procedure technological and methodological completely contact-less to aid clinical practice as support for the diagnosis and treatment of neurodevelopmental disorders. The whole system is undergoing testing in the clinical centers and will undergo a specific assessment according to functional ISO standards for medical devices and tools for human-centered design.

Section I - Background

1. Preliminary Concepts

Currently it is estimated that one child out of 68 to be identified as suffering from Autism Spectrum Disorders (ASD) and related Neurodevelopmental Disorders (NDD) and these figures are rising at an alarming rate [1]. The current scientific literature reports many studies and practices showing the relevance of technology as a support to ASD in the fields of diagnosis, therapy and education. The technology is playing an increasingly important role in supporting the behavioural observation making available objective data and measurement about behavioural atypias distinctive of such disorders. Current solutions are anyway inadequate in terms of flexibility, adaptability and integration with ICT (Information and Communication Technology) subsystems having a limited ability to correlate information, furthermore they are often invasive and with reduced usability.

In this section the most important concepts of ASD will be described, along with the clinical and technological approach for the study of infant's cry and movements related to NDD to introduce the new method presented in this thesis.

Autism Spectrum Disorders (ASDs) are complex neurodevelopmental disorders that cause problems with social interaction and communication. ASDs are associated with motor development problems, more or less blurred, and with perceptual and sensory brain areas. These symptoms are usually evident and diagnosed in early childhood or adolescence [2].

These neurodevelopmental disorders result in the alteration of communication skills, behavioural and social interaction. Autistic children are different from those considered with a "typical development" for the activities carried out, fixed and repetitive interests, usually stereotyped.

Currently, ASDs are not diagnosed before the second year of life but an early identification of ASDs would be crucial in light of findings indicating that early interventions are much more effective than interventions starting in later childhood.

Yet 50% of parents of children with ASDs report that they suspected a problem before their child was 1 year of age [3] thus a more precocious diagnosis seems possible. So far, the study of the early symptoms of ASDs is mainly based on retrospective analysis of home videotapes, usually recorded during children's first birthday party. Recently, prospective studies of infant siblings of children with ASDs have been performed, using the Autism Observation Scale for Infants [3-8], which measures visual attention, response to name, social babbling, eye contact and sensory behaviours. Neither home video analysis nor high-risk infants studies allows the detection of consistent abnormalities earlier than 1 year of age. It is worth to notice that in other populations of infant at risk for neurodevelopmental disorders, the combination of neurophysiological and behavioural measures has allowed to evidence precociously atypical maturation of some neural circuits

associated with auditory processing in infants at risk of later developmental problems [9, 10]. Since the diagnostic evaluation of ASD children is based on the assessment of unusual social interaction, imitation, play, insistence on sameness and preference for fixed routines, verbal and nonverbal communication [11], it is not surprising that a reliable diagnosis is not possible before two years of age. For this reason, the focus of this thesis is on the automatic analysis on age-specific motor and vocal repertoires, which are known to be later impaired in children with ASD and that have been found altered in other different infant disorders. In fact, the general movements analysis in infancy revealed disturbances that could be detected clearly at age of 4-6 months of age [12] whereas abnormal cry features occur in several neural disorders [13]. However, while the assessment of movement patterns have already been performed in ASD children by home-videos, very few studies have investigated the specificity of cry in infants with ASDs [14, 15] and the correlation between these two responses have never been reported before. This is very peculiar considering the connection between crying and the functioning of the brainstem and limbic system, both areas compromised in children with ASD [16, 17]. In conclusion, the association of general movements and cry analysis is desirable and of great relevance, since they both reflect the development and the integrity of the central nervous system and they can be exploited for early clinical diagnosis of several pathologies since they are easy to perform, cheap and completely non-invasive.

1.1 Autism Spectrum Disorders

ASDs are a group of complex disorders of brain development. They are characterised at various levels of complexity and severity by difficulties in social interaction, verbal and nonverbal communication and repetitive behaviours. Until 2013, for the Diagnostic and Statistical Manual-IV (DSM-IV) published by the American Psychiatric Association, ASDs include autistic disorder, childhood disintegrative disorder, pervasive developmental disorder not otherwise specified and Asperger syndrome [11]. In most cases, ASDs symptoms appear in the first year of life, when it is not yet possible to carry out a reliable diagnosis. This is because the diagnostic tests currently used and considered as gold-standard for ASD are based on the assessment of unusual social interaction, imitation, play, insistence on sameness and preference for fixed routines, verbal and nonverbal communication that require a conscious and active interaction with the patient hardly recognizable in the newborn [11].

Autism was initially described by Kanner in 1943 [18] as a triad of deficiencies and was defined as specific behavioural signs and impairments in socialising and communicating. At present, autism comprises a spectrum of disorders and autistic subjects are classified based on different kind of capabilities that depend on their individual intellectual and social development. A lack of social and communication skills are amongst the earliest and most common signs of autism and ASD children generally do not share attention with other people and have difficulties to understand the mental state of others. There are some different theories about autism difficulties, one of these called “theory of mind” explains some deficits in autism and

considers that autistic sufferers have difficulties to interpret and interact with their environment and the people in it [19].

In comparison to a "typical" children these difficulties cause a lack of understanding of their own and others motor intentions. The true mechanisms involved in autism are still not completely understood but it is believed to be related to abnormal neurotransmission and specific brain regions, in particular limbic system, amygdala region, cerebellum and sub cortical areas that mediate motor control [20]. The impairment of mental functions and difficulties in social interaction give reasons for finding non-invasive techniques that might aid in the early diagnosis of ASD. Kanner's triad and other studies do not focus on the early development of children, i.e. before their first year of life. Instead, some of the most common symptoms are evident in the first years of life and include sensory processing difficulties [21; 22], limited social interaction [23], deficits or delays in language development [24], and unusual or problematic behaviours [25]. Parents of infants with autism often report sensory peculiarities early in development. Because of the heterogeneity of symptoms, the wide range in their severity and the spectrum of functional deficits, the term Autism Spectrum Disorders (ASDs) is now being used to account for the differences seen in these children [26].

The new DSM-V [2] introduces a different definition of Autism Spectrum Disorders. One of the major changes in DSM-V is the classification of a group of disorders which usually occur in childhood and adolescence. Autism Spectrum Disorder now encompasses the distinct categories of "classic" Autism, Asperger's Disorder, a rather loosely formulated disorder referred to as Pervasive Development Disorder (PDD), and Childhood Disintegrative Disorder defined in DSM-IV. Some major changes in exclusion criteria involve removal of much of the references to language delays or disorders. More specific criteria most likely will exclude some children previously diagnosed ASD defined as Pervasive Developmental Disorder. It will thus be possible to diagnose more clearly an Attention-deficit Hyperactivity Disorder, who typically has milder social skill deficits avoiding being misdiagnosed or "over-diagnosed" as children suffering from an Autism Spectrum Disorder [27, 28].

The new horizon for the diagnosis of ASDs should be the possibility to recognize emotion in facial expression, tone of voice and level of interest in reciprocal social interaction as suggested by a more accurate paediatric neuropsychological assessment or evaluation.

Assessment of sensory-perceptual integration skills, fine and gross motor skills, attention, memory and executive functions along with differential diagnosis will help to determine the level of disability and guide treatment or Individualized Educational Plans (IEP).

1.1.1 Prevalence of ASD

One of the main difficulties to give estimations about prevalence of ASD in a historical perspective is the fact that our understanding of autism has changed over the past decade. One of the changes has been the appreciation that several closely-related disorders exist; they share the same essential features but differ on specific symptoms, age of onset, or family history. After the introduction of the DSM-IV criteria, a dramatic increase in the prevalence of autism has been observed. In the document of "European commission health &

consumer protection directorate-general - 2005" [29], is reported that: "according to the existing information, the age specific prevalence rates for 'classical autism' in the EU could be estimated as varying from 3.3 to 16.0 per 10 000. But these rates could increase to a range estimated between 30 and 63 per 10 000 when all forms of autism spectrum disorders are included. Debate remains about the validity and usefulness of a broad definition of autism."

The document of World Health Organization of 2013 [30] reported for Europe prevalence, the median rate is 61.9/10 000 (range 30.0–116.1/10 000) and for America, the median rate is 65.5/10 000 (range 34–90/10 000).

The prevalence rates are influenced by changing diagnostic criteria and broadening of the concept of ASD, therefore the introduction of the DSM-5 will lead to new statistics.

There is a well-defined increased prevalence in males, with an affected male–to–affected female ratio of approximately 4:1 [31, 32].

There is a 19 % of probability that younger siblings of children with ASD develop autism, a rate significantly higher than the general population [33]. This result prove that if there are two children with ASD in a family, the risk of the third sibling developing ASD increases to more than 32 %. The risk of an ASD diagnosis for male infants who had an older sibling with ASD is almost three times greater than the risk for female infants (26 % compared to 9 %).

1.1.2 Causes of Autism

ASDs are associated with numerous comorbidities and disabling symptoms, such as aggressiveness and self-injurious behaviours, for which behavioural and psychopharmacologic interventions are the basis of treatment [34].

ASDs are known to be extremely heritable, however their common genetic causes remain largely elusive because of the complex behavioural phenotypes and multigenic aetiology of these disorders [35]. Many researchers support the hypothesis that autism is heterogeneous and that different symptoms, behaviours, and levels of severity displayed by the children may be representative of different types of autism. This is the reason to refer to autism like a spectrum of disorders. No one gene has been linked to autism [36, 37], but it has long been theorized that it might be an x-linked disorder because of a high ratio of males [38]. Most researchers today believe that autism may begin with a combination of genetic vulnerabilities and environmental triggers [39]. Autism however is believed to be prenatal in its origin. Many researchers have suggested that there is a genetic mutation (or mutations) that underlie or cause autism. Statistics suggest that there is a 60-90% chance of having a sibling with an ASD if another child is affected. Geneticists studying autism have discovered specific abnormalities on chromosomes 2, 5, 7, 11, 15, and 17 [40 - 44].

Some studies investigated the relation between ASD and neurotoxins. A study was completed that examined tissue samples obtained from 700 children with autism to test for the presence of heavy metals (ex. mercury) due to the "vaccine" theory of autism. While it has been suspected that an immune system response was activated as a result of the injections since some children show chronic inflammation in their brains [45], this

theory has not been supported as a cause of ASD [46 - 50].In addition, the influence of opioids was also examined due to their association with being a by-product of gluten and casein in the diet [51]. Increased levels of these substances have been found in the urine of some children with autism [52, 53].

As already mentioned, aetiology of ASDs is still unknown although several studies identified abnormal neurotransmission in specific brain areas, such as limbic system, amygdale, cerebellum and sub cortical areas that mediate motor control [19] which up to now have been detected only for older infants.

1.1.3 Early diagnosis

The diagnosis of ASD is not made until the first clear signs of the disease appear. This normally happens at the end of the 2nd year of life, when the lack of communicative skills and social deficiencies become evident [8, 54, 55]. In fact, ASD diagnosis is based on the identification of behavioral symptoms, which are reliably evident only after 24 months [6, 56].

The symptoms of ASDs typically are present before 2 years of age and often are accompanied by abnormalities in cognitive functioning, learning, attention, and sensory processing [57]. This is the reason why sometimes it is possible to make a diagnosis from 18 to 36 months.

The assessment of indexes of infant's health during the first six months of life is strongly suggested as their alteration could be considered as early markers of ASDs.

ASDs are also associated with movement disorders, more or less blurred, and with perceptual and sensory brain areas. We observe that, although very peculiar in the relationship between crying and functioning of the brain stem and limbic system [58], to date the correlation between crying and general movements has not been highlighted.

Several studies focused on the importance of early diagnosis in ASD, aiming to identify the first symptoms of these disorders [6, 8, 56, 59 - 68]. This is of relevance because an early diagnosis could improve intervention effectiveness [8, 68, 69]. However, this is not an easy task, also because the severity of these disorders may vary, presenting at least four degrees of severity [2, 11]: 1) autistic disorder (most severe degree); 2) childhood disintegrative disorder; 3) pervasive developmental disorder-not otherwise specified; 4) Asperger syndrome.

One of the characteristic ASD indicators is the delay in language development. Accordingly, vocal development of child diagnosed with ASD, or at high risk of ADS because younger siblings of a diagnosed ASD subject, was widely investigated [70 - 73].

The study of the early symptoms of ASDs is mainly based on retrospective analysis of home videotapes, usually recorded during children's first birthday party [56], but this method presents some limits. This method however presents some limits. First, existing data refer to the assessment of a restricted number of infants (e.g. 10-12) [3-5]. Second, only presence or absence of autistic symptom behaviours was considered. Overall, the method is not standardized because of methodological differences in quality of recording and observational setting. Clinical research on newborn patient would take considerable advantage with the aid

of a non-invasive and marker-less technique that might aid in ASD early diagnosis, enabling a more effective intervention. A recent study support a strict relationship between cry and motor system development [74].
Autistic symptomatology cannot be completely recovered. However, ASD deficits may be in many cases attenuated by means of rehabilitation programs, which resulted more effective if started in the very first years of life [8, 54].

1.2 Infant Cry

Crying is the infant's earliest form of communication and recent studies connect this primitive form of language to autism disorders. Cry involves activation of the central nervous system and requires a coordinated effort of several brain regions, mainly brainstem and limbic system. For this reason, the crying has recently aroused interest as an early sign of potential disturbances and pathologies involving the central nervous system. Thus, cry analysis could contribute to the assessment of the newborn health state. Being completely non-invasive, it is an appealing approach for early ASD diagnosis [75].
The analysis of this signal carried out perceptually requires special skills of the clinician who must be an expert in this field. It is also clear that listening to the crying may lead to incomplete and sometimes conflicting clinical indications as it might be carried out under different conditions (often of interest is the crying of pain, but the crying of hunger or generic discomfort are also explored) and in a short amount of time. This problem can be overcome by recording the crying, even for quite long temporal durations, which can be listened to and interpreted by the clinician later.
The microphone recording, however, presents some problems. The microphone should be of the best quality, possibly unidirectional and placed at a predetermined distance from the newborn; the recording should take place in an acoustically isolated environment; the acquisition (on the computer) should be made using an appropriate audio board and at a sampling frequency high enough to avoid loss of information.
Besides all this, the problem that arises is that of the ability of the clinician to listen and analyse perceptively recordings lasting up to several minutes. In these recordings there are dozens and dozens of "significant events" (i.e., the so-called crying episodes or cry units), but also many parts of no interest (moans, breath, background noise, etc.), and the pace of work of the doctor is often incompatible with the required work to perform the manual selection of the cry units. Finally, to date there are no standards for the above, making it difficult if not impossible an objective assessment of the characteristics of neonatal crying.
All these considerations explain in part why the analysis of the infant crying, even if recognized of great clinical interest, is not widespread nor it is became a part of the protocols of neonatal screening. Recently, thanks to technological development that allows the analysis of the audio signal with high reliability, research is making progress under this respect.
The study of the newborn infant crying has had an outstanding growth in the last decades. In the 1960s Wasz-Höckert and his group began the first spectrographic studies of crying and categorized them into cry of anger, pain and hunger [76 -79]. Michelsson et al. described the crying of infants with neonatal asphyxia, [80, 81], cleft palate [82] and with bacterial meningitis [83] being the first major descriptions referring to the

crying spectral analysis. Lester in 1976 studied the crying of malnourished infants [84], the effects of maternal use of drugs such as marijuana or cocaine and its effect on the production of crying [85 - 87]. Later, Lester and Corwin described the crying of infants died of sudden death [88]. Some authors tried to detect differences between crying of premature and term infants and with respect to gender but no significant differences in the fundamental frequency (F_0) have been found [89, 90].

From the functional standpoint, current theories of physiology of sound production have been used for explaining the categorization of the cries and their analysis [91]. In the case of newborn infants, due to high perinatal risk, excessive levels in the production of crying have been reported, observing also high irritability, physiological instability and delay in the response [92]. In neonates with hyperbilirubinemia the brainstem nuclei are affected altering the auditory pathway, causing an increase in the production of cries and in the values of F_0 and the first formant [93 – 95]. Studies also include psychological aspects of crying, mother-child interaction, and semantic content according to its characteristics depending on the conditions in which crying occurs, such as pain, stress, colic, cramps or cold. Early automatic methods exist based on Fourier analysis [96] and on parametric techniques [97, 98] for the study of features such as F_0, resonance frequencies, duration, etc. Some studies show that crying reflects the neurophysiological integrity of the infant. Infant cry variation may be related to organic and functional changes as *cri du chat* [99], Down syndrome or hyperbilirubinemia [100]. Similarly, it has been attempted to associate infant cry with future events such as global cognitive development or sudden death [88, 101].

Previous studies have shown that preterm infants and infants with neurological conditions have different cry characteristics such as fundamental frequency, when compared to healthy full-term infants [102 - 106]. Research has been carried on to study possible differences between full-term and preterm infants in their neuro-physiological maturity and the possible risk to the brain for the premature infant caused by prolonged deoxygenation due to crying [107].

The spectral-phonographic analysis of the newborn baby crying has received special attention. Most of the studies are focused on the analysis of full term newborns, premature, with neurological, metabolic, or chromosomal alterations, as well as those with congenital anomalies among others [94, 108]. In recent studies the subjective auditory analysis of voice and speech is supported by automatic acoustical analysis [107 - 111, J1 – J5]. The procedure to automatically find the class to which an infant cry sample belongs to is known as Automatic Infant Cry Recognition (AICR) [110]. The automatic infant cry classification process is, in general, a pattern recognition problem, similar to the Automatic Speech Recognition (ASR). From the infant's cry (input pattern), the goal is to classify the kind of cry or pathology detected on the baby [111].

The most significant acoustical parameters of infant crying are the fundamental frequency (F_0) and the first two resonance frequencies of the vocal tract (F_1 and F_2). F_0 reflects the regularity of the vibration of the vocal folds, F_1 and F_2 are related to the varying shape of the vocal tract during phonation and thus to its control and F_3 concerns the length of the vocal tract [109].

Both are extremely relevant also in the study of language development in the infant [112]. Few papers describe the F_0 developmental pattern in the first months of life. Gilbert et al [113] analysed the variation of F_0 (mean, median and standard deviation) in hunger cries of 4 male infants during the first 12 months of life and found an increase of F_0 median in the first 5 months and a decrease at the 12th month. Lind and Wermke [114] reported no significant difference in F_0 mean in spontaneous cries of male healthy infants in the first three months of life. Rothganger [115] found that the mean fundamental frequency of crying varied considerably from 441.8 to 502.9 Hz in 25 infants recorded at the 3rd –5th day of the 1st, 3rd, 6th, 9th and 12th month. Baeck and De Souza [90] found a 380-435 Hz range of F_0 when analysing 30 male and female healthy infants recorded at birth and every two weeks until the 6th month of life.

This not exhaustive list shows the noticeable scientific interest for the analysis of infant cry. Unfortunately, the variety of cries (feeding, anger, and pain), the lack of dedicated software for infant crying recording and analysis and the difficulties of recruiting a large number of subjects did not allow defining normative values for F_0. One reason is that this study requires standardized and appropriate recording settings and a preliminary perceptual analysis of the waveform to extract relevant sequences, the so-called crying episodes or cry units (CUs). However, the workload of clinicians is often incompatible with this analysis, as it requires the manual selection of a large number of cry units and the removal of moans, breath, background noise, from recordings lasting up to several minutes. Moreover, the acoustical analysis is commonly performed through commercial or open source software tools like MDVP™ [116] or Praat [117, 118] which however were developed for the analysis of adult's voice and singers. Thus, their proper use with infant high-pitched quasi-stationary cry signals requires manual setting of some parameters and thus some technical skill [119].

The most relevant clinical parameter is given by the F_0, which reflects the regularity of the vibration of the vocal folds of the newborn, and in particular, the "melody" of the cry, that is the temporal trend of F_0 within a cry episode.

The huge number of works on the estimation of F_0 shows how this parameter was and is still the subject of increasingly extensive research. More recently, the "shape" of the melody has been classified into several categories and the percentage of "shapes" belonging to one or the other category is considered as an additional relevant parameter. Four main categories have been identified [120 - 125]: 1) *rising*- slow increase to a high F0 (200– 300 Hz and more) followed by fast decrease; 2) *falling* - fast increase to a high F0 (200–300 Hz and more) followed by slow decrease; 3) *plateau* - slight increase and slight decrease, i.e. cries characterized by a limited F0 variation (less than 100 Hz); 4) *symmetric* - high F0 increase and decrease (200–300 Hz and more), both of about the same duration.

As already mentioned, the study of neonatal cry, has its origins several decades ago, when the technology was limited and it was therefore mainly based on the perceptual analysis made by the clinician through listening to the cry and visually analysing the recorded signal and its spectrogram [9, 84, 88, 126, 127]. The Fast Fourier Transform (FFT) is implemented in the MDVP™. While most of the research is devoted to the estimation of F_0 with traditional approaches such as FFT and cepstrum [126 - 130], few papers address the

RFs estimation: in several papers F_1 is estimated with FFT [88, 97, 129, 131, 132] and recently FFT is applied for F_1-F_3 estimation [130]. In [132] FFT, power spectrum and LPC methods are compared with results comparable only for F_1. Later several approaches were tested on synthetic [97, 98, 107, 112] and real signals and a fully automatic adaptive parametric approach was developed, named BioVoice [107, J3-J5, C1-C5], recently successfully applied in [131]. As for F_0, the difficulty in the estimation of the RFs is mainly linked to the quasi-stationarity and the very high range of frequencies of interest in the newborn cry which requires sophisticated adaptive numerical techniques characterized by high time-frequency resolution. The reason of this gap is their variability and to fill this gap is one of the aims of this thesis.

Thus, it is highlighted the need to develop dedicated software tools that could provide automatically the main parameters of the infant cry.

In the last years several authors propose classification methods for a wide range of newborn pathologies based on the analysis of infant cry. Poel et al [133] present results concerning the classification of neonate crying sounds into "normal" and hypoxia-related disorder with Radial Basis Function Neural Networks with an overall classification performance of 85%. Lederman et al. [134] propose the classification of infants with cleft palate based on parallel HMM (PHMM) for coping with age masking, based on a maximum-likelihood decision rule. The proposed algorithm yields an average of 91% correct classification rate in a subject- and age-dependent experiment. Mijovic et al [135] propose Empirical Mode Decomposition (EMD) technique to assess the existence and extent of decoupling in term neonates and its possible relation to clinical pain expression. In [136] Sahak et al. applied Combined Support Vector Machine (SVM) and Principal Component Analysis (PCA) to recognize the infant cries with asphyxia with a classification accuracy of 95.86%. Zabidi et al [137] applied a new algorithm to optimize Mel Frequency Cepstrum Coefficients (MFCC) parameters, in order to extract an optimal feature set for diagnosis of hypothyroidism in infants using Multi-Layer Perceptrons (MLP) neural network. Nonaka et al. [138] used a hidden Markov model architecture, in which state likelihoods are estimated either with Gaussian mixture models or by converting the classification decisions of a support vector machine. The algorithm yields up to 95% classification precision (86% average) to identify expiratory and inspiration phases from the audio recording of human baby cries. In Hariharan et al [139] a General Regression Neural Network (GRNN) is employed as a classifier for discriminating infant cry signals. Two classes of infant cry signals are considered such as normal cry signals and pathological cry signals from deaf infants. In [139] and [140] two types of radial basis neural networks such as Probabilistic Neural Network (PNN) and General Regression Neural Network are employed as classifiers for discriminating infant cry signals. Three classes of infant cry signals are considered: normal cry signals, cry signals from deaf babies and babies with asphyxia. Etz et al [141] propose a decision tree to classify infant cries in order to find differences between infants with normal development, hearing impairment (HI) and unilateral cleft lip and palate, while Farsaie Alaie et al. [142] apply Gaussian mixture models (GMMs) to distinguish between healthy full-term and premature infants, and those with specific medical problems with a true positive rate of 80.77% and a true negative rate of 86.96%.

Finally, in [143] three different types of infant cries are considered: hunger, pain, and wet diaper. Gaussian mixture models (GMMs) are used to classify the above-mentioned cries.
In this thesis classification methods for crying analysis are tested with WEKA [144] application using Neural Network and Random Forest method. Results are reported in Section III.
Studies about crying and autism will be explain better in the Chapter 2 where a systematic literature review is presented.

1.3 General Movements Assessment

Besides cry, great interest is paid to newborn spontaneous (i.e. not induced or stimulated) movements, known as general movements (GMs). Regular GMs are gross movements, involving the whole body with a variable sequence of arm, leg, neck and trunk movements rising and waning in intensity, force and speed. They begin and end gradually and may last from a few seconds to several minutes. The majority of sequences of extension and flexion of arms and legs is complex, with superimposed rotations and often slight changes in direction. These added components make the movements fluent and elegant and give the impression of complexity and variability [145 - 148].
At the end of pregnancy and during the first two months following childbirth, GMs are commonly referred to as *writhing movements*, characterised by small to moderate amplitude and by slow to moderate speed. Fast and large extensor movements may occasionally break through, particularly in the arms. Typically, such movements show an elliptic shape and give the impression of a writhing quality of movement. At the age of 6 to 9 weeks, the form and features of GMs of normal infants change from the writhing type to a fidgety pattern [147]. The two features can coexist for a few weeks. *Fidgety movements* (FMs) are circular movements of small amplitude and moderate speed and variable acceleration of neck, trunk and limbs in all direction. They are continual when the infant is awake, except during focused attention. FMs may be seen as early as 6 weeks of age but usually occur around 9 weeks and are then present until 15 to at most 20 weeks [147, 148]. They may last from a few seconds to several minutes or longer. What is particular about them is the variable sequence of arm, leg, neck and trunk movements. They wax and wane in intensity, force and speed and show a gradual beginning and end. The majority of sequences of extension and flexion movements of arms and legs are complex, with superimposed rotations and often slight changes in direction of the movement. These added components make the movements fluent and elegant and give the impression of complexity and variability [147].
Abnormal Writhing GMs are present during the prenatal period and during preterm, term and first two months after birth. They are classified as: Poor-Repertoire of GMs (PR: a sequence of the successive movement components is monotonous and movements of the different body parts do not occur in the complex way as seen in normal GMs). Cramped-Synchronised GMs (CS: these appear rigid and lack the normal smooth and fluent character, all limbs and trunk muscles contract and relax almost simultaneously). Chaotic GMs (Ch: movements of all limbs are of large amplitude and occur in a chaotic order without any fluency or smoothness and they consistently appear to be abrupt).

Abnormal Fidgety GMs are Absent "F-" (FMs are never observed from ages 6 to 20 weeks, however other movements are commonly observed; CS infants never show FMs) or Abnormal "FA" (FA look like normal FMs but their amplitude, speed and jerkiness are moderate or greatly exaggerated). After 20 weeks infants may have abnormal spontaneous motor repertoire.

In the last decades, Prechtl´s method [147] was proven to be a highly accurate and specific marker-less diagnostic tool for the qualitative assessment of GMs, preventing discomfort for the baby without interfering with other measures. To date GMs are commonly assessed by expert clinicians (paediatrician/neurologist/child psychiatrist), scoring their variety and complexity according to strict protocols but solely through a perceptive examination.

Beside perceptual evaluation, marker-based techniques have been applied to the analysis of GMs [149, 150]. In addition to quite expensive equipment, (building three-dimensional models requires at least two high-speed cameras), marker-based techniques require the application of markers in well-defined anatomical landmarks (often the joints) that, although of reduced invasiveness in adults, can make the newborn movements less spontaneous or even hinder them. Despite these drawbacks, marker-based techniques allow to obtain very reliable models thanks to accurate data processing. Marker-based analysis for the study of the infants' movements was performed by Coluccini and colleagues [149] with the aim of assessing movements at different stages of development (7, 10, 12 weeks). Other applications concerned the evaluation of signs of spasticity in newborns at risk [151] or the analysis of spontaneous movements in control cases [152, 153].On the other hand, marker-less techniques can be applied on movies recorded with any good quality camera under an appropriate and standardized protocol. Adde et al. [150, 154] assessed fidgety movements identifying the centroid of motion of the subject by a video processing software tool for the diagnosis of cerebral palsy. The centroid of motion is plotted on a diagram called motiongram (motion diagram). In a sample of 140 subjects, the method has shown a sensitivity of 81.5% and a specificity of 70.0%. These studies will be explain better in the Chapter 2 where a systematic literature review is presented.

GMs and crying could be considered early indicators of neurodevelopmental disorders, since they are both related to a functional neurotransmission in specific brain areas. These putative early indicators have been analysed in several studies [67, 75]. Preterm infants and infants with neurological conditions have different cry characteristics when compared to healthy infants [154, 155]. Several studies [147] prove that in newborn infants affected by different brain lesions spontaneous motility loses elegance, fluency and complexity.

Though the perception of the clinical specialist is undoubtedly the most accurate method to assess infants' crying and movements, the procedure of finding and analysing crying episodes and video frames of interest out of the whole recording is operator-dependent. Moreover, the huge amount of recorded data makes a detailed analysis often prohibitive in daily clinical practice being highly time consuming even for trained and qualified clinicians. Finally, clinicians make use of several acquisition devices and software tools not specifically designed for clinical use: that is, they have to use different hardware and software tools and even resort to pen and paper to manage patient data, record and process audio and video signals to obtain parameters of interest, collect and save results, fill tables with diagnostic indexes, scores, etc. Thus, the lack

of a unique software tool that aids the clinician in the many different steps from data acquisition to clinical results makes this approach not yet widespread.

Being cheap and completely contact-less, both cry and GMs automatic analysis have recently aroused interest as early signs of potential disturbances involving the central nervous system including ASDs [1]. The purpose of this thesis is to contribute to the development of such methodologies.

2. State of the Art

2.1 Cry and Autism: a systematic review

One of the main symptoms of the ASD or of Neurodevelopmental Disorders (NDD) is a severe impairment in vocal communication skills. Since the first way of communication in infant is the cry, several studied investigated the relationship between cry and ASD. The result of this study is that there is evidence that cry of infant with ADS is significantly different from that of normal ones, although further studies are required to better standardize clinical protocols and methodologies for the analysis of cry.

The first studies analysed the cry of ADS infants by investigating, with different means, the reaction of adults exposed to the cry. Subsequent studies investigated differences in ADS infant cry by applying several voice analysis techniques to the cry, through retrospective and, more recently, prospective studies.

The purpose of this section is to review systematically the state of art of studies from 1993 to date investigating the crying as early indicator of ASDs before the 3rd year of life.

2.1.1 Study design

Articles published on scientific journals from 1993 to December 2014 focusing on the study of cry in infants affected by ASDs, or at high risk of ASDs, were reviewed. Infants were considered at high risk of ASD if they were younger sibling of autistic children. Only studies based on the analysis of cry recordings were included, whether retrospective or prospective. Those studies based on parent's memories or descriptions of ASDs infant cries were excluded.

2.1.2 Search strategy

An electronic search using Ovid Medline, PubMed and Scopus databases was performed. Additionally, a linear investigation among references of retrieved papers was performed too. This search was performed combining with boolean operators the following keywords: ''cry''; ''crying''; ''cries''; ''infant''; "infants"; ''baby''; "babies"; "new-born"; "newborns"; ''newborn''; "autism"; "autistic"; "ASD". Two investigators analysed independently the title, abstract or the full text of each paper to assess the coherence with the objective of the study. In case of disagreements a third researcher reviewed the paper to achieve consensus.

In this study, different acronyms were used: ASD for Autism Diagnosis, TD for children with Typically Developing, HR for children at High-Risk of ASD because siblings of autistic children and LR for Low-Risk infants (children before 36 months that could be not received an autism diagnosis).

2.1.3 Inclusion, exclusion and classification criteria

Relevant studies published only in English were included if they fulfil the following criteria:

- There is a control group: comparing infants with ASDs, or HR, versus normal infants;
- There is HR and LR data validation for the diagnosis of ASDs;

- They focus on cry episodes (CUs) or cry sequences (epochs or excerpts) registered from infants in the first two years of life (before a clinical ASD diagnosis);
- Results of the study were supported by an appropriate statistical analysis.

Finally, the studies included in this review were classified in three categories, according to their design and methodology:

- Retrospective studies, that is those that analysed retrospectively infant cry from video recorded by parents during the first two years of life;
- Prospective studies, i.e. those analysing the cry of high risk infants (younger siblings of a child with a diagnosis of ASD) from video recorded by operators or parents for the study;
- Perceptive studies, that is those investigating the reaction of adults to the cry of infants with or without ASD.

2.1.4 Results

The strategy adopted allowed to find 35 papers [15, 155 – 186, J4, C10]. The number of papers included/excluded and their final classifications are reported in Figure 2.1. Reasons why some papers were excluded in each step of this study are further detailed in Tables 2.1 and 2.2.

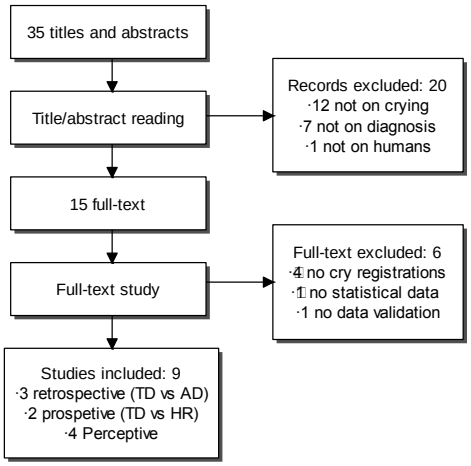


Figure 2.1 Study flow diagrams

The studies included are described in three tables: Table 2.3 retrospective, Table 2.4 prospective and Table 2.5 perceptive studies. For each study, these tables report the goal, the design, the description of control and observed groups (sample size of, IQ mean, gender composition, mean age at the cry registrations), the information regarding the diagnosis of ASD (age at the diagnosis and tests used), the dataset of signals analysed (i.e. number of episodes and durations), the software tools used to feature the signals, the features investigated (in time domain, frequency domain, time-frequency domain), the main results (in bolt those statistically significant), the main limits reported by the authors of the studies.

2.1.4.1 Retrospective studies

Three retrospective studies were found. In these studies, the cry sequences were extracted from home video recordings of "routine" life's situation in the first or second year of child's age. All infants recruited were free from other pathological conditions (e.g., seizures, Fragile X syndrome), and visual or hearing impairments. In the first two studies all the children were firstborn. All participants of retrospective studies were recruited among the population of patients at the Observation and Functional Diagnosis lab (ODFLab) at the University of Trento, since the authors of the three studies were from the same group.

In these studies the autistic children were diagnosed after the 3rd year of life (mean month: 36,77 $\pm$ 1,6). The diagnosis of autistic disorder was performed according to the "Diagnostic and Statistical Manual of Mental Disorders"(*DSM-IV-TR*) [187], and "Autism Diagnostic Observation Schedule – ADOS" [188, 189] criteria. The mean Intelligence Quotient (IQ) of infants was measured with the "Griffiths Mental Development Scales" and it was different in the two groups (TD and ASD) in the 3 studies. The ADOS and the *Griffiths Mental Development Scales* were administrated by three different trained clinical psychologists, to reduce potential bias. These clinical psychologists were involved only in the diagnostic assessment and were unaware of the purpose of the study. The group was etiologically heterogeneous, but children were excluded if facial or body features suggested a genetic syndrome (e.g., children with Down syndrome were excluded).

In all studies a researcher, who was blind to the diagnoses of the participants, edited the videotapes and selected only the excerpts of the videos were children were crying. In Table 2.4 the materials, methods and main results of these three studies are schematically reported.

The first retrospective study was published in 2009 by Esposito et al. [180]. In this study, a comparison between 12 TD, 12 Delay Development (DD) and 12 ASD infants' cry at 12 months was presented. Three types of information were extracted by home videos for the analysis: infant cry sounds, infant movements, and mother behaviour during cry episodes. According to the aim of this review, only infant cry sounds were reported. Because of limited videos' quality, only one useful cry episode was extracted for 12 infants (4 in each class) and 2 cries were extracted for the remaining 18 infants (6 in each class). The mean duration of each epoch of cry was 32s and no significant differences in the length of cries were found.

The authors focused only on time domain features and particularly on the durations of: *screams*, defined as cry episodes with frequencies between 1-2 kHz; *moans*, defined as those cry episodes with no harmonic vibrations; *aspiration/expiration* phases, defined as cry sounds that follow the breathing rhythm; and *pauses* between 1s and 10s among consecutive cry episodes. All the durations were normalized to the length of the cry episode in which they were recorded. The main results of this study was to observe a **s**tatistically significant difference in screams, aspiration/expiration phases and pauses. Particularly, once normalized, the duration of screams of ASD infants were significantly longer than screams of normal infants. Conversely, pauses and expiration/aspiration phases resulted significantly shorter in ASD infants. No significant differences were observed in moans. The authors reported some study limitations that could have affected

the significance of the achieved results as: the limited number of subjects, the heterogeneity of ASD severity and the low quality of audio recording in amateur settings.

In 2010, Esposito et al. [155] investigated the developmental trajectory of the fundamental frequency (F_0) of cry-episodes in TD child in infancy, and whether or not this trajectory is the same in children with ASD. A total of 240 epochs of crying was extracted (80 for each group) from 8 registrations from each participant: 4 registered during the first year of life (about 5 months of life) and 4 during the second year (about 18 months of life). The recordings were extracted by videotapes registered during family play situations or special events such as parties. Digital signal processing and the acoustic measurements were accomplished using the Praat software [118]. The mean duration of each epoch of crying was 24 seconds (SD = 9.3) and no significant difference in the length of the cries was found among the three groups (TD, DD, ASD). The authors focused on the F_0, which is defined as the basic frequency of the cry, perceived as a "pitch". The minimum and maximum F_0 values were set in the Praat software between the 200-700 Hz, according to [115].The sampling rate of the audio signal was set to 44,100 Hz, and the signal was low pass filtered at 10,000 Hz [190]. In this study, a long-term average spectrum (LTAS) was employed to provide spectral information for the crying episode. From all the LTAS of all the cry episodes, the First Spectral Peak (FSP) was computed and used to estimate the average F_0 [191]. Prior to data analysis, univariate and multivariate distributions of the FSP scores were examined for normalcy, homogeneity of variance, outliers and influential cases [192]. Since the FSP scores were normally distributed no transformation was applied to solve problems of non-normalcy. The distance of each case from the centroid was evaluated to screen for multidimensional outliers [118, 192]. The authors concluded that children with ASD had significantly higher fundamental frequencies than normal ones at 18 months. Moreover, in normal children a change of the F_0 trajectory was observed, while no changes were observed in children affected by ASD.

In 2010, Esposito et al. [181] repeated the study described in [155] enrolling more subjects. A total of 210 sequences of crying (epochs) were extracted (70 for each of the three groups) from 5 registrations from each participant at about the 18th month of life. Also in this study, the recordings were extracted by videotapes. The mean duration of each epoch was 22 seconds (SD = 7.3) and no significant difference in the length of the cries was found among the three groups (TD, DD, ASD). Selection, classification and featuring of signal were already described in [155]. FSP scores were normally distributed. The distance of each case from the centroid was evaluated to screen for multidimensional outliers as described in [192]. The authors concluded that children with ASD had significantly higher fundamental frequencies than normal ones at 18 months. Moreover, in normal children a significant change of the F_0 trajectory was observed, while no changes were observed in children affected by ASD. In addition to the 2009 study [180], the cry were also organized in three main categories, according to the context: feeding, changing diaper, pain. However, no significant differences emerged among the different contexts (feeding, changing diaper, pain) and the three categories were finally analysed as one.

No additional limits were reported compared to previous studies [155, 180].

2.1.4.2 Prospective studies

Two prospective studies were identified. The main difference with respect to the retrospective studies was that these papers examined the acoustic characteristics of cries in younger siblings of a child with ASD (HR) and a low risk group (LR).

The validation of these studies was performed at around 36 months of life, according to the DSM-IV-TR [187] and ADOS [188, 189] criteria. More analysis criteria were used with respect to those used in the retrospective studies. These criteria are reported in Table 2.4. In the same table materials, methods and results of both studies are also reported.

The first prospective study was published in 2012 [156] and its aim was to characterize cries of infants considered at risk for autism compared with the cry of low-risk infants at 6 months. Older siblings were evaluated prior to enrolment with the Autism Diagnostic Observation Schedule [188, 189] by a trained clinician. High risk infants were enrolled in the study if the older sibling scored above the threshold for autism on the ADOS. Following the study, HR infants were finally evaluated at 36 months of age to document developmental functioning and to validate diagnosis. Developmental functioning for the HR group was assessed using the Mullen scales of early learning [193].Three infants of the HR group were classified as ASD at 36 months of age by a trained clinician blind to previous study data. The prospective analysis was performed on cry registration from audio-video tapes made by one of the researchers at the infants' home at 6 months of age (+/- 3 days), while the infants were involved in routinely activities. Recordings were made using a digital video camcorder and a wireless microphone and transmitter that was embedded in a vest the infants wore over their clothing. Placement of the microphone on the infants' clothing allowed for some degree of standardization of the distance to the infant's mouth. Potential causes of the cry were also estimated (frustration, seeking attention, hunger, startles, pain-related vs. non-pain-related). Episodes of cry and fuss vocalizations were identified from video recordings by a blind research staff. Only those cry episodes without background noises that would interfere with the analysis (e.g. adult talk, sounds from toys, or other environmental noises) were considered suitable for the analysis and both pain-related and non-pain-related cries were included. Cry excerpts of 60 seconds were manually selected, including 30 sec before and 30 seconds after the start of the cry. The first utterance (expressed vocalization of distress between two pauses) of each cry was analysed. The method used for acoustic analysis was validated in [86 - 88]. Sheinkopf and colleagues measured F_0 and phonation. F_0 is defined as the basic frequency of the cry and phonation was referred to cry segments (% of measurement blocks) resulting from harmonic vibration of the vocal folds. Other variables such as cry amplitude and formants (F_1 and F_2) were not analysed because outside the focus of the study. Anyway, the authors estimated these parameters: F_0, Variability of F_0, phonation, hyperphonation, utterance duration, average energy/amplitude, variability of energy/amplitude, F_1 and F_2.

Independent *t*-tests were performed on the mean values to investigate group differences in acoustic cry features, with separate analysis being run for non-pain- and pain-related cries. Results and limits of this study are reported in Table 2.4. Two infants in the HR group were diagnosed with ASD once 3 years old. The

analysis of their cries revealed that they produced higher pitches in pain-related cries and the lowest average phonation in both pain-related and non-pain related cries. However, two subjects were not sufficient to provide statistical evidence.

In 2013, Esposito et al. [182] examined the cries extracted from the separation phase of the Strange Situation Procedure in a sample of both HR and LR toddlers.

The aim of this research was to investigate the acoustic characteristics of cries elicited in a standardized social interaction context with the Strange Situation Procedure (SSP) [194]. The SSP is a gold-standard measure of infant attachment security which consists of a series of separations and reunions with the caregiver designed to activate the infant's attachment behavioural system. 27 participants were enrolled in a larger longitudinal study investigating the early social and emotional development conducted in the Miami metropolitan area. LR and HR groups did not differ significantly on either the age of the older siblings (older brother age was around 4 years for HR and LR groups). Children were excluded from the study if they had a gestational age below 37 weeks, or major birth complications. Additional inclusion criteria were: (1) administration of the SSP, and (2) audio recordings of cries for acoustic measurements that did not contain background noises that would interfere with the acoustic analysis (e.g. adult talk, sounds from toys, or other environmental noises). The mean age of SSP administration was 15 months. HR infants had one (and two of them had two) older siblings with a diagnosis of ASD that was confirmed via a DSM-IV-TR and ADOS [188, 189]. LR siblings had no reported history of ASD in their first degree relatives, and all of their older siblings received a cut-off score lower than 9 (indicating no evidence of ASD) on the Social Communication Questionnaire [195]. Three infants from the HR group met criteria for an ASD when 3 years old. Infant-mother attachment security was assessed at 15-18 months. The SSP was a 25-minute procedure. Cries that expressed vocalization of distress were extracted from the second separation episode of the SSP, during which the infant was left alone. All infant cries started within the first 30 s of the separation. All the recordings were made in the same assessment room with an overhead omnidirectional microphone. A researcher who was blind to the status of the children extracted all epochs of crying without noise. An expert audio analyst judged appropriate all selections. Cries were broken into epochs if there was a pause between cry utterances of greater than 5 s. An average of 4.5 epochs of cry from each participant were analysed (no differences emerged for the number of epochs for each group at each time point). The authors analysed the following features: number of epochs, F_0, F_0max, F_0 range, duration of the epochs and of the first utterance. For acoustic analysis, they used Praat [118]. Authors made a sound quality control and digital signal processing, which was not detailed in the paper. The analysis was performed as in previous study of Esposito et al [155, 181]. The authors estimated the average fundamental frequency (F_0) of the epochs of crying, the Maximum Pitch (F_0 Max, highest level of the FSP), variability of Pitch (F_0 Range, range of F_0 across the cry epoch), and the duration (average time in seconds). All these measures were extracted for the entire sequence of each child's crying (excluding the inter-utterances interval), and for the child's first cry utterance. Inferential statistics for number of epochs, F_0, F_0 max, F_0 range, and duration from ANCOVAs at mean age of 15 months were calculated. The results and the limits of the study are reported in Table 2.4.

2.1.4.3 Perceptive studies

Four studies investigating adult reaction to infant cries with ASD were identified in literature and grouped as perceptive studies. These studies were considered relevant for this review as they aim to investigate if ASD cries are different from those of normal infants thought the reaction of adults. In all these studies the cry sequences were extracted from home video recordings about "routine" life's situation in the first or second year of child's age. In Table 2.5 the materials, methods and main results of these four studies are reported. Also in this study an acoustical characterization of the cries was performed, using Praat [118] as in the retrospective studies [155, 181]. F_0 was estimated by averaging the first spectral peak of the spectrum of each cry episode for each subject. To compare ASD and TD cries, a general linear model (GLM) and Tukey HSD Post hoc tests [155, 181] were used.

In 2011, Esposito et al. [185] published a study focusing on how adults perceived distress in children with different developmental conditions. The authors focused on different clues: cry, motion and facial expression. The children were diagnosed as in [155, 180, 181]. Differently from previous studies, infants with a diagnosis of Pervasive Developmental Disorder not otherwise specified or Asperger Syndrome were not included in the study. A different threshold for the Griffiths Mental Development Scales was adopted (62 instead of 57 as in [181], 64 as in [180] or 59 as in [155]). One researcher, who was blind to the purpose of the study and to children diagnosis, gleaned the registrations. Three sequences were extracted from one video for each of the 18 infants. Therefore, 54 sequences were selected, lasting 15 sec each. Scenes were selected if only the child was visible (no other persons), with the full body visible and with the child facing the camera. The cry epochs extracted were compliant with previous findings [155, 181] showing an increased F_0 at the 18^{th} month in ASD group. 42 women were enrolled to study their reaction to the cries of the two groups of children. For each epoch, the 42 women were asked to rate if the infant seems distressed. The *distress* level was measured with a 7 steps scale going from *lowest level of distress (1)* to *highest level of distress (7*). Moreover, the women were asked to judge if infants were perceived as normally developed or not, with a similar scale (1-*less typical*, 7-*more typical*). The authors reported that the cry of ASD infant was perceived as more distressed that the cry of TD infants. This difference was statistically significant. In addition, the perception of typicality of infants via the cry resulted significant. The TD infants were judged more typical than the ASD ones. These results were justified by the authors as confirming the significant differences in the cry of ASD infants at 18 months. F_0 was higher in ASD than TD ($p<0.05$). Limitations identified by the authors, a part from those regarding the use of videos, were in the limited number of nullipara women enrolled.

In 2012, Esposito and colleagues [183] repeated the study in [185] with a wider number of adults recruited in Italy and in Japan. A total of 80 adults recruited in the urban area of Trento, Italy and 80 adults recruited in the urban area of Chiba, Japan. The participants in the two countries had a comparable socioeconomic status, ranging medium to high as calculated using the Hollingshead four factor index [196]. Cry stimuli were retrospectively extracted from 20 home videos of unedited cries of 20 firstborn 13-month-olds who belonged to one of two groups: Autism Spectrum Disorders (10 ASD) and Typically Developing (10 TD). The

database of home videos, the modality of cry selection and the cry processing did not differ significantly from the previous study [185]. The selection, the classification and the featuring of signal were as in [155, 181]. Significant differences in F_0 emerged among the stimuli. This results were in line with their previous findings [185] indicating that cries of ASD usually have higher F_0 ($p<0.05$). Cries of children with ASD were judged to express more distress than cries of typically developing children. Neither differences for country nor interaction effects emerged. One limitation of this study is the selection of only two cultures representative of East and West. A further limitation of this work was the use of retrospective home videos to extract cry stimuli. Unfortunately, retrospective studies do not allow identifying and describing the developmental level of children at the time of video recording.

In 2012, Venuti et al. [186] carried out a new study using fMRI to measure brain activity during adult processing of infants' cries listening. They investigated the changes when the adult was listening cry from children with diagnosed ASD and those of children typically developing (TD). Twenty-one healthy adults around 31-35 years of age were recruited through the University of Trento: 11 primiparous parents of TD children, 1 non primiparous father and 9 non-parents. Inclusion criteria were an age between 18 and 40 years, ethnically homogeneous, and with no children older than 3 years old. Exclusion criteria were the presence of neurological or psychiatric disorders, substance abuse/dependence, psychotropic medications, and pregnancy. Prior to the experiment, candidates were screened by a neurologist to check compatibility with MRI scan. A total of 20 acoustic cry sequences of natural cry episodes (10 ASD and 10 TD) from infants were used. Excerpts were extracted from 16 home videos of 16 firstborns taken at 13 months of age. They belonged to one of two groups of children: 8 autistic disorders (ASD) and 8 typically developing (TD). The database of home videos, the modality of cry selection and the cry processing did not differ significantly from the previous study [185]. The selection, the classification and the featuring of signal were as in [155, 181]. During functional scanning, participants passively listened to the acoustic stimuli presented binaurally at ~ 75 dB SPL using headphones. Subjects underwent a single fMRI run in which stimuli were presented in a blocked design. Acoustic stimuli of each category lasted 10 sec, with an inter-stimulus interval of 14 sec during which no stimuli was presented. They found that ASD cries elicited enhanced activity in brain regions associated with verbal and prosodic processing. Authors suggested that the reason was to be found in the alteration of acoustic patterns of ASD cries. In fact, cries determine an increase of activity in brain regions associated with emotional processing indicating that ASD cries also elicit more negative feelings and may be perceived as more aversive and/or arousing. Specifically, cries of children with ASD activated temporo-parietal regions implicated in sound and voice processing as well as brain areas involved in parenting behaviours and, more generally, with processing of emotional stimuli, including thalamus, putamen, and insula. The authors found that cries of children with ASD compared to those of TD children specifically activate brain regions critical for second level acoustic processing in addition to regions associated with the analysis of basic acoustic features. Listening to the cries of children with ASD called for deeper and more effortful auditory attention and comprehension, and in particular comprehension of "emotional content" which may be compromised in the cries of infants with ASD. The authors reported some limitations. First, a

stronger test of the specificity of brain responses to ASD cries might involve comparison of cries of children with other developmental disorders and additional types of child cry. Second, their sample included a mix of females and males, parents and non-parents, and this sample composition modified the generalizability of their findings but the number of participants was too small.

In 2013, Esposito et al. [184] repeated the study published in 2012 [183] with a major number of cases, recruiting 160 adults in Italy an in Japan. 80 Italian adults recruited in urban Trento (North of Italy) were enrolled: 40 parents of typically developing children (20 fathers and 20 mothers around 33 years of age) and 40 non-parents (20 males and 20 females around 28 years of age). 80 Japanese adults were recruited in urban Chiba: 40 parents of typically developing children (20 fathers and 20 mothers around 31 years of age) and 40 non-parents (20 males and 20 females around 29 years of age). Cry selection and analysis was not different from [183]. The 20 audio files were presented randomly to participants (recorded at 44,100 Hz with a stereo resolution of 32 bit) using a personal computer and a headset. Participants were asked to rate the level of distress expressed in the cries (by which we mean perceived distress of the child) and to rate the level of distress they felt (by which we mean distress felt by the adult observer) while listening to the cries. The scales of distress were described in [183]. The authors found no differences among parents and non-parents in judging and how the distress was perceived from cry episodes of child. They found that cries of children with autism are perceived more negatively from two groups.

Many previous studies [183, 185, 186] have pointed to fundamental frequency as the principal negative influence in perceptions of crying and they found that pause length has the strongest impact on perception of distress compared to fundamental frequency and the number of utterances. It is likely that previous research [183, 185, 186] has missed the importance of pauses because vocalizations have mainly been manipulated in terms of F_0 rather than the absence of vocalizations (i.e., pauses).

Episodes of crying in the two groups differed in the duration of pauses (a silence longer than 250 ms within the episode of crying) and in the number of utterances. F_0 of cries were higher for ASD than for TD children. These results were in line with previous findings indicating that cries of ASD usually have different waveform modulations in terms of shorter pauses ($p<0.01$) and fewer utterances ($p<0.05$) as well as higher fundamental frequency ($p<0.05$) [155, 180, 181, 185].

In the Italian group, for both ASD and TD cries, the variable that explained the distribution of expressed and felt distress best was length of the pauses. Specifically, cries with pauses shorter than 1.5 s generated higher levels of expressed and felt distress than those with longer pauses. Next to the length of the pauses, F_0 higher than 608 Hz and than 618 Hz resulted in higher levels of expressed distress and felt distress respectively. The effect of number of utterances was not statistically significant.

The same result was found for Japanese group but with pauses shorter than 2 seconds, that generated higher levels of expressed distress than those with longer pauses. Next to the length of the pauses, episodes of cries with more than 7 utterances resulted in lower levels of expressed or felt distress; those with F_0 higher than 515 Hz resulted in higher levels of expressed distress, and those with F_0s higher than 518 Hz resulted in higher levels of felt distress.

Pause length and number of utterances were as expected correlated, statistically significant correlations between expressed distress and felt distress scores were found in both the Italian and Japanese samples. In this study, authors analysed only audible features of episodes of crying. For this reason, it would be informative to replicate the present study using cries recorded with a large range of sounds. Another limitation of the study it was the recruitment of only parents of typically developing children. A later study might assess parents of children with ASD.

Table 2.1. Full papers exclusion Criteria

	Year	Title	Author of study	Exclusioncriteria
1	2008	How is crying perceived in children with Autistic Spectrum Disorder	Esposito G., Venuti P. [15]	No statistical analysis was presented 'Listen-and-Response' procedure, with questionnaire, was designed to test whether the atypical structure of crying episodes of children with ASD could bias parents' perception.
2	2012	Automatic newborn cry analysis: A Non-invasive tool to help autism early diagnosis	Orlandi, S., Manfredi, C., Bocchi, L., & Scattoni, M. L. [C10]	The topic of the study was pertinent but there was a data validation for the ASD diagnosis of high and low risk groups.
3	2012	Autism spectrum disorder and autistic traits in the avon longitudinal study of parents and children: Precursors and early signs	Bolton P.F., Golding J., Emond A., Steer C.D. [157]	Irrelevant for cry analysis - To chart the emergence of precursors and early signs of autism spectrum disorder (ASD) and autistic traits in the Avon Longitudinal Study of Parents and Children.
4	2012	50 years ago in the journal of pediatrics: The cry latencies of normal infants and those with brain damage	Accardo P.J.[158]	Irrelevant for cry analysis and no specific for ASDs- Cry latency is defined as the time elapsed between the application of a painful stimulus and the onset of crying.
5	2013	Prenatal, perinatal and neonatal risk factors of Autism Spectrum Disorder: A comprehensive epidemiological assessment from India	Mamidala M.P., Polinedi A., P.T.V. P.K., Rajesh N., Vallamkonda O.R., Udani V., Singhal N., Rajesh V.[175]	Irrelevant for cry analysis - A total of 25 factors were evaluated by unadjusted and adjusted analysis in this study., also cry but only with questionnaire for parents and there were not specific conclusions about crying.
6	2013	Effective pre-processing of long term noisy audio recordings: An aid to clinical monitoring	Orlandi S., Dejonckere P.H., Schoentgen J., Lebacq J., Rruqja N., Manfredi C.[J4]	Irrelevant for cry analysis and it is not specific in ASDs- A method for automatic identification of cry was presented.

Table 2.2. Exclusion Criteria

	Year	Title	Author of study	Exclusioncriteria
1	1994	A case of a cholelithiasis in a very low birth weight infant	Ban H., Higuchi R., Noda E., Koike M., Kusumoto S., Keiichirou S. [159]	Notpertinent- AtrialSeptalDefect
2	1997	Offspring-Induced Nurturance: Animal-Human Parallels	Stern J.M. [160]	Notpertinent- Animal/Human study
3	2003	Social and cardiac responses of young children with autism	Sigman M., Dissanayake C., Corona R., Espinosa M. [173]	Not pertinent - study of cardiac distress and autism. Kids watch infants /children that are crying
4	2004	Childhood vaccinations anno 2004. II. The real and presumed side effects of vaccination	Rumke H.C., Visser H.K.A. [164]	Notpertinent - vaccinations and autism
5	2004	Straight-chainacyl-CoAoxidasedeficiencypresenting with dysmorphia, neuro developmental autistic-type regression and a selective pattern of leukodystrophy	Kurian M.A., Ryan S., Besley G.T.N., Wanders R.J.A., King M.D. [167]	Not pertinent - acyl-CoAoxidase deficiency and autism
6	2004	The lessons of MMR	Simon E.N. [162]	Notpertinent - vaccinations and autism
7	2004	Childhood vaccine finance and safety issues	Giffin R., Stratton K., Chalk R. [170]	Notpertinent - vaccinations and autism
8	2004	Pharmacovigilance on MMR vaccine containing L-Zagreb mumps strain	Phadke M.A., Patki P.S., Kulkarni P.S., Jadhav S.S., Kapre S.V. [165]	Notpertinent - vaccinations and autism
9	2005	Communicating the benefits and risks of	Kimmel S.R., Wolfe R.M.	Notpertinent - vaccinations and autism

		vaccines	[169]	
10	2006	Signs of autism in the first year of life	Senator [163]	Notpertinent - only English abstract
11	2006	Transgressors, victims, and crybabies: is basic moral judgments pared in autism?	Leslie A.M., Mallon R., DiCorcia J.A. [166]	Not pertinent - social problem and autism
12	2007	Addressingimmunizationbarriers, benefits, and risks	Kimmel S.R., Burns I.T., Wolfe R.M., Zimmerman R.K. [168]	Not pertinent - social problem and autism
13	2007	Routine vaccines across the life span	Zimmerman R.K., Middleton D.B., Burns I.T., Clover R.D., Kimmel S.R. [161]	Notpertinent - vaccinations and autism
14	2007	The declaration of nutrition, health, and intelligence for the child-to-be	Katzen-Luchenta J. [176]	Not pertinent - social problem and autism
15	2009	The Hunter'sHopeKrabbe Family Database	Duffner P.K., Jalal K., Carter R.L. [171]	Notpertinent - Krabbedisease
16	2009	Best breastfeeding and formulas	Tomas L. [172]	Notpertinent - breastfeeding
17	2009	Is cranio-sacral therapy useful in the management of crying babies? Commentary	Bradley E., Finlay F. [177]	Notpertinent - therapy vs crying
18	2010	Vaccination of children - A systematic review	Ortqvist A., Blennow M., Carlsson R.-M., Hanson L., Lindberg A., Lindqvist L., Magnusson M., Nilsson L., Norlund A., Nyren O., Olcen P., Olin P., Silfverdal S.-A., Sawe J., Soderstrom A., Trollfors B [174].	Notpertinent - vaccinations and autism
19	2011	Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting	Betancur C. [178]	Notpertinent -ASD genetic
20	2011	A controlled study of the risk factors and clinical picture of children with Autism in an Egyptian sample	AbdElhameed M.A., AbdElbaky A.E.O., Kamel E.A. [179]	Not pertinent - Social problem and autism

Table 2.3 Characterization of cry in infants with ADS versus infants typically developing: retrospecitve studies.

AD: Autism Diagnosis; TD: typically developing, DD: developmental delay
**statistically significant with p < .05, ** statistically significant with p < .01*
[+]FOPCI: Free from Other Pathological Conditions (e.g., seizures, Fragile X syndrome) or Impairments (i.e. visual or hearing impairments).
[++]high risk (HR): siblings of children with autism.
‡ADOS: Autism Diagnostic Observation Schedule (Lord et al. 1989; Lord et al. 2000); *DSM-IV-TR:* Diagnostic and Statistical Manual of Mental Disorders (APA 2000)*; IQ:*Intelligence Quotient measured with the "Griffiths Mental Development Scales"; GLM: Generalized Linear Model (Tabachnick & Fidell, 1996)

	Study focus/design	**Sample size, study groups&gender (F/M), IQ mean (standard deviation)**	**Diagnosis of ASD: age at the diagnosis (months) and tests used**	**Dataset and software tools**	**Featuring of signals**	**Results** *(in bolt statistically significant results)*	**Limits**
(Esposito and Venuti 2009) [180]	- Characterization of cry sequences at 12 months - Analysis of audio from home videos	30 FOPCI[+] first-borns: - 10 AD (5/5) - 10 TD (5/5) - 10 DD (5/5) AD Mean IQ‡ = 64 (11).	*Age*: 38.4 *Tests:* - DSM-IV-TR‡ -ADOS‡	2/4 cry epochs from each infant (mean length 32s): - 21 feeding cries (7AD/6TD/8DD), - 13 changing diaper (4AD/5TD/4DD) - 14 startled (5AD/5TD/4DD) Software used: - ObsWin, version 3.2 - General Sequential Querier software for Windows, version 4.1	Time domain Durations, normalized to the length of cry, of: - *Screams*(cry episodes with frequencies between 1-2 kHz); - *Moans* (cry episodes with no harmonic vibrations); - *Aspiration/expiration* phases, (cry sounds that follow the breathing rhythm); - *Pauses* of 1to10s between consecutive cry episodes.	*Test*: -ANOVA - Tukey HSD post hoc tests. In AD: *1)*longer - ***screams*** * - ***aspiration/expiration*** ** *2)*shorter - ***pauses**** 3) No significant differences - *moans*	-Number of subjects -Heterogeneity in ASD severity -Low quality of audio recording in amateur settings.
(Esposito and Venuti 2010a) [155]	-characterization of fundamental frequency (F0) trajectory at 5 and 18 months -analysis of audio from home videos.	30 FOPCI[+] Infants - 10 AD (5/5) - 10 TD (5/5) - 10 DD(5/5) AD Mean IQ‡ = 59 (11).	*Age*: 35.2 *Tests:* - DSM-IV-TR‡ -ADOS‡	240 cry epochs (mean length 24s) 8 cry epochs from each infant: - 4 at the 5th month - 4 at the 18th month From play situations or special events Software used: - Praat	Frequency domain F0 was estimated by averaging the first spectral peaks of the spectrum of each cry for each subject	*Test*: - GLM - Tukey HSD post hoc tests. 5 months: No significant differences 18 months: In AD: *1)higher* - ***F0*** *	-Number of subjects -Heterogeneity in ASD severity -Low quality of audio recording in amateur settings. -No number of cries - No Resonances Freq.
(Esposito and Venuti 2010b) [181]	- characterization of F0 at 18 months -analysis of audio from home videos.	42 FOPCI[+] infants - 14 AD (7/7) - 14 TD (7/7) - 14 DD (7/7) AD Mean IQ‡ = 57 (11).	*Age*: 36.7 *Tests:* - DSM-IV-TR‡ -ADOS‡	210 cry epochs (mean length 22s) from different cries: - 68 feeding cries (24AD/25TD/19DD), - 75 changing diaper (26/23/26) - 67 pain cries (20/22/25) Software used: - Praat	Frequency domain F0 was estimated by averaging the first spectral peaks of the spectrum of each cry for each subject	*Test*: - GLM ANOVA - Tukey HSD post hoc tests. In AD: *1)higher* - ***F0**** *2)*No significant differences - among feeding, changing diaper, pain cries	-Number of subjects, -Heterogeneity of ASD severity -Low quality of audio recording in amateur settings - No number of cries -No Resonances Freq.

Table 2.4 Characterization of cry in Low-Risk (LR) versus High-Risk (HR) infants: prospective studies.

AD: Autism Diagnosis; TD: typically developing, DD: developmental delay; HR: High-Risk;LR: Low-Risk
**statistically significant with p < .05 ** statistically significant with p < .01*
[+]FOPCI: Free from Other Pathological Conditions (e.g., seizures, Fragile X syndrome) or Impairments (i.e. visual or hearing impairments).
[++]high risk (HR): siblings of children with autism.
‡ADOS: Autism Diagnostic Observation Schedule (Lord et al. 1989; Lord et al. 2000); *DSM-IV-TR:* Diagnostic and Statistical Manual of Mental Disorders (APA 2000)*; MSEL:* Muller scale of Early Learning Assessment at their motor and linguistic development (Mullen 1995) ; - SCQ Social Communication Questionnaire (SCQ; Berument et al. 1999);

	Study focus/design	**Sample size, study groups&gender (F/M)**	**Diagnosis of ASD: age at the diagnosis (months) and tests used**	**Dataset and software tools**	**Featuring of signals**	**Results** *(in bolt statistically significant results)*	**Limits**
(Sheinkopf et al. 2012) [156]	- Characterization of cries at 6 months. -Cries were recorded in infant house by one of the authors according to a standardized protocol.	39 FOPCI[+] infants (28 analyzed): - 11 LR (15/16) - 17 HR (10/8)	*Age*: 36 *Tests:* - DSM-IV-TR‡ -ADOS‡ -MSEL‡	First utterance met in a registration 45 min and lasting longer than 0.5s. Cry classification: - 12 pain related cries (5 LR and 7 HR) - 16 not pain related cries (6 LR and 10 HR) Software used: - Develop ad hoc	-Cries were filtered above 5 kHz and digitized at 10 kHz. - A fast Fourier transform to compute the log magnitude spectrum for each 25-msec analysis block within a cry unit was applied. Time domain - Phonation, - Utterance duration - Average amplitude (Loudness) - Amplitude variation (range) Frequency domain - F0 - F0 variation - F1 and F2 - Hyperphonation (F0>1000 Hz)	*Test:* - t-Test Pain-related cries 1) higher in HR: - ***F0**** - *F0 variation* 2) smaller in HR: - *Amplitude variation* 3) No ***hyperphonation*** in LR 4) no significant differences - *Phonation* - *Amplitude* - *F1* - *F2* No significant differences: - *in non-pain-cries* - in gender	-Not utilize standard methods to elicit cries. - Difficulties to distinguish the causes of cry (pain o not pain) -Some missing data on the measure of developmental functioning for 3 AD children. - Problems in F1 and F2 estimation. - Manual analysis for cry-episodes extraction -Small sample size
Esposito et al. 2013a) [182]	- Characterization of cries at around 15 months during Strange Situation Procedure (SSP). - Cries were recorded in a room of the hospital according to SSP protocol.	27 FOPCI infants (about 15) - 14 LR - 13 HR	*Age*: 36 *Tests:* - DSM-IV-TR‡ -ADOS‡ -Mullen Scales‡ - SCQ.‡	First utterance and whole epochs met in a registration 25 min. (Cries were broken into epochs if there was a pause between cry utterances of greater than 5 s) 159 epochs analysed Mean number of epochs per infant: - LR= 2.71 (±0.99) -HR = 2.50 (±2.28) Software used: -Praat	Cries were low pass filtered at 10,000 Hz Acoustic features were estimated by A long-term average spectrum (LTAS), F0 was the first spectral peaks of the spectrum of each cry. Time domain: - Mean duration Frequency domain: - F0 mean - F0 max - F0 variation (range)	*Test:* - ANCOVA First utterance HRvsLR 1) higher in HR - ***F0 **** - ***F0 max **** 2)longer in HR - ***Duration **** 3)No significant differences - *F0 variation*	- Small sample size - Cries recorded in the same assessment room with an overhead omnidirectional microphone.

Table 2.5 Characterization of how adult perceive cry of ADS and typically developing infants: perceptive studies.

AD: Autism Diagnosis; TD: typically developing, DD: developmental delay
statistically significant with p < .05; ** statistically significant with p< .01; * statistically significant with p < .001*
[+]FOPCI: Free from Other Pathological Conditions (e.g., seizures, Fragile X syndrome) or Impairments (i.e. visual or hearing impairments).
[++]high risk (HR): siblings of children with autism.
‡ADOS-G: Autism Diagnostic Observation Schedule - Generic (Lord et al. 2000); *DSM-IV-TR:* Diagnostic and Statistical Manual of Mental Disorders (APA 2000)*; IQ:*Intelligence Quotient measured with the "Griffiths Mental Development Scales"; ADI-R- Autism Diagnostic Interview-Revised; GLM: Generalized Linear Model (Tabachnick & Fidell, 1996); IQ(WPPSI-II):Wechsler Preschool and Primary Scale of Intelligence-II.
[a]There is a tendency but the result is not statistical significant

	Study focus/design	**Sample size, study groups&gender (F/M), IQ mean (standard deviation), IQ [WPPSI-II] mean (standard deviation), adult sample size**	**Diagnosis of ASD: age at the diagnosis (months) and tests used**	**Dataset and software tools**	**Methods**	**Results** *(in bolt statistically significant results)*
(Esposito et al. 2011) [185]	- Characterization of how adults perceived distress in AD vs TD infants. -Adults listened the cry extracted by home videos registered when the infant was 18 months old.	18 FOPCI[+] first-borns - 6 AD (3/3) - 6 TD (3/3) - 6 DD (3/3) AD Mean IQ‡ = 62 (3) TD Mean IQ[WPPSI-II]‡= 102(5) 42 Italian non-mother women with similar schooling	*Age*: 36 *Tests:* - DSM-IV-TR‡ -ADI-R‡ -ADOS-G‡	54 epochs of 18 cry-sequences were analyzed. First 15 seconds of each recording was used. Software used for F0: -Praat	Using visual a 7 step scale the women rated: - the infant distress perceived via the cry - the typicality level of the cry The 7-point Likert-type scales ranged as: - 1 (lower distress) to 7 (highest distress) - 1 (less typical cry) to 7 (more typical cry)	*Cries of AD were associated to:* 1) ***higher distressed infants**** 2) ***lower typical cries** *Discussion* The authors attributed these differences the ***higher F0**** in AD
(Esposito et al. 2012) [183]	- Characterization of how adults perceived distress in AD vs TD infants. -Comparison between Italian and Japanese adults - as (Esposito et al. 2011) but cries registered at the 13 month of life	20 FOPCI[+] first-borns - 10 AD (5/5) - 10 TD (5/5) AD Mean IQ(WPPSI-II)‡ = 96 (7) TD Mean IQ[WPPSI-II]‡= 106(3) 160 adults (50% Female): - 80 parents (50% Italian) - 80 non-parents (50% Italian)	*Age*: 36-40 *Tests:* - DSM-IV-TR‡ - ADI-R‡ -ADOS-G‡	20 first utterances of cry-sequences were analysed. First 15 seconds of each recording were used. Cries occurred after a feeding and when the baby was left alone. Software used for F0: -Praat	Using visual a 7 step scale the adults rated: - the infant distress perceived via the cry - the adult distress felt of the cry The 7-point Likert-type scale ranged as: - 1 (lower distress) to 7 (highest distress)	*Cries of AD were associated to:* 1) ***higher distressed infants***** 2) ***higher distress felt***** No significant difference were observed between the two groups of adults *Discussion* The authors attributed these differences the ***higher F0**** in AD
(Venuti et al. 2012) [186]	- Characterization of how adults perceived distress in AD vs TD infants cries. -Adults listened the cry extracted by home videos registered when the infant was 13 months old. - Measure of dynamic brain activity during adult listening of infants' cries, using fMRI.	16 FOPCI[+] first-borns - 8 AD (4/4) - 8 TD (4/4) IQ *not given* 21 adults - 12 F (6 primiparous mothers of TD); - 9 M (5 primiparous fathers of TD);	*Age*: 36 *Tests:* - DSM-IV-TR‡ -ADI-R‡ -ADOS-G‡	20 cry sequences (10AD/10TD) of 10s were analysed. Software used for F0: -Praat	*Dynamic brain imaging* - participants passively listened to the acoustic stimuli during fMRI - cries were presented binaurally at ~ 75 dB SPL using Serene Sound (Resonance Technologies, Northridge, CA) headphones, with stereo quality sound (40 Hz to 40KHz frequency response) and passive scanner noise attenuation (30 dB) - Acoustic stimuli of each category lasted 10 sec, with an inter-stimulus interval of 14 sec during which no stimuli was presented. *Behavioural analysis* Participants rated on a 4-point Likert-type scale (ranging from not at all to extremely) their feeling of distress when hearing AD and TD cries.	*Dynamic brain imaging* AD cries, compared to TD cries, elicited: - enhanced activity in those brain regions involved in verbal and prosodic processing. - increased activity in the left inferior frontal gyrus/anterior insula. *Behavioral analysis* - Autistic cries may elicit more negative feelings and may be perceived as more aversive and/or arousing. *Correlations* Authors report a significant correlation between (1) the mean difference in ratings of distress perceived for the two types of cries and (2) the change of activation in the left inferior frontal gyrus/anterior. *Discussion* The authors attributed these differences to the ***higher F0**** and ***shorter pauses**** in AD

(Esposito et al. 2013b) [184]	- Characterization of how adults perceived distress in AD vs TD infants cries in the caregiver-child interaction. - Adults listened the cry extracted by home videos registered when the infant was 13 months old	20 FOPCI⁺ first-borns - 10AD (5/5) - 10 TD (5/5) AD IQ‡ and TD IQ(WPPSI-II): typical range - 160 adults (50% Female): - 80 parents (50% Italian) - 80 non-parents (50% Italian)	*Age*: 36-40 *Tests:* - DSM-IV-TR‡ - ADI-R‡ - ADOS-G‡	20 first 15s of cry sequences reordered after feeding were analysed. Pauses between episodes were at least 250ms. Mean number of episodes per infant: -AD= 4.92 (±2.39) -TD= 7.33 (±2.33) Software used for F0: -Praat	Using visual a 7 step scale the participants rated: - the infant distress perceived via the cry - the distress felt while listening the cry The scales ranged as: - 1 (lower distress) to 7 (highest distress)	*Cries of AD were associated to:* ***1) higher infant distress***** **2) *higher felt distress***** No differences among parents and non-parents in judging *Discussion* The authors attributed these differences to the ***higher F0*, shorter pauses**** and ***lower number of cry episodes*** in AD

2.1.5 Discussion

In this paper studies investigating the potentiality of crying as early indicator of ASDs before the 3rd year of life were reviewed. Three different categories of studies were identified: retrospective studies, prospective and perceptive studies. At present the most used is the study through retrospective video analysis on the first year of a child's life. The second technique is to carry out investigations, in the first months of life, on siblings of children already diagnosed with autism. In fact, in this case, the percentage that the second child of a family has this syndrome is equal to 19% [33].

The problem with retrospective analysis could be the lack of a standardized protocol to standardize the type of analysed crying.

For newborn, phonation is based on the development of the coordination of the larynx and, therefore, the study of cry signals can lead to the detection of disturbances of brain functions. There are important characteristic parameters of acoustic emission: fundamental frequency and its time variations, vocal tract resonance frequencies, length and intensity of cry episodes.

Some studies [155, 156, 182, C10] investigate the crying through acoustic analysis to get objective information on the state of health of newborn babies as well as of children a few weeks old. The fundamental frequency was analysed in the 5 retrospective and prospective studies [155, 156, 181, 182, C10] and 4 perceptive ones [183-186]. In [155, 156 181, 182, C10], at 5, 6, 13 and 20 months of life, were reported values similar to Rothgänger studies [115] and they found that the cry of children with ASD had a higher F_0.

Esposito [155] found a decrease in the F_0 in TD children between 5 and 20 months and in TD and DD children there was a change of the F_0 trajectory in infancy, in children with ASD there was no change.

In [182] the analysis of the first utterance of crying indicated a higher fundamental frequency in the HR group, as well as a higher maximum fundamental frequency, at 13 and 15 months of life.

Sheinkopf's [156] results showed that HR infants produced pain-related cries with significantly higher mean F_0 than the LR group.

In Esposito retrospective studies [155, 180, 181] information is missing about the reason that elicited cry: it was reported only "after one hour of the usual nap" and a research assistant selected "only the scenes were the child was crying".

In [C10], F_0 at 10 days, 6, 12 weeks of newborn's life was estimated. The results showed that in most cases F_0 was lower in high-risk infants than in normal subjects in hunger and boring cries, but validation and comparison between LR and HR newborns was performed on 2 HR infants only.

In Esposito studies [155, 180 - 185] the Praat software was used for the analyses. Praat is a free voice analysis tool often used by clinicians developed for the analysis of the adult's voice and not specifically for cry, thus you may get wrong results if it is not applied properly, especially for F_1 and F_2 estimation.

Resonance frequencies (F_1 and F_2) were studied only in Orlandi [C10] and Sheinkopf [156]. In [C10], F_1 and F_2 reached higher values in HR newborns with respect to LR ones. Sheinkopf [156] did not found differences in the two groups, (after the validation of ASDs at 3 years of age), but authors claim to have encountered problems in F_1 and F_2 estimation.

Important results were found about the length and number of cry-episodes. Orlandi [C10] found that HR cases show less cry episodes with long pauses between them. The method of BioVoice tool, used in [C10] appeared robust and well-structured also for the F_1 and F_2 estimation.

Most papers focus on F_0 estimation while other important cry features (such as resonance frequencies and the number and length of cry units) are seldom considered. Thus the poor number of studies about these parameters does not allow to perform a proper Meta-analysis.

The first Esposito's study [180] reported some result about aspiration and expiration phases and different pauses. It was found that infants with autism show less duration of pauses than infants with developmental delays and those who are typically developing.

In [182], it was reported that the first cry utterance of HR toddlers was longer than that of LR toddlers but the total duration of HR toddlers' cries was of shorter duration than that of LR toddlers. Moreover, HR toddlers show a higher F_0 with respect to LR.

All studies [155, 156, 180 - 186] concern newborns validated at 3 years of age for ASD diagnosis, only Orlandi [C10] reported data with a first experimental validation at 12 weeks, carried out by general movements analysis. The cases analysed in this paper are today validated: one of the HR children included in the case study has a Language Delay; all other cases were diagnosed as TD. The comparison for 39 LR and 10 HR infants is reported in the Sect. III of this thesis and showed trajectories as in [C10].

In general, all results suggest that vocal control may be affected in infants at risk for autism.

At present, anyone has defined a standard protocol for the identification of the cry-episodes and for the type of crying to be analysed.

More studies are based on the analysis of structured cries at 18-20 months, when the cry is already a form of communication and expression of a need. Over the past two years studies are focusing towards the first 6 months of life when the crying can be considered spontaneous or a marker of a newborn state (e.g. feeding, sleeping or colics).

Unfortunately, the studies presented several limitation. First of all, perceptual and acoustical studies are based on small populations. There is no standardized method for the identification of cry-episodes

Retrospective home videos were used in 7 studies and a huge number of methodological problems were encountered (e.g., difficulty in controlling variables such as the age of subjects and the length and variability of content and structure of the audio/video tracks). A problem related to retrospective studies of children with ASD (or with other psychopathologies) consists in the impossibility of clearly describing the developmental level of a child at the time of the video recording. However, retrospective video analysis currently appears to be an excellent option for accessing early periods in development, months or years before a child is diagnosed with autism [15, 66, 67].

Cry pitch deeply influences caregivers' perceptions: cries at very high frequency are normally perceived as more aversive and distressing than lower frequency ones [15, 180, 181, 185-186]. These findings have emerged from a series of experimental procedures in which acoustic parameters of the cry (duration and/or fundamental frequency) were estimated.

This systematic review highlights the interest in crying as early indicator of ASD but also highlights the lack of standardized procedures for newborn cry analysis, as well as that of an appropriate recording system and, finally, the need to establish normative ranges for children with a typical development that could help the clinical diagnosis.

The aim of this thesis is to provide such tools emphasizing the importance of methods of analysis dedicated to this type of signals and the definition of stringent registration protocols.

2.2 General Movements: video based techniques

The analysis of the infant GMs is based on perceptual techniques and requires an intense training course and a lot of practice to determine the characteristics necessary to the description of the general movements.

Therefore, this analysis is a subjective method and in recent years, several teams of clinicians and engineers have tested marker-less and marker based techniques, independent of the operator, searching an automatic identification of these movements in order to describe the GMs with quantitative parameters.

The purpose of this section is to review systematically the state of the art of studies from 1993 to date investigating the video-based techniques for the GMs analysis.

2.2.1 Study design

Articles published on scientific journals from 1993 to December 2014 focusing on the analysis of GMs with video-based methods were reviewed.

2.2.2 Search strategy

An electronic search using Ovid Medline, PubMed and Scopus databases was performed. Additionally, a linear investigation among references of retrieved papers was performed too. This search was performed

combining with boolean operators the following keywords: ‘‘infant’’; “infants”; ‘‘baby’’; “babies”; “new-born”; “newborns”; ‘‘newborn’’; “neonatal”; “movement”; “motion”; “motor”; “general movement”; “spontaneous movement”; “video analysis”; “automatic analysis”; “computer-based”; “marker-less”; “video-based”.

Two investigators analysed independently the title, abstract or the full text of each paper to assess the coherence with the objective of the study. In case of disagreements a third researcher reviewed the paper to achieve consensus.

2.2.3 Inclusion, exclusion and classification criteria

Relevant studies published only in English were included if they fulfil the following criteria:

- They included newborns or infants in the first 6 months of life;
- They focus on automatic GMs analysis with video-based methods.

2.2.4 Results

The strategy adopted allowed to find 1550 papers but only 52 pertinent. The number of papers included/excluded are reported in Figure 2.2. In the flow diagram are reported the reasons for the papers excluded in each step of this review. 10 full-texts excluded are reported in Tables 2.6 [152, 153, 197 - 204]. 18 full-texts [154, 205 - 221] were pertinent for the automatic analysis of movements but 12 full-texts were classified as “other infant movements” and not focusing in GMs analysis. These 12 papers reported video-based techniques applied to motion segmentation of epileptic movements and seizure detection [210 - 221].

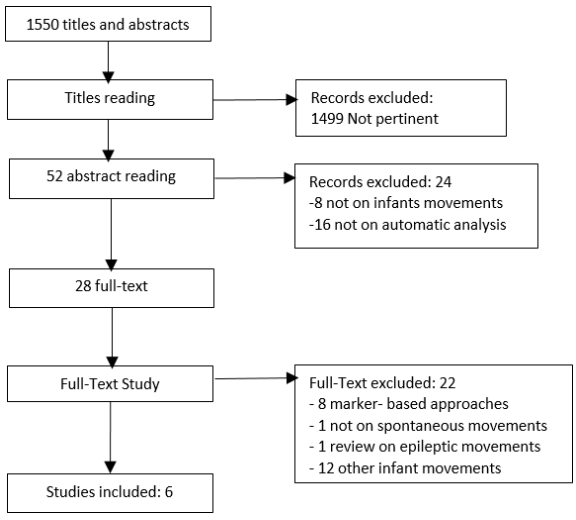


Figure 2.2 GMs - Study flow diagrams

Results and details of GMs studies are described in Table 2.7. These studies used computer vision techniques for the kinematic analysis of GMs. The first paper published in 2009 [154] tried to detect fidgety movements in 82 infants at risk of cerebral palsy. 137 video recording were analysed with a specific software tool (General Movements Toolbox – GMT). GMT showed the motiongrams and the

trajectories of some motor index obtained with a frame differencing. Motiongrams are explained in details in Chapter 7.6.

The quantity of motion of pixels inside a region that exceeded a specified threshold and the centroid of this region were calculated frame by frame. The standard deviation of the centroid allowed to discriminate the fidgety with higher sensitivity (around 80%) obtaining similar results comparing with perceptual analysis reported in [222]. Adde in 2010 [205] repeated the previous study on 30 children at high risk of Cerebral Palsy and calculated the index CPP (Cerebral Palsy Predictor) that allowed to discriminate against children with Cerebral Palsy with sensitivity of 80%. A third study [206], published in 2012, reported a motion tracking method based on optical flow tested on 82 children. The algorithm used was the Horn-Shunck method [223] for the optical flow estimation using for the motion parameters extraction. After the position identification of interest points of image and the velocity estimation in the first frame it is possible to estimate new position and new velocity in the next frames. This method has been shown to discriminate Cerebral Palsy with accuracy above 90%. Rahmati [207] used a tracking method of limb segments based on mean-shift algorithm combining a kernel-based method with a Bayesian filter. An elliptical shape was used for the identification of limbs.

In 2013, another work by Adde [208] tested the predictor of Cerebral Palsy to identify differences between term and preterm infants. In 2014 Rahmati et al. [209] used a new limbs tracking method based on LDOF algorithm (Large Displacement Optical Flow). Motor parameters extraction was based on features identified in marker-based studies [224]. For the first time the area outside the standard deviation, the periodicity and correlation coefficient were estimated with marker-less analysis.

The correlation coefficient of limbs was greater in children with CP. This method was applied on 78 children (14 with Cerebral Palsy diagnosis) in fidgety period. The algorithm was successfully compared with marker-based techniques.

Table 2.6. Full papers exclusion Criteria

	Year	Title	Author	Exclusion criteria
1	2014	Jerky spontaneous movements at term age in preterm infants who later developed cerebral palsy. [197]	Kanemaru, Watanabe, Kihara, Nakano, Nakamura, Nakano, Taga, Konishi	Marker-based method.
2	2014	Computer-based analysis of general movements reveals stereotypies predicting cerebral palsy.[152]	Philippi H, Karch D, Kang KS, Wochner K, Pietz J, Dickhaus H, Hadders-Algra M.	Marker-based method.
3	2012	Increasing selectivity of inter-limb coordination during spontaneous movements in 2-to 4-month-old infants. [198]	Kanemaru N, Watanabe H, Taga G	Marker-based method.
4	2012	Introduction of a method for quantitative evaluation of spontaneous motor activity development with age in infants.[153]	Disselhorst-Klug C, Heinze F, Breitbach-Faller N, Schmitz-Rode T, Rau G	Based on 3D electromagnetic sensors for movement detection.

5	2012	Kinematic assessment of stereotypy in spontaneous movements in infants.[199]	Karch D, Kang KS, Wochner K, Philippi H, Hadders-Algra M, Pietz J, Dickhaus H	Movement analysis based on electromagnetic tracking system.
6	2011	Analysis of adductors angle measurement in Hammersmith infant neurological examinations using mean shift segmentation and feature point based object tracking.[200]	Dogra DP, Majumdar A, Sural S, Mukherjee J, Mukherjee S, Singh A	Study of induced movements, not spontaneous.
7	2008	ENIGMA - Enhanced interactive general movement assessment [201]	Berge P.R., Adde L., Espinosa G., Stavdahl O.	Movement analysis based on signals obtained from sensors.
8	2008	Time series analysis of spontaneous upper-extremity movements of premature infants with brain injuries. [202]	Ohgi S, Morita S, Loo KK, Mizuike C	Acceleration of limbs is measured by tri-axial accelerometers.
9	1999	Kinematic and qualitative analysis of lower-extremity movements in preterm infants with brain lesions [203]	Van Der Heide J.C., Paolicelli P.B., Boldrini A., Cioni G.	Marker has been used to keep track of points of interest.
10	1999	Automatic Detection of Seizures and Spikes.[204]	Gotman, J	Review

Table 2.7: Marker-less General Movements analysis in newborn .

Abbreviations: CP – cerebral palcy, FMs –Fidgety Movements, NT – normal term, NRT near term newborn, ROI –region of interest, PTI-.Preterm infant, HRI- High risk infants, NICU - Neonatal Intensive Care Unit, GMA - general movement assessment, MGA - *median gestational age*, *MBW* - median birth weight, TIA-arterial ischaemic stroke, HIE - hypoxic-ischemic encephalopathy, PTA-post-term age, CPP-cerebral palcy predictor (CPP=(a *Qmean)+(b*Qstd) + (c *Cstd), where a=–2.6(1.5),b=+2.4(1.6), and c=+1.8(1.1).), F+ fidgety present, F- fidgety absent, Fa- fidgety abnormal in nature, HFTI- healthy full-term infants, Q- quantity of motion, C-centroid of motion, std-standard deviation, V-velocity, A-acceleration, TP-true positive, FP- false positive, TN-true negative, FN- false negative, FR-frequency rate, SVM-support vector machines, Ga-gestional age, BW-birth weight, GMFCS- Gross Motor Function Classification System

Author	**Study Focus/Design**	**Simple size (M/F)/ protocol/recordings data set**	**Diagnosis of CP: method and age at the diagnosis**	**Software and Tools**	**Methods**	**Featuring of signal**	**Results**	**Limits**
H.Rahmati, 2014 [209]	-To provide an analytic tool to study general movements and to predict CP with a new markerless segmentation method. -To compare markerless and marker method with motion tracking electromagnetic sensors.	-78 infants (14 confirmed diagnosed with CP) Age 10-18 w PT **Protocol** Infant placed on a standard mattress with rigid, transparent walls. **Recordings** n.a. **Video sequences** n.a.	*Method:* n.a. *Age:* 2 children: 24 months 12 children: 5 years	**Video recorder** Sony DCR-PC 100E stationary digital video camera, installed at a distance of about 110 cm above the infant, recorded the videos with a frame rate of 25 frames/s. **Motion capture system** Minibird motion sensors, electro-magnetic tracking system.	*Perceptual analysis* GMA by clinicians *Automatic analysis* 1) Marker-based Movements of the infants (x, y, and z) were captured with 25 Hz using 6 miniBird motion sensors. 2) Video-based Infants were registered by a video camera. **Motion segmentation** - Dense trajectories of movement for the whole image. **Motion parameters extraction** - Graph-based segmentation algorithm for trajectories of individual body parts. - Computation of one single trajectory for body part. **Motion Classification** SVM-Support vector machines	- Area out of standard deviation from moving-average - Periodicity - Correlation coefficients between trajectories of limbs.	Video-based method vs sensor-based to predict CP: *Sensor data* Sensitivity - 50% Specificity - 92% Accuracy - 85% *Motion-segmentation* Sensitivity - 50% Specificity - 95% Accuracy - 87%	

Adde.L 2013 [208]	-To determine the influence of post-term age at the assessment on sensitivities and specificities for FMs and CP at 2 years. - To show that the use of 2 video recordings improves the accuracy of the computer-based detection of FMs and prediction of CP.	52 infants (24/28) - 19 HFTI - 33 PTI from NICU divided in: 8 extremely PTI (GA<28w,BW<1000g) 10 HRI (GA>28w, BW>1000g) 15 LR/MR-I **Protocol** Infant lying in a supine position, awake and active. **Recordings** 2 videos for each infant PTMA: - 11 wks (9-14 wks) - 14 wks (13-17 wks) **Video sequences** 50 s-5 min	*Method:* European Classification System of CP [53] *Age:*24 months	**Video recorder** Stationary overhead digital video camera (Sony DCR-PC100E). **Analysis tool** General Movement Toolbox (GMT) **Statistical software** **R (**R foundation for statistical Computing)	*Perceptual Analysis* FMs classified by a trained GM observer based on the Prechtl approach of GMA *Automatic Analysis* **Motion parameters extraction** Automatic analysis is performed on video recordings with the method introduced in [30] in 2 different moments during fidgety period. Accuracy in CP prediction was tested with respect to both analysis performed.	- Quantity of motion(Q) Sum all active (white) pixels in the motion image divided the total number of pixel of the image: ✓ Qmean ✓ Qmedian ✓ Qstd - Centroid of motion (C) Spatial centre of the active pixels in the motion image: ✓Cmean ✓Cmedian ✓Cstd - Cerebral Palsy Predictor (CPP)	Same GMA at 11 and 14 wks: - 9 infants F- - 43 infants(F++, F+, F+/-) - No infants with Fa 1) AUC with respect to FM identification for each variable (1st rec/2nd rec /mean): - C_{SD} 0.83 / 0.62 / 0.90 - Qmean 0.68 / 0.69 / 0.77 - Qstd 0.62 / 0.61 / 0.71 - CPP 0.74 / 0.78 / 0.87 - AUC with respect to CP prediction x each variable (1st rec/2nd rec /mean): - C_{SD} 0.82 / 0.81 / 0.88 - Qmean 0.64 / 0.54 / 0.68 - Qstd 0.55 / 0.53 / 0.60 - CPP 0.74 / 0.77 / 0.85	
Rahmati H, 2012 [207]	-To develop a new video-based tracking approach	n.a.	n.a.	n.a.	*Perceptual Analysis* n.a. *Automatic analysis* - Bayesian filtering and kernel-based tracker based on mean shift is used. A new feature space is obtained that uses color information as well. - Updating the kernel profile using information of the previous frame to have a better representation of the target.	Motion information from video captured from infants were been extracted.	The proposed method is compared to the Zivcoviz method [37] for video-based object tracking Experiment results shows the improved performance of the proposed method.	The method has not been tested on any population, or at least it is not reported in the article.
Stahl A. 2012 [206]	-To investigate the classification accuracy of a system exploiting the motion information of the infants, using a state-of-the-art variational optical flow methods.	82 infants -67 normal -15 diagnosed with CP **Protocol** Infant awake, active and comfortable, placed on a standard mattress with rigid, transparent walls. **Recordings** 136 (1-2 recordings for each infant). PTA 10–18 w **Video sequences** n.a.	*Method:* n.a. *Age:* 2 children: 24 months 13 children: 5 years	**Video recorder** Sony DCR-PC 100E camera at 110 cm above the infant. **Software** The feature extraction and selection process along with the training and classification were carried out using C++ programs along with a Matlab/Octave environment. (GPU CUDA)	*Perceptual analysis* n.a. *Automatic analysis* **Motion Parameters extraction** 1) Motion extraction: Optical flow computation and tracking. (Werlberger algorithm) for trajectories 2) Feature extraction: Discrete Wavelet (Meyer wavelet function) and frequency analysis of trajectories. Wavelet coefficients are used for power spectrogram of trajectory. **Motion Classification** SVM (10-fold cross validation)	Feature values extracted by trajectories and used for classification: - Absolute motion distance - Relative frequency - Magnitude of the wavelet coefficients.	Automatic classification of CP: -accuracy(93.7 ±2.1)% -sensitivity(85.3 ± 2.8)% -specificity (95.5 ±2.5)% TP= 12. 8±4, FP=3.0±1.7, TN=64.0±1.7 FN=2.2±0.4. .	- The FR (25 frames/s) is sometimes not enough for accurate of computation. -It happens that tracked grid points get lost.
Adde L, 2010 [205]	- Investigation of the predictive value of a computer-based video analysis of the development of CP in young infants.	30 HRI (13/17) MGA: (31 ± 6) wks **Protocol** Infant were placed supine on a standard mattress wearing a nappy and a bodysuit. **Recordings** - 1 recording at 13wks PTA **Video sequences** 50s- 5 min	*Method:* GMFCS *Age:* 28 children: 4-7 years 1 child: 26 months 1 child with spastic CP died at 12 months before validation	**Video recorder** Stationary digital video camera (Sony DCR-PC100E) placed above the infant. **Statistical software:** SPSS version 17.0 (SPSS Inc, Chicago	*Perceptual analysis* FMs were classified by a trained GM observer based on the Prechtl approach of general movement assessment. ***Automatic analysis*** **Motion parameters extraction** 1) The image is cropped so that only the mattress and infant are shown. 2) The 'motion image' is calculated using frame differencing.	- Quantity of motion(Q) - Qmean - Qmedian - Qstd - Centroid of motion (C) - Cmean - Cmedian - Cstd	13 of 30 infants (5/8) were identified as having CP. (Variable- mean total/mean normal/meanCP) Qmean- 1.5 / 0.7 / 2.2 Cstd- 2.6 / 3.0 / 2.3 Astd- 1.2/1.5/1.1 Vstd- 7.8 / 9.1 / 6.8 Qmean – 2.4 / 1.5 / 3.0 Qstd – 2.6 / 2.1 / 3.0 - Q,Qmean, Qmedian and Qstd significantly >	

				IL, USA).			NG - Cstd velocity, and acceleration > CPG AUC with respect to CP: - CPP - 88 - Qmedian - 85 - Cstd - 84 - Astd - 83 - Vstd - 81 - Qmean - 81 - Qstd - 71	
Adde L 2009. [154]	-To Investigate the ability to detect non-fidgety versus FMs by using the GMT (General Movements Toolbox) -To describe the usability of motiongrams in the study of FMs	82 at risk **f**or CP (37/45) - *High risk* (n=32) - *low risk* (n=50) - *48* PTI (58.5%) MGA = 29.5 wks (23-36wks) *MBW* = 1910 **Protocol** Infants placed in supine position during active wakefulness **Recordings** 137 video recordings *MA* = 13wks (10-18 wks) *ML*= 3.3 min **Video sequences** 0.5–5 min **Video editing** Disrupting movements were omitted.	Long term neurological outcome was not yet collected at the time of the study.	**Video recorder** Stationary overhead digital video camera (Sony DCR-PC100E). **Software** Videos analysed General Movement Toolbox (GMT)	*Perceptual Analysis* GMA on fidgety period. *Automatic Analysis* **Motion parameters extraction** 1) image cropping 2) the motion is identified by the change for each pixel between two frames. 3)Two different filtering techniques were tested on 20 video recordings containing both normal and abnormal qualities of GMs: a) simple low pass filter b) same low pass filter applied after a spatial noise reduction - Method b) was chosen after visual inspection of videos.	Three quantity of motion variables and 5 centroid of motion variables selected for classification: - Qmean - Qmax - Qstd - Cxmean - Cymean - Cstd - Vstd - Astd	Perceptual analysis results: - 27 F- -110 F+ -None Fa (Variable- Mean F+/ mean F-) Qmean (%) - 2.95 / 1.79 Qmax (%) - 32.70 / 29.04 Qstd (%) - 3.20 / 2.41 Cxmean - 4.65 / 4.49 Cymean - 4.31 / 4.01 Cstd - 2.17 / 2.82 Vstd - 6.35 / 8.29 AStd - 1.03 (0.03) 1.35 (0.07) FMs detection (sensitivity/specificity): - Qmean 81.5 / 44.4 - Qstd 81.5 / 44.4 - Vstd 81.5 / 56.0 - Astd 81.5 / 46.4 - Cstd 81.5 / 70.0	-Only on CP risk populations.

2.2.5 Discussion

There are not many research teams interested on automatic analysis of general movements as emerges from this review. This is because this kind of analysis is very complex and based on perceptual analysis concepts. Often protocols recordings are not appropriate and there are not specific standard settings.

For this reason, the indices and the characteristics of motion mostly describe global features: the amount of motion in [154], the comparison of segments as the correlation between the speed of the limbs [206] or the symmetry between the limbs [219]. Among the various methods perhaps the most suitable and most widely used also in the papers described above is that of the optical flow which allows to have a greater quantitative information on movement. This approach has been used in various works on both GMs and other movements and has proven to be a good technique of assessment allowing extraction of useful motor indices for the classification and discrimination of the different kind of movement. These techniques provide quantitative assessment of the parameters for the discrimination of various patterns of GMs in fidgety period (after the 8th weeks of infant's life). The study of Adde [205] has shown that the analysis of the video-based GMs during fidgety helps identify children who later develop Cerebral Palsy with a precision of 85% - 90% and a specificity from 88% to 96%. This is an important result because a markerless approach avoids the use of sensors or markers that might annoy the child or complicate routine tasks in a neonatal intensive care unit.

Marker-based studies, more accurate, allow the assessment of objective parameters to be considered in marker-less analysis and therefore were used as reference for the techniques described in Chapter 5 of this thesis. However, for these systems it is necessary to perform data acquisition in special environments and

with a bulky and expensive equipment, such as the opto-electronic systems. Instead, video-based techniques are much cheaper than marker-based techniques. The analysis carried out on other kinds of patients suggest the use of 3D webcam sensor. Given the length of the recording, specific compression algorithms will be provided to avoid problem of the data considerable size.

Section II - Material and Methods

3. The Acquisition System

Presently, no automatic tool is available for early detection of ASDs. Traditional techniques for the diagnosis of neurological disorders are recently complemented by contact-less methods that provide a semi-quantitative assessment of the patient status. The assessment of infant's behaviour based on the analysis of audio and video recordings is appealing thanks to its unobtrusiveness and to the affordable costs of the equipment. First results were obtained from prospective studies carried out on high-risk infants, based on home videotapes [3, 5, 56] However they suffer from different quality of recordings due to non-uniform video recording systems and low quality of the video clips.

Instead, the acquisition procedure proposed here is standardized, totally not invasive and contact-less, thus minimizing the ethical issues involved in the recruitment of low-risk subjects and high risk infants.

First, the acquisition protocol for the recording of the data used in this thesis will be described.

Then the acquisition system named Audio/Video Infant Recorder (AVIR) will be presented.

3.1 Acquisition protocol

The GR3 project involved the recruitment of a set of low-risk subjects (LR) and a set of high risk subjects (HR) to be followed prospectively. LR are children with no familiarity with ASDs while HR are siblings of at least one child with ASD. In fact, these infants are particularly at risk of developing autism or ASD [6] and this population is arguably the most clearly defined high risk group [225, 226].

The protocol requires the recording of a set of personal data and medical history, together with the acquisition of a video clip and an audio track [227, 228].

The recording tool allows collecting clinical data and information about family members along with weight, length and head circumference of the newborn at each recording session.

Informed consent was obtained from all parents. The protocol was approved by the local ethical committee (Istituto Superiore di Sanità, IRCCS Fondazione Stella Maris and IRCCS Bambino Gesù). The recruitment of the newborns obtained the ethical approval from the Istituto Superiore di Sanità ethical committee in Rome (CE/11/308).

Each subject was involved in a set of measures, scheduled every six weeks, starting a few days after birth up to the 24th week of life. Specifically at: 10 days, 6, 12, 18, 24 weeks after birth. Each session involved the following steps:

• Video recording of the spontaneous movements of the subject. The video recording is performed while the subject is awake and not crying. The perceptive clinical assessment of general movement patterns assumes that the video recording lasts a few minutes (usually 3-5).

• Cry recordings. Only spontaneous cry episodes are considered, therefore excluding pain-related stimulation. The acquisition procedure requires the recording of at least two minutes of crying. However in some cases the records were shorter for reasons related to the situation of the infant. Nevertheless also these recordings have been analyzed.

• Filling of a set of questionnaires: Italian Questionnaire of Temperament; Bayley Scales of Infant Development; the first child vocabulary, MacArthur - Bates Communicative Development Inventory and the Modified Checklist for Autism in Toddlers (*M-CHAT*).

At the age of 6, 12 and 24 months the parents of the enrolled children were invited at Stella Maris and at Bambino Gesù Hospitals to evaluate the child's development through play sessions, structured interviews and clinical tests such as M-CHAT and ADOS 2 - toddler module. At 12 months cells samples of the oral cavity were also picked up by buccal swab in high-risk children and in their older siblings already diagnosed as affected by ASD.

3.1.1 Inclusion Criteria

The inclusion criteria referred to were the following: gestational age greater than 37 weeks, caesarean or natural birth, both males or females (including twins), APGAR index >7 at the first, fifth and tenth minute, age between few days after birth and 24 weeks for the analysis of crying and general movements, age between 6 months and 24 months for the development evaluation protocol.

The APGAR score, the very first test applied to newborns, is assessed in the delivery room right after the baby's birth. The test was designed to quickly evaluate a newborn's physical condition and to understand if there is an immediate need for extra medical or emergency care. Although the APGAR score was developed in 1952 by an anesthesiologist named Virginia Apgar, it is also referred to as an acronym for: Appearance, Pulse, Grimace, Activity, and Respiration. The APGAR test is usually applied to a baby twice: once at 1 minute after birth and again at 5 minutes after birth. Sometimes, if there are concerns about the baby's condition or the score at 5 minutes is low, the test may be repeated ~~for~~ a third time at 10 minutes after birth.

Five factors are used to evaluate the baby's condition and each factor is scored on a scale from 0 to 2, 2 being the best score:

1. Appearance (skin color)
2. Pulse (heart rate)
3. Grimace response (reflexes)
4. Activity (muscle tone)
5. Respiration (breathing rate and effort)

Doctors, midwives, or nurses combine these five factors into the APGAR score, which will be between 10 and 0, 10 being the highest score, though rarely obtained.

The inclusion criteria were the same for LR and HR children. For a baby to be included in the HR group he/she must have at least a brother or a sister with a diagnosis of Autism Spectrum Disorders.

3.1.2 Exclusion Criteria

The main exclusion criteria consist in: preterm birth, APGAR index < 7, impairment of other neurological systems (sensory, extrapyramidal, oculomotor, cerebellar, vegetative) and specific clinical conditions such as mother's intake of toxic substances; severe cardiac, pulmonary, renal, hepatic, endocrine or haematological impairments; genetic diseases; chronic infections or malignant neoplasms; acquired immunodeficiency syndrome or seroconversion for HIV; confirmed diagnosis of psychiatric illnesses. The exclusion criteria were the same for LR and HR children.

3.1.3 Setting and working assumptions

The acquisition system was designed for being used in the patient's home, thus minimizing the discomfort for the involved subjects and the impact of the external environment on children habits. Hence, the basic requirement is the ease of transportation and assembly of the system.

Indeed all recordings were carried out at the patient's home after agreement with parents on the best time for registration based on the behaviour of the child.

The kind of crying recorded in the project is that for hunger. For this reason, the recordings were made during daytime hours after waking up from nap and before feeding. While waiting, usually the operator collected personal data and medical history of the newborn. Upon waking of the baby, the operator could register the cry. After that, the infant was nursed and comforted by the mother. Finally, the general movements (GMs) were video recorded.

If it was not possible to perform the procedure described above, the operator recorded the personal data, measurements and audio / video data according to the needs of the newborn and the mother, that is recording of the GMs first and then of crying. The GMs are completely spontaneous movement and for that reason they must be registered in a calm and waking state, without stressing or disturbing the child. Usually, after a few minutes the baby starts crying for boredom or hunger. In this way it is possible to register crying.

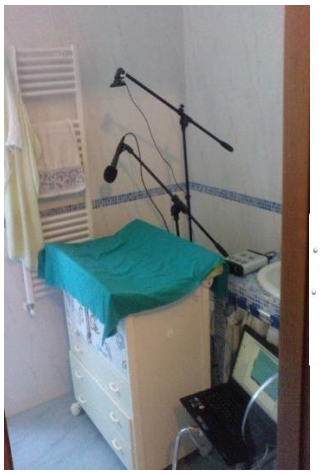
If the baby does not cry spontaneously, the operator may elicit the event through the measurement of head circumference or body length. In fact it has been found that the contact with the yardstick causes a feeling of discomfort.

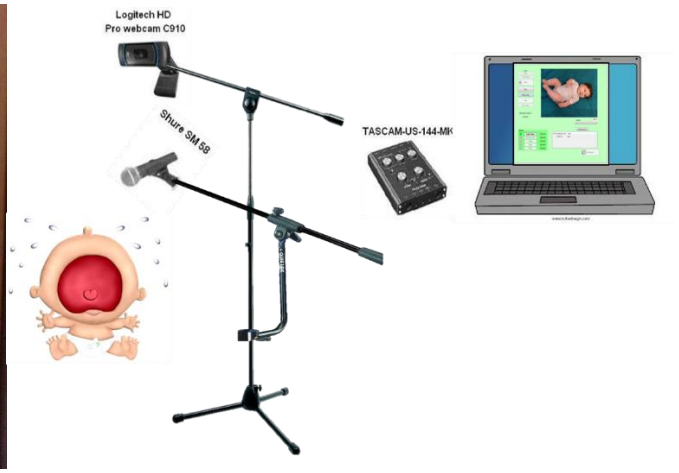
With the overall procedure the main difficulties have been encountered during the first recording session.

The room used for recording had to fulfil standard requirements, that is it must be silent and isolated from outside noise sources, with steady, diffuse and attenuated lighting and a temperature suitable to accommodate the child with a bodysuit with short sleeves or with the diaper only. To simplify the subsequent video processing, a green bed sheet was always placed on the surface chosen for the child placement. This surface could be a bed or a changing table. In the event that the room was too cold or in uncomfortable situations for the newborn the recording was carried out with the bodysuit or onesie.

The specific request of the experimental protocol was to have arms and legs naked. However in many recordings this was not complied as will be explained in the experimental Section. In case it was not possible

to register according to these requirements it was accepted to make recordings with just hands and feet naked. For requirements related to the implemented techniques of video analysis, the use of a neutral colour for bodysuit or onesie was recommended.





3.1 3.2

Figure. 3.1 and 3.2. Example of acquisition room and acquisition equipment: Logitech HD pro webcam C910, Tascam US-144-MK2 for external audio acquisition, Shure SM58 and a microphone stand with a second arm. Laptop characteristics: 2GHz processor, 4 GB RAM and 500 GB for the hard disk driver.

3.2 The recording tool

The systems of audio and video recording currently available, although of high quality, require the manual user intervention for the determination of the duration of the recording and the manual annotation, usually on paper, of the additional information about the state of calm, waking and crying of the subject. Moreover, such systems are currently diversified both as regards the equipment used and the procedures of annotation.

Instead, the standardized acquisition protocol proposed here is based on a customized recording procedure. For these reasons both the standardization of the equipment and the design and the development of a new software tool was necessary.

The recording tool consists of a laptop (2GHz processor, 4 GB RAM and 500 GB for the hard disk driver) connected to a high-speed USB video camera (Logitech HD pro webcam C910), able to provide a 1280x1024 pixel video stream, an external audio board (Tascam US-144-MK2), and a professional unidirectional microphone (Shure SM58).

In the figures 3.1 and 3.2 an example of the acquisition setting and the system components is presented. The system reported in Fig.3.3 shows the acquisition window of the developed software tool, named *Audio Video Infant Recorder* (AVIR) that was installed on the laptop [C10].

According to the protocol AVIR allows the recording of clinical and medical history data, of a video clip for GMs and an audio track for crying. It allows to manage several recording sessions for each subject and the set up of a database to avoid data loss or mislabelling. AVIR is written in C++ language using the MFC

architecture for the user interface, and the OpenCv image processing library for video acquisition and recording.

The user interface includes a main window for selecting the patient in the database, a dialog box for the editing of data in text form and an acquisition window.

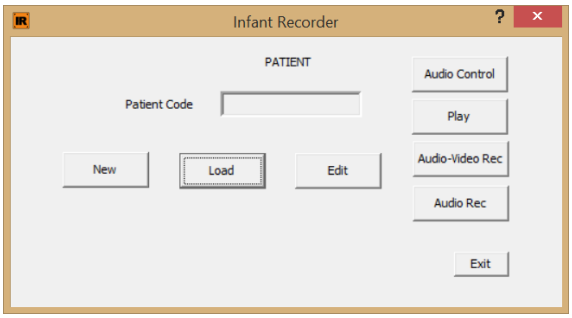


Fig. 3.3 Main interface of AVIR – Audio/Video Infant Recorder

The user can mark the time frames corresponding to different behaviours of the newborn. The markers are stored together with the multimedia tracks and are used to automatically extract the relevant part of the signals to be used in the following steps. These markers are referred to awake, crying or sleep infant's status during the recording as mentioned in 3.3.2.

AVIR was used by clinicians for the acquisition of all recordings made to date in the GR3 project.

3.2.1 Infant Clinical Data

AVIR allows adding infant, parents and siblings data in specific windows. Infants data are saved with a unique code defined by the user. Moreover the user fills up a form for clinical and personal data concerning: name, surname, gestational age, birthday, birth place, social security number (SSN), APGAR index, weight, length and head circumference at birth, possible twin birth, number of siblings, gender, type of childbirth (caesarean or natural), assisted fecundation, incubator information, breastfeeding, possible jaundice, blood group and Rh, fetal distress and kangaroo mother care information.

About parents information AVIR allows to insert the following data: name, surname, birthday, phone number, address, marital status, qualification, socioeconomic context, cultural level, blood group and Rh, height, weight, age at childbirth, number of children, smoke and alcohol consumption information.

About siblings it is possible to fill fields with name, surname, gender, blood group and Rh, clinical information about possible pathologies and possible ASD therapies if the sibling is autistic. An example is reported in fig. 3.4.

3.2.2 Recorder Tool

This is an integrated environment for capturing multichannel audio/video signals, entering notes and contextualized information through the wizard and flexible setup of the system.

The user interface includes a main window for selecting the recording devices and an acquisition window. The acquisition window, shown in figure 3.5, allows previewing both video and audio recordings and checking their quality to avoid blurry or distorted images and ensure a good audio quality avoiding both saturation and too low signal amplitude. Through this window the user can mark the time frames corresponding to different behaviours of the newborn, assess the duration of the audio signal for the acoustic analysis as well as the calm-awake status for the GMs analysis. It is also possible to make multiple recordings on the same patient.

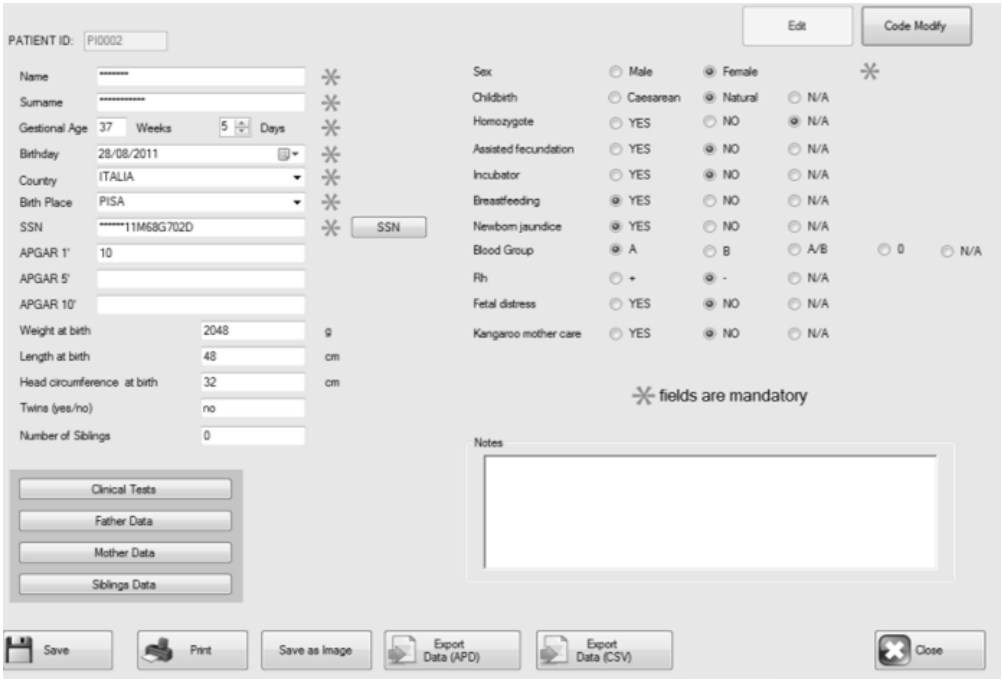


Fig. 3.4 Part of the Infant data interface

Fig. 3.5 The acquisition window

4. Acoustical analysis of the infant cry

As already mentioned in Sect.1 the acoustical analysis of the infant cry is a non-invasive approach to assist the clinical specialist in the detection of abnormalities in infants with possible neurological disorders.
To date, the analysis is most often carried out with a perceptive examination based on listening to the cry and visually inspecting the signal waveform and its spectrogram. However, this approach is operator-dependent and requires a considerable amount of time often prohibitive in daily clinical practice. Thus, the scientific community is paying special attention to techniques devoted to the accurate automatic analysis of the cry.
In Chapter 1 and 2, the physiological motivations and several approaches for fundamental frequency (F_0) and resonance frequency (RFs) estimation are mentioned and described. The difficulty in the estimation of F_0 and RFs is mainly linked to the quasi-stationarity and the very high range of frequencies of interest in the newborn cry that require sophisticated adaptive numerical techniques characterized by high time-frequency resolution.
In this chapter two software tool for the cry analysis are presented. The first one, named BioVoice [107, J3-J5, C10], implements a parametric approach for F_0 and RFs estimation and during the PhD it was improved and tested; the second one named Newborn Cry Analysis (NeCA) proposes a new method for the frequencies estimation based on the wavelet transform approach. It was developed during the PhD project.
Both systems are implemented under Matlab® r2014a environment but could be adapted for any embedded processor.

4.1 Newborn Cry Analysis: BioVoice Tool

BioVoice [107], a multi-purpose voice analysis tool, was developed under Matlab® r2014a. At the Biomedical Engineering Lab., Department of Information Engineering, Università degli Studi di Firenze. It allows the analysis of audio recordings of any length coming both from adults and newborns. BioVoice, is provided with a user friendly interface for uploading the audio file(s) and the type of analysis (adult, singer, newborn, etc.). Several recordings can be uploaded and analysed sequentially with a considerable time-saving with respect to the most widely used tools, both commercial and freely available. Moreover, BioVoice does not require any manual setting thus being well suited also for non-expert users.
Thanks to its high resolution characteristics, the BioVoice tool is used within the GR3 project for newborn cry analysis.
Within this PhD work, some upgrade and improvement of BioVoice was carried on concerning newborn cry analysis especially for recordings of long duration (even several minutes). Moreover, I tested its capabilities on synthetic and real signals. It has been successfully compared with most commonly used software tools on synthesized signals [J4].

Fig 4.1 shows the menu for the choice of age, gender, kind of signal and other options. Fig. 4.2 shows the main interface with a cry signal displayed along with the voiced parts automatically selected by BioVoice (dotted line). (the original figures are in the colour map).

As concerns newborn cry, BioVoice allows the estimation of the following acoustic parameters (and related statistics): fundamental frequency (F_0), the first three resonance frequencies of the vocal tract (F_1-F_3), Power Spectral Density (PSD), spectrogram and time duration of each cry episode.

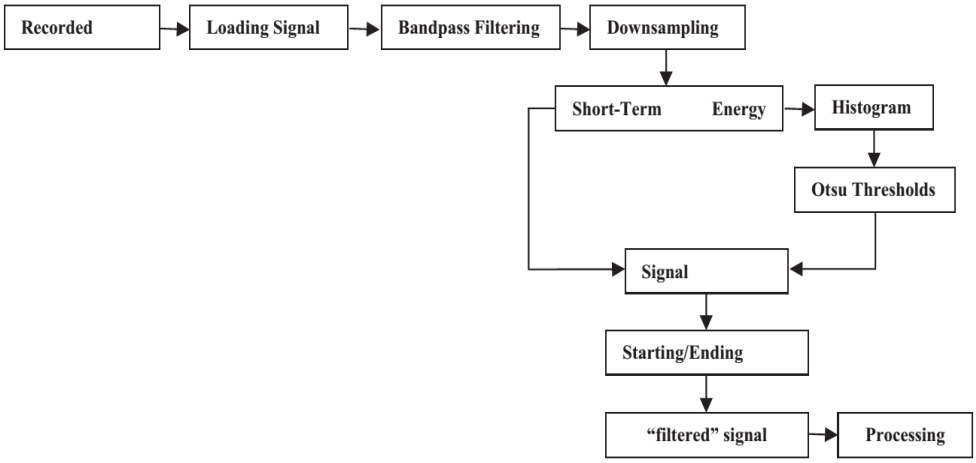
Detection of cry units CU (voiced frames lasting at least 260 ms) is performed using a robust Voiced/Unvoiced (V/UV) detection procedure [J3, J4] that avoids incorrect splitting of a single event into several intervals. This in fact often occurs in the case of irregular and quasi-stationary signals as newborn infant cries are. The methods and steps of analysis implemented in BioVoice are described below.

Cry Units detection:

This is a relevant technical challenge in the case of irregular and quasi-stationary signals as newborn infant cries are, but is indeed a key point that makes the subsequent analysis and estimation of acoustic parameters reliable: only the significant parts of the signal must be taken into account, also avoiding the improper splitting of a single event into several ones that often occurs with commercial and free software tools.

Basically, ambient microphone recorded sounds are composed of two kinds of events. The first one is made of "silence" events characterized by an energy level much lower than the sound (vocalic) intervals but which may also contain short spikes of high energy, however not comparable with a vowel sound; the second one is made up of "sound" events that, according to the application, include voice, snore, cry, but also "other" events such as breathing, ambient noise, etc. The set of both "silence" and "other" events is called here "unwanted" events. It is assumed that "unwanted" events have lower energy and/or shorter duration than "sound" events, that is the recording has been made with the necessary attention to the surrounding environment, which, although not soundproof, must meet the minimum requirements of noise level mentioned earlier.

The problem of extracting only the "sound" intervals of an audio signal is of basic importance in biomedical applications of voice analysis, as well as detecting the starting and ending points of a "sound" event. In [J4] a new method for the automatic detection of meaningful sound events from the whole audio signal was presented named Long Term Audio Analyzer (LTAA). The aim is the fast extraction of the "sound" intervals of audio signal only, while "unwanted" segments are removed. LTAA is depicted in figure 4.1 and described below.



.

Figure 4.1 Flow-chart of the LTAA procedure

To detect voiced and unvoiced (V/UV) frames of the recorded signals (that in the newborn cry correspond to the cry units (CUs)), a two-threshold iterative Otsu method is applied to the histogram of the signal energy. The procedure is an improved version of that in [J5]. Moreover, here, the procedure is tested on synthetic signals that simulate a broader range of voice signals, and results are compared to those obtained with other existing software tools.

The use of high quality microphones can improve signal acquisition; however noise reduction is needed to eliminate interferences and disturbances. Therefore a robust pre-processing step is required. Since most common human laryngeal sounds have main components in the frequency range 50–1000 Hz, the recorded sound is band-pass filtered by a Butterworth filter of order 5 and a cut-off frequency of 50–1000 Hz. This frequency range can be modified acting on the interface. After the filtering step, the signal can be down sampled to a convenient value to speed up signal processing (default setting 11.025 kHz).

Thanks to its computation speed, one of the most used techniques in speech processing is based on Short-Term Energy measure (STE), that increases during "sound" events and decreases during "unwanted" episodes [109]. Thus suitable thresholds on the signal energy are required to determine the boundaries (starting and ending points) of "sound" segments in order to separate "sound" from "unwanted" events.

The pre-processed signal is divided into windows of suitable length (default: 20 ms) with 50% overlap between adjacent windows. In each window the Short-Term Energy is computed as:

$$\mathrm{STE} = \log_{10}\left(\frac{\sum_{i=1}^{n} s(i)^2}{n} + k\right) \qquad (1)$$

where n is the number of samples in the window, s is the signal and k is a small constant to avoid log(0).

In the figure 4.2, a frame of newborn cry (about 10 s) from a healthy full-term female newborn (age: 18 weeks) is shown.

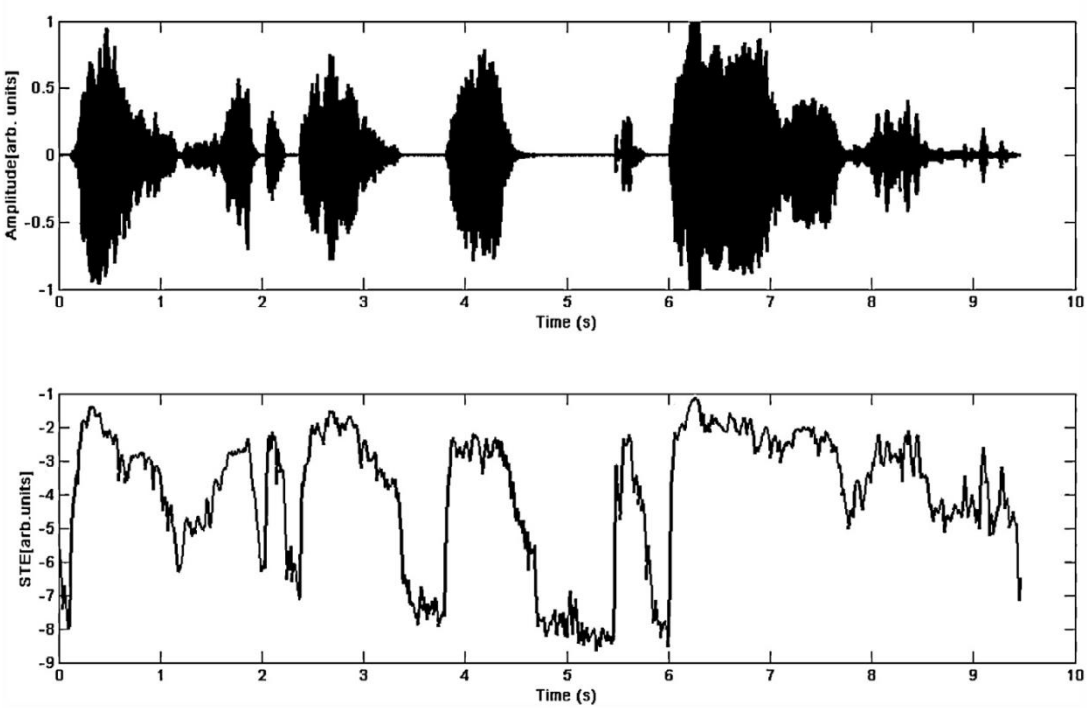


Figure 4.2 newborn cry – healthy full-term female, aged 18 weeks. Signal sample (top) and its energy (bottom). High STE values correspond to voiced events.

STE requires setting a threshold that defines the lowest energy level for a sound signal. Here, the selection of the optimal threshold to separate "sound" and "unwanted" signal frames has been performed using a modification of Otsu's method [229], commonly used in computer vision and image processing to automatically extract objects from their background. Otsu's is a nonparametric and unsupervised method for picture segmentation. It assumes that the image contains two classes of pixels or bi-modal histogram (e.g. foreground and background) and evaluates the optimum threshold that separates the two classes so that their combined spread (intra-class variance, defined as a weighted sum of variances of the two classes) is minimal. An optimal threshold is selected so as to maximize the separability of the resultant classes in grey levels. Otsu showed that minimizing the intra-class variance is identical to maximizing inter-class variance which is expressed in terms of class probabilities ωi and class means μi. The class probabilities are computed from the histogram. The relationship between the within-class and between-class variances has been exploited to generate a recursion relation that permits a much faster calculation, the class probabilities and class means being computed iteratively. This idea yields an effective algorithm with the advantage of reduced processing time [230]. The procedure is very simple and an optimal threshold (or set of thresholds) is selected automatically and stably, not based on a local property but on a global property of the histogram. The range of its applications is not restricted to the thresholding of the grey-level picture, but may also cover other cases of unsupervised classification in which a histogram of some characteristic (or feature) discriminative for classifying the objects is available. Taking these points into account, the method was applied here to (mono-dimensional) audio signals, as the most simple and standard one for automatic threshold selection that can be applied to practical problems. This approach has already been successfully applied to heart sound

signals [231]. In LTAA the signal energy is computed on each signal frame and Otsu's method is iteratively applied to the energy histogram to obtain two thresholds: the upper one tu and the lower one tl. Specifically, a first upper threshold tu is detected that identifies the event. Then a second Otsu thresholding is performed from level zero to level tu to obtain a lower threshold tl to find onset. Using a double threshold avoids incorrect splitting of a single event into several intervals in the case of highly irregular signals. The histogram reports up to 2000 levels, a reasonable compromise between sufficient detail and calculation speed. The procedure is described here in detail:

The signal is divided into frames of fixed lengths, that depend on the type of signal (adult male/female, infant, singer, etc.). Typically the length of the frame is chosen as 20 ms. On each frame the short time energy (STE) is evaluated using Eq. (1). This value is stored in an "energy vector" E. Each element of E thus contains the average energy of the STE signal in that specific frame. On the whole vector E, minimum and maximum energy values are found. The histogram of the energy is then computed and the Otsu method is applied to find the threshold th for the histogram I (which is divided into a convenient number nl of levels). In this application, we use the basic concept of Otsu of maximizing the separation between the classes and the compactness of each class. To do this, the variance between the classes and the internal variance are calculated separately. This step requires to identify the number of occurrences, given by the sum total of all occurrences of the histogram for all 2000 levels. Then, the total momentum of the histogram is evaluated as well as the average and the variance for each level of the histogram. Then, the global variance is evaluated. The first (upper) threshold tu corresponds to the threshold value in which the variance is maximum.

Next the Otsu method is applied again for the calculation of the lower threshold tl. It is evaluated using the same approach as before, but the histogram levels stop at the first threshold tu. Once obtained, the two thresholds are multiplied by a factor d given by the ratio of the differences between the maximum and minimum values of energy and the number of levels of the histogram. This gives the upper (tu) and the lower (tl) thresholds required to determine if a frame is voiced or not. The minimum value of the signal energy is added to tu to guarantee that tu is above the minimum. A schematic representation is reported in Fig. 4.3.

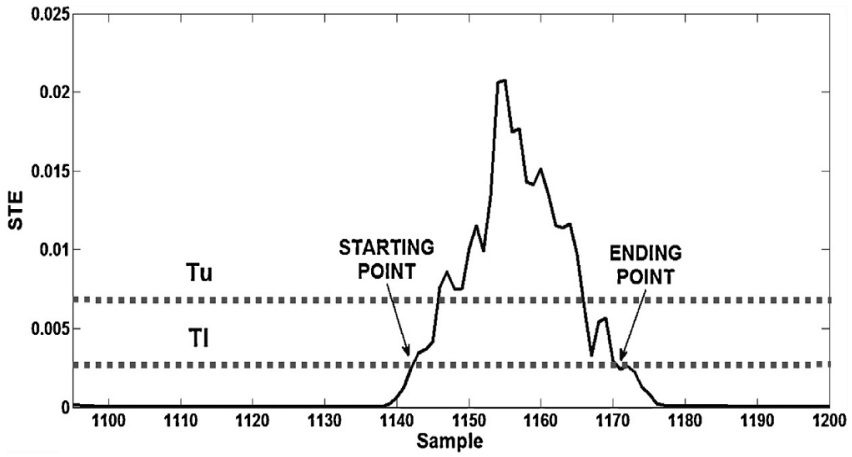


Figure 4.3 The starting and ending points of a "sound" event. When the STE curve overpasses tu, the first point under tl. (on the left side of the curve with respect to tu) is detected. This corresponds to the starting point of the event; when the STE curve falls down tl, the ending point of the event is found.

Once the thresholds have been obtained, the separation of the voiced/unvoiced intervals of the signal in each frame is carried out.

To do this, the energy vector E is scanned and for each element of the vector, it is verified that the value is above tu or not. If it is, then we proceed with the determination of the voiced part. To do this, all the elements preceding the current value of the energy are checked and the first element for which the energy value falls below tl is set as the starting sample of the voiced part. To determine the end of the voiced part, the elements of the energy vector subsequent to the current one are checked: the first element that falls below tl is set as the ending sample of the voiced part. All the frames included in the voiced part are considered voiced. To keep track of voiced and unvoiced frames, a "voice" vector V is created made up of 0 and 1. If the frame is voiced, the corresponding element of V is set equal to 1; otherwise, it is set equal to 0. If the value of energy of the current frame does not exceed tu, but is between tl and tu, the current frame is considered voiced and the corresponding element of V assumes the value 1. If the value of energy of the current frame does not exceed tu and is not between tu and tl, but is below tl, then that frame is considered unvoiced, and the corresponding element of V assumes the value 0. Finally, to guarantee continuity of the voiced frames, all "isolated" voiced frames (i.e. single values equal to 1 in the vector V that are preceded and followed by 0) are reset as unvoiced. At the end of this step, all the "sound" events in the audio signal are extracted and their starting and ending points are stored for further elaboration.

The results show that this approach is convenient for a first fast screening of long-lasting recordings and also allows to exclude vowel sounds such as infant whining that are not of interest in the analysis. The performance of the method was also assessed by qualitative inspection based on audio listening giving comparable results. At present, the algorithm requires about 1 minute to analyse five minutes of crying, but processing time depends on the number and length of CUs detected in the signal. To demonstrate its effectiveness, the method was compared to existing software tools commonly used in biomedical applications using synthetic signals of adult and infant voices [J4].

Moreover during the first PhD year the implemented method was compared with an autocorrelation-based (AC) approach both as concerns accuracy and processing time. With the AC-based method a signal frame is selected as voiced if the maximum of the autocorrelation function on that frame, γmax, is larger than a threshold value, linked to F_0. Results show that the new method performs better that the AC-based one.

The duration of a CU is measured as the time span separating the amplitude onset and offset points obtained with the V/UV selection procedure described above [J4]. In the literature different time durations are considered for CUs, ranging from 60 to 500 ms [122, 129, 134, 235, 236]. However, CUs of very short duration do not allow the assessment of some relevant features such as their melodic shape. Moreover, inspiratory sounds that have a duration less than 200 ms must be disregarded [235]. For these reasons, audio analysis is performed here on CUs longer than 260 ms. However the user can set the proper threshold at any other value as shown in Fig. 4.4 and Fig. 4.5.

Fundamental Frequency estimation: The strength of the method implemented in BioVoice comes from the adaptive procedure for the local definition of the length of each signal frame on which the acoustic parameters are estimated: the higher the F_0 the shorter the length of the frame. F_0 is estimated in the previously selected CUs with a two-step procedure that was shown to outperform other methods [97, 98]. Simple inverse filter tracking (SIFT) is applied first to signal time windows of short and fixed length M=3Fs/Fmin, where Fs is the signal sampling frequency and Fmin is the minimum allowed F_0 value for the signal under consideration (for newborn cry: Fmin=150 Hz). For a more accurate F_0 estimation, in the second step F_0 is adaptively estimated inside [Fl, Fh] where Fl, Fh are respectively the lowest and the highest previously estimated F_0 values. A variable window length for analysis is applied, inversely proportional to the changing F_0. Thus, very short time windows (5 - 15 ms) are obtained, locally dependent on F_0 variability. In each time window the signal is bandpass filtered in the range 150 - 1200 Hz with the Mexican Hat CWT and its periodicity is extracted by means of the average magnitude difference function (AMDF) approach. In case of fast and abrupt F_0 changes this procedure was shown to give enhanced results with respect to standard methods [107]. Results on F_0 estimation were compared to those obtained with the classical Fast Fourier Transform algorithm (FFT) and with other software tools such as PRAAT [117, 118], showing the better performance of the proposed method.

Hyperphonation

BioVoice also allows the analysis of hyperphonated CUs. The problem of hyperphonation, typical of neurological diseases and pain cry, is well known and is usually defined as F_0 exceeding 1000 Hz [114, 115, 156]. In this thesis this is not a relevant point, as feeding cries commonly do not exhibit such high F0 values. Nevertheless, the system is able to detect possible hyperphonated CUs. With BioVoice a CU is defined as hyperphonated if more than 50% of the estimated F0 values are considered outliers (that will be defined later in section 4.3) and their average is above 1000 Hz. In this case, BioVoice alerts the user in the results table that this CU is hyperphonated.

Resonance frequencies: Even if formant frequencies (vowel frequencies) cannot be found in infant cries due to the incomplete development of the vocal tract, resonance frequencies (RFs) reflect important acoustical characteristics. For RFs estimation and tracking a robust parametric technique is used, obtained by peak picking in the power spectral density (PSD) evaluated on the same adaptive time windows described in the previous section. [Manfredi 2000, 2008, 2009] For PSD estimation, on each varying time window, an autoregressive (AR(q)) model is used of order q. The relation $q \cong Fs/2$ (where Fs is the sampling frequency in kHz), coming from acoustic and physiological constraints, was applied giving an enough detailed spectrum while preventing spectral smoothing and consequent loss of spectral peaks. According to the relationship reported above, q=22 with Fs=44100 Hz as in our experiments. This choice has already been proved effective in many applications [112, 237]. This approach was found more robust and with higher resolution capability than the traditional FFT-based technique [13, 81-83].

Co-ordinates of PSD maxima on each time window corresponding to formants, as well as their mean and std value on the whole signal, were evaluated. Specifically, the first three RFs (F1, F2 and F3) are estimated and tracked.

Figure 4.4 – BioVoice signal properties menu. The length of voiced frames can be selected by the user. For newborn cry the range is commonly set between 150 ms and 500 ms

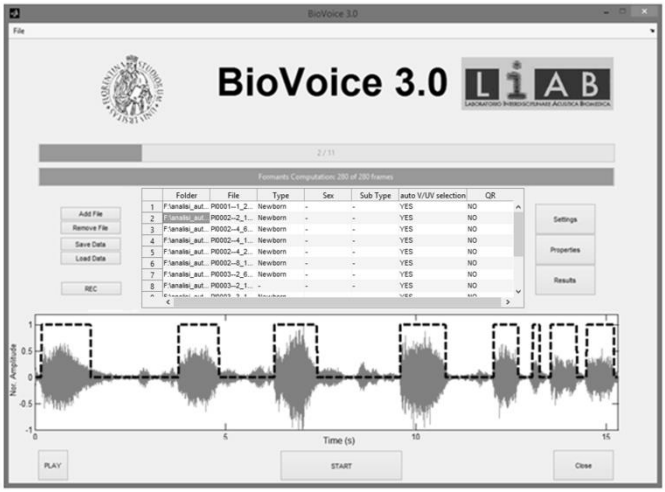


Figure 4.5 – BioVoice user interface. The cry episodes of the selected recording (2th in the list) obtained with the implemented V/UV selection algorithm are displayed.

Statistical parameters

Through skewness, kurtosis and percentiles it is possible to describe the shape of F0. In fact, skewness is a measure of the asymmetry of the data around the sample mean.

The skewness of a distribution is defined as:

$$s = \frac{E(x-\mu)^3}{\sigma^3}$$

where μ is the mean of x, σ is the standard deviation of x, and E(t) represents the expected value of the quantity t. If skewness is negative, the data are spread out more to the left of the mean than to the right

(falling). If skewness is positive, the data are spread out more to the right (rising). The skewness of the normal distribution (or any perfectly symmetric distribution) is zero (symmetric).

Kurtosis is a measure of how outlier-prone a distribution is.

The kurtosis of a distribution is defined as:

$$k = \frac{E(x-\mu)^4}{\sigma^4}$$

where μ is the mean of x, σ is the standard deviation of x, and E(t) represents the expected value of the quantity t. Kurtosis computes a sample version of this population value. The kurtosis of the normal distribution is 3. Distributions that are more outlier-prone than the normal distribution have kurtosis greater than 3; distributions that are less outlier-prone have kurtosis less than 3.

In the most cases the shapes are presented by curves and the idea it is to prove how a F_0 shape is far from a Gaussian curve. This is the reason why it is necessary to calculate the percentiles.

A percentile is a measure at which that percentage of the total values are the same as or below that measure. For example, 90% of the data values lie below the 90th percentile, whereas 10% of the data values lie below the 10th percentile. The percentiles are calculated in a similar way as quantiles. Quantiles are values that divide a (part of a) data set into four groups containing an approximately equal number of observations. The total of 100% is split into four equal parts: 25%, 50%, 75% and 100%. The percentiles and quartiles are computed as follows:

1. The f-value of each value in the data set is computed:

$$f_i = \frac{i-1}{n-1}$$

where *i* is the index of the value, and *n* the number of values.

2. The first quartile Q1 (or lower quartile) is computed by interpolating between the f-values immediately below and above 0.25, to arrive at the value corresponding to the f-value 0.25. This is the same thing as the twenty-fifth percentile.

3. The third quartile Q3 (or upper quartile) is computed by interpolating between the f-values immediately below and above 0.75, to arrive at the value corresponding to the f-value 0.75.

4. Any other percentile is similarly calculated by interpolating between the appropriate values.

The interquartile range, IQR, is defined as Q3-Q1.

From the BioVoice analysis we obtained F_0 vectors varying from 4 to 698 points and a length variable from 260 ms to 5,96 s. We considered only F_0 vectors with more than 10 values. In a first step, percentiles were estimated at 25th, 50th and 75th only, afterwards to obtained a more accurate-shape description we calculated the 10th, 20th, 30th, 40th, 50th, 60th,70th,80th 90th and 100th percentiles. In this way we obtained 10 values for each CU.

Results report: BioVoice computes the main acoustic parameters described above and related statistics (mean, median, standard deviation, maximum and minimum of F_0 and F_1-F_3, curtosis, skewness, percentiles) that are saved in separate folders, one for each recording, in an easily usable format (jpeg, excel, txt).

Moreover it computes the number and duration of each CUs in each recording, the vocalic percentage, the number and length of the voice breaks. Moreover, the values obtained for each recording were averaged; BioVoice provides a report of the averages of all the cases analysed.

Finally BioVoice provides several plots (F_0, spectrogram, resonance frequencies F_1-F_3, PSD).

Some example are shown in figure 4.6, 4.7 and 4.8.

The parameters obtained with BioVoice are saved in an Excel spreadsheet along with the following: the filename or patient's code, the recording session, the number of CUs in the recording, starting and ending time instant of each CU, the time instant of the occurrence of F_0 maximum and minimum, kurtosis, skewness, 10% percentiles (thus 10 values), 25th, 50th and 75th percentiles of F_0, the number of estimated F_0 values, the number of possible outliers (F_0 > 900 Hz) and a flag indicating if the CU is hyperphonated. Each row of the spreadsheet is thus made up by 39 parameters.

Figure 4.6 shows the V/UV selection from the first 35s of a cry recording. Figs. 4.7 and 4.8 respectively show F_0 and the spectrogram with F_1-F_3 superimposed for the first cry unit of the example in Fig.4.6. Original figures are in the colour map.

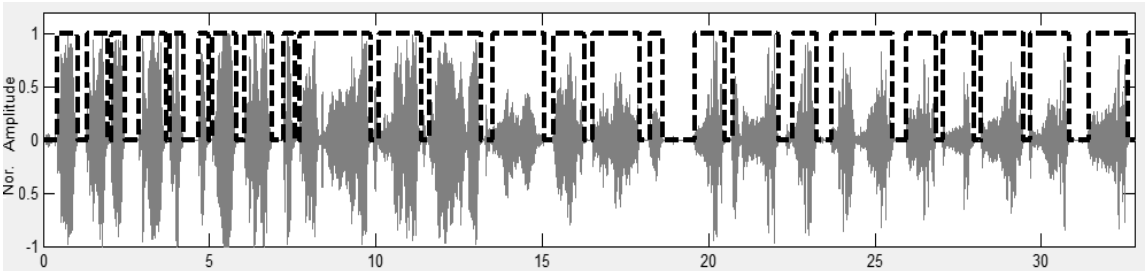


Figure 4.6 V/UV selection for a cry recording of 35s of duration.

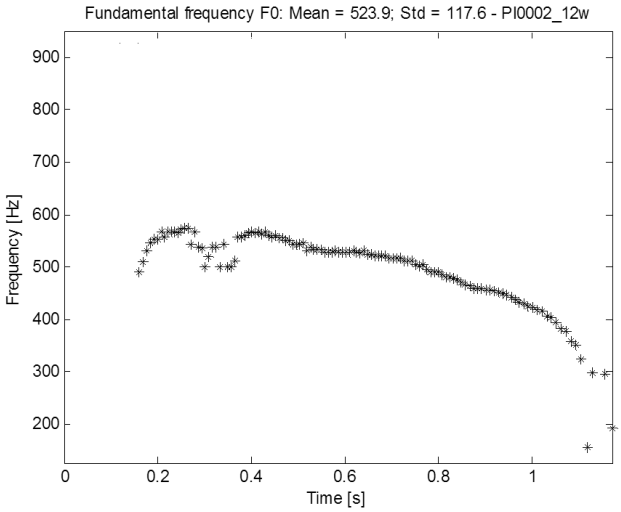


Figure 4.7 Time evolution of F_0 for the first CU of the example shown in Fig. 4.3.

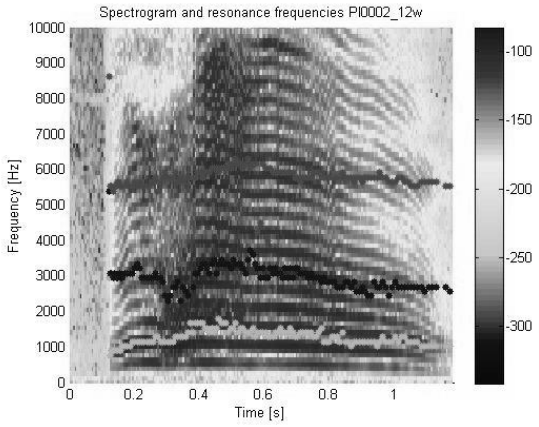


Figure 4.8 Spectrogram and the first three resonance frequencies F1-F3 for the CU shown in Fig. 4.4

4.2 Newborn Cry Analysis: NeCA Tool

During PhD work a new software was developed for the estimation of F_0 and RFs based on wavelet transform. The wavelet approach is in fact particularly suited to the study of neonatal cry, characterized by quasi-stationarity and very high range of frequencies of interest, thanks to its time-frequency high resolution characteristics and low computing time. The new tool was also useful for testing BioVoice capabilities. This section presents a first attempt to apply wavelets to the analysis of newborn cry. The necessary steps for the selection of the best mother wavelet were implemented both for discrete and continuous wavelet transform (DWT and CWT). This choice was compared with mother wavelets reported in the literature. Specifically, in [238] the Meyer wavelet is selected for the best discrimination between "normal" and pathological cries while in [239] and in [240] a discrete wavelet transform with Daubechies mother function is applied. However in these papers the wavelet transform was not used for the estimation of any acoustical parameters as in this PhD work.

For the continuous wavelet (CWT) existing work on adult voice signals is considered, where the Mexican Hat [241] and the complex Morlet mother wavelet were applied for the F_0 and formants estimation [242]. It must be pointed out that also in this case no application exists to newborn cry.

The implemented approach, named NeCA (Newborn Cry Analysis) is developed under Matlab R2014a. NeCA is tested and compared on synthetic signals with BioVoice and PRAAT as will be reported in what follows.

To select the method for the F_0 and RFs estimation a comparison between Discrete and Continuous Wavelet Transform was performed. The detection of the CUs, where F_0 and RFs are estimated, is performed with BioVoice.

Continuous wavelet transform

The wavelet transform filters a signal *f*(t) with a shifted and scaled version of a prototype function ψ(t), the so-called "mother wavelet", a continuous function in both the time domain and the frequency domain [243]. The Continuous Wavelet Transform (CWT) of f(t) is defined as [244];

$$CWT(a,b;f(t),\psi(t)) = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{a}}\psi^*(\frac{t-b}{a})dt \qquad (1)$$

Where $a \epsilon R - \{0\}$ is the scale parameter, $b \epsilon R$ is the shift parameter and ψ* is the complex conjugate of ψ. Unlike the Fourier Transform, the CWT allows a time-frequency representation of a signal with very good time and frequency localization. The scale parameter a is related to the width of the analysis window: it either dilates or compresses the signal. The shift parameter b locates the wavelet in time. Varying a and b allows locating the wavelet at the desired frequency and time instant [244]. The relationship between a and the frequency is given by the so-called pseudo-frequency (F_a) in Hz, defined by the following equation:

$$F_a = \frac{F_c}{a\Delta} \qquad (2)$$

where Δ is the sampling period, and F_c is the wavelet central frequency.

The Mexican Hat CWT is defined as:

$$\psi(t) = \frac{2}{\sqrt{3}}\pi^{-\frac{1}{4}}(1-x^2)e^{-\frac{x^2}{2}} \qquad (3)$$

For each time window and in the frequency band of interest (Table 4.1) the highest coefficient of the CWT matrix is found. The autocorrelation (AC) is computed on the row of the matrix that contains this value, which corresponds to the optimal scale. F_0 is given by:

$$F_0 = F_s/\tau \qquad (4)$$

Where τ refers to the position (lag) of the maximum of the AC.

The estimation of $F_1 - F_3$ is performed in a similar way, with different ranges for the band-pass filter as reported in Table 4.1.

The complex wavelet transform provides a more accurate representation of the oscillatory components within the signal without introducing fluctuations in the coefficients [244].The complex Morlet wavelet is described by the following relationship [245]:

$$\psi(t) = \frac{1}{\pi^{\frac{1}{4}}}e^{j\omega_c t}e^{-\frac{t^2}{2\sigma_t^2}} \qquad (5)$$

where: $\frac{1}{\pi^{\frac{1}{4}}}$ is the normalization term of the energy; $\omega_c = 2\pi F_c$ is the center frequency of the wavelet; $\sigma_t$is the standard deviation (SD), that is the scale parameter which determines the amplitude of the wavelet. In fact $\omega_c\sigma_t$ sets the link between the bandwidth of the wavelet and its frequency F_c. For a wavelet family this amount must be constant. For the Morlet wavelet, the latter must assume values such that [241, 242]:

$$\omega_c\sigma_t \geq 5 \qquad (6)$$

The method applied is similar to that used for the Mexican Hat. Moreover the following relationships are taken into account:

$$\omega_c = 2\pi F_c \qquad (7)$$

$F_b = 2{\sigma_t}^2$ (8)

Where F_b is the bandwidth of the wavelet. Comparing the frequency ranges and on analogy to [241] the values of F_c and the corresponding values of F_b were set as in Table 1. Specifically, for each F_c relative to each frequency band, F_b was computed with $\omega_c \sigma_t = 5$ and according to Eq. (7) and (8).

Table 4.1. Frequency bands of interest in newborn cry, center frequency Fc and bandwidth Fb for the complex Morlet

Frequency band [Hz]	F_c [Hz]	F_b [Hz]
F_0 [200 - 800]	1	1.27
F_1 [800 - 2100]	0.8	1.98
F_2 [1500 - 3500]	0.75	2.25
F_3 [3400 - 5500]	1.5	0.56

Discrete Wavelet Transform

The DWT allows the decomposition of the signal into approximations (low frequency components) and details (high frequency components) [244, 246]. At each decomposition level a "downsampling" is performed, i.e. the removal of one out of two samples. Therefore the maximum level *n* of decomposition is given by:

$n = log_2 N$ (9)

Where N is the number of samples of the signal.

The DWT applies dyadic scales iteratively dividing Fs by 2. It is therefore necessary to set the level at which one wants to decompose the signal. Its choice is linked to the sampling frequency F_s and to the range of frequencies from F_0 to F_3.

Also with the DWT it is necessary to choose a mother wavelet. According to the literature [239, 240] the Daubechies (db) wavelets is applied here. The best results for F_0 were obtained with a mother wavelet db5 up to level 5 (1378.125 Hz); for F_1, F_2 and F_3 with a mother wavelet db3 and decomposition level 4 (2756.25 Hz), 6 (5512.5 Hz) and 8 (5512.5 Hz) respectively. In the literature references are found only for the estimation of F_0 and F_1 [239, 240] that agree with the levels mentioned above.

Estimation of F_0

For infants F_0 values are usually in the range 200 Hz - 800 Hz [115].

For the CWT (Mexican Hat and Complex Morlet) the proposed method involves the following steps:

1. Application of a bandpass filter FIR with Kaiser window properly sized according to Table 4.1;
2. Application of the CWT to the signal. A pxq matrix M of coefficients is obtained, where p = maximum value of the scale and q = number of frames of the signal;
3. Location of the scale (line) of M corresponding to the coefficients of maximum modulus and estimation of F_0 according to eq.(4).

For the DWT transform the steps are:

1. Application of Daubechies DWT with mother wavelet db5 until the fifth level of decomposition;
2. Computation of the AC of the vector of the coefficients of approximation;
3. Computation of the maximum of the AC and estimation of F_0.

Estimation of RFs

Typical values for the first three RFs are approximately 1000 Hz, 3000 Hz and 5000 Hz [7-9]. Significant deviations from these ranges may be related to pathological conditions of the central nervous system.

The estimation of F_1 –F_3 is carried out with a procedure similar to that used for F_0 but with different ranges for the band-pass filter, according to Table 4.1. Both the Mexican Hat and the complex Morlet CWT are tested. The RFs estimation by means of DWT is similar to that of the CWT but without the bandpass filter.

Validation of method

For F_0 both CWT and DWT wavelets were tested on a sine wave at 450Hz:

$y(t) = \sin(450\,t) + e(t)$ (10)

For the RFs F_1-F_3 a sum of three sinusoids is considered on analogy to [242]:

$y(t) = 5 * \sin(1000t) + 10 * \sin(3000t) + 15 * (5000t) + e(t)$ (11)

White noise e(t) set at 5% of the signal amplitude was superimposed through the Audacity® open source tool. The signals were sampled at 44100 Hz. The results were compared with those obtained with BioVoice and PRAAT.

Besides MDVP™ [116], a commercial software widely used in clinical practice, recently the freely available online tool named PRAAT is widespread [117]. For F_0 estimation it implements a method based on the AC applied to a time window of fixed size while Linear Predictive Coding is applied for the RFs estimation. PRAAT allows the analysis of a single audio file at a time. For proper use, and especially with newborn cry RFs, it requires the manual setting of some parameters. Therefore its use must be made with caution [119]. Thus in this work the best parameters for PRAAT were preliminarily tested and set. Specifically, the range for F_0 was set at 200-800 Hz while for F_1-F_3 the maximum range was set up to 11025 (Fs/4) with the estimation of 5 formants instead of 3. The use of default values (5500 Hz and 3 formants) leads to wrong results.

MDVP, BioVoice and PRAAT have been compared both with regard to the selection of CUs [J4] and for the estimation of F_0 in the presence of jitter and noise. The results show that BioVoice is able to estimate the CUs with the greatest accuracy and is also the most robust against jitter [J3].

Estimation of F_0

A preliminary test was carried out on the sinusoid in Eq. (10) without adding noise.

On each time window of length 10 ms the scale parameter *a* was allowed to vary in the range 1÷55. This choice is related to a reasonable frequency range for F_0: 200 Hz-1050 Hz [3] and up to 5.5 kHz for RFs. Therefore the Mexican Hat CWT was applied with a = 55, $\Delta = 1/F_s = 1/44.1$ s, Fc = 0.25 Hz. Consequently

Fa = 200 Hz according to Eq. (2). Thus on each time window a CWT matrix (size: 55x441) of real coefficients is computed. The row index of its maximum entry corresponds to the 'optimum' scale value, *a*. The AC of this row gives the estimated value of F_0 according to eq. (4). The best result is given by the Mexican Hat which estimates F_0 at exactly 450Hz, while the complex Morlet overestimates it of about 15 Hz. With the db5 DWT the best results are obtained with an underestimation of 5-10Hz.

These first results show that the CWT Mexican Hat allows to obtain better results than the other methods. Thus it was also tested on Eq. (10) with additive white noise and compared with BioVoice and PRAAT.

Table 4.2 shows the results obtained with the three approaches. The CWT has the best performance, as well as PRAAT (set with optimal parameters) though with a slightly higher standard deviation (STD), while BioVoice slightly underestimates F_0 (0,26%).

Table 4.2 – F_0 estimation. Comparison of BioVoice, PRAAT and CWT Mexican Hat on a synthetic signal (sinusoid at 450Hz with 5% white noise)

Method	**F_0 mean**	**STD**
Mexican Hat	450,00	0,00
BioVoice	448,81	2,08
Praat	450,00	0,88

Estimation of RFs

On analogy to F_0, a preliminary test was applied with the synthetic signal in Eq.(11) without added noise. The CWT Mexican Hat provides the best results for F_1 (only 2 Hz of overestimate) while underestimates F_2 (250 Hz) and overestimates F_3 (500 Hz). The CWT Complex Morlet gives the best results, with 5 Hz of underestimation and overestimation for F_1 and F_2 respectively and overestimation of about 90 Hz for F_3. The DWT provides worse results overestimating all three RFs (100 Hz, 200 Hz and 900 Hz, respectively). The CWT Complex Morlet was then applied to the signal in Eq.(11) with 5% white noise. Table 4.3 shows that the CWT Complex Morlet provides good results especially for F_1 and F_2. All methods give comparable results although with significant differences on STD. BioVoice gives the best results, with the lowest STD for all RFs.

Table 4.3 – F_1-F_3 estimation. Comparison of BioVoice, PRAAT and CWT Complex Morlet on a synthetic signal (sum of sinusoids with 5% white noise)

	F1 mean	**F1 STD**	**F2 mean**	**F2 STD**	**F3 mean**	**F3 STD**
Morlet_CWT	1024,09	91,00	2971,27	170,41	5163,62	411,32
BioVoice	985,47	5,12	2956,42	8,11	5050,56	11,20
PRAAT	1120,50	387,37	3068,69	346,26	5019,35	147,02

According to these results NeCA implements the Mexican Hat CWT for F_0 estimation and the complex Morlet CWT for RFs estimation. Results concerning the cry data are reported in the section devoted to the experimental results.

4.3 Melody analysis

As reported in Sect. II, the newborn cry melody (time-domain F_0 contour) has recently gained much scientific interest [120-125]. There are several kinds of melody shapes. Schönweiler et al. [124] classified the melody shapes of crying whose main categories were: falling, rising, rising-falling, flat.

- The Falling (F) curve is characterized by a rapid rise in the first part of the interval and a slow descent until the end.
- The Rising (R) curve has a pattern symmetrical to F. In this case, therefore, it has a slow climb up to the last part of the interval, where there is a steep slope that ends with the end of the episode.
- The rising-falling (henceforth called symmetrical (S)) curve is described by an increasing-decreasing trend of F_0 around a maximum approximately positioned in the middle of the CU.
- A flat curve (henceforth called plateau P) is described by an almost constant F_0 (variation less than 100Hz) value around its median.

Fig. 4.9 shows such ideal patterns, while Fig. 4.10 highlights some melody shapes in a real cry

However, there are other possible melody shapes. Várallyay [123, 125] found up to 77 different shapes but out of them there were 20 which include the 95% of melodies.

Várallyay [123] also pointed out that there are three main attributes of a crying segment which should be used for the automatic segmentation:

- The amplitude (i.e. energy) is quite high.
- The duration is longer than a few tenths of seconds.
- The spectral structure is regular.

In this work we will restrict the classification to the 4 main shapes (R, F, S and P) described above.

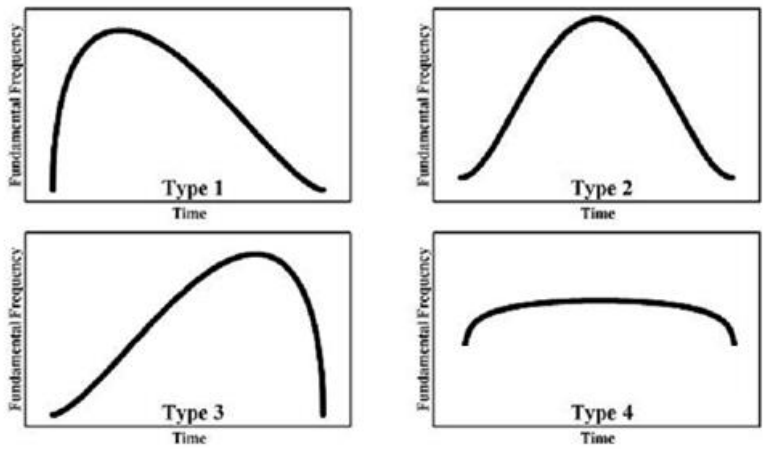


Fig. 4.9 plots modified from [120]. Top left: falling; bottom left: rising; top right: symmetric; bottom right: plateau

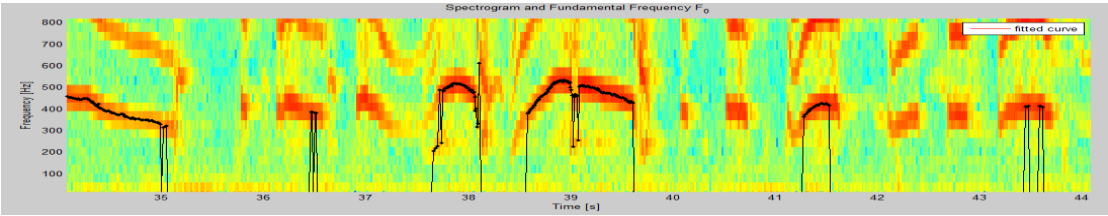


Fig. 4.10 Several shape of F_0 are shown in the spectrogramm of a cy signal of a LR infant at 10 days of life.

As reported in Sect. II to date there are few studies [111, 123, 125 J2] concerning the automatic assessment of the melody, resulting in semi-automatic approaches. During the last PhD year a preliminary study on the melody was carried out for the four principal categories listed above [120, 123]). A first attempt was based on both time-domain and frequency-domain approaches. Several methods were tested such as spline functions for fitting and k-means (described in chapter 4.4.2 for classification in time-domain and Fourier coefficients (from 4 to 16) in the frequency domain, however the results were unsatisfactory and are not reported here.

Thus a new method was implemented that is summarized here, the whole procedure being quite complex due to the high time-frequency variability of F_0 (an example is shown in-fig. 4.10).

The first step of the automatic analysis is a test on the outliers (very irregular points) in the first 5% and the last 5% of F0: if the standard deviation in these parts is greater than 20 Hz then they are cut off. This stems from a thorough visual analysis where it was noted that F_0 often contains much "noise" in those areas, probably due to the difficulty of the infant to initiate phonation.

To classify the melody shape it was necessary to delete the F0 outliers in each CU. Here outliers are defined as three interquartile ranges below the 25th percentile or above the 75th percentile.

This is performed with the following iterative method (Tukey's method) based on empirical thresholds.

In 1977, Tukey [247] developed a helpful method based on boxplot that makes no assumptions about the distribution nor depends on a mean or SD standard deviation. The lower quartile (q1) is the 25th percentile, and the upper quartile (q3) is the 75th percentile of the data. The inter-quartile range (IQR) is defined as the interval between q1 and q3. Tukey [247] defined q1-(1.5*iqr) and q3+(1.5*iqr) as "inner fences", q1-(3*iqr) and q3+(3*iqr) as "outer fences", the observations between an inner fence and its nearby outer fence as "outside", and anything beyond outer fences as "far out". Later, High [248] renamed the "outside" potential outliers and the "far out" problematic outliers. The "outside" and "far out" observations can also be called possible outliers and probable outliers, respectively. Although Tukey's method is quite effective when working with large data sets that are fairly normally distributed, many distributions of real-world data do not follow a normal distribution, as our data are. In this case, q1-(1.5*iqr) was used for a mild outlier detection.

After that, the CUs of length <260 ms are discarded and classified as "too short". This value, that can be changed by the user, was set according to the literature and allows to discard too short episodes that therefore

may be improperly classified as CUs but that often correspond to the period of inhalation of air and not to cry.

To perform the classification of CUs according to the basic melody shapes we defined three ideal reference curves (equations) corresponding to F, R and S respectively, with which each CU is compared. These equations and the corresponding coefficients come from the interpolation of a number of points large enough to obtain a sufficiently smoothed curve (10-20). To make the comparison, these ideal curves are normalized both in time and in amplitude (frequency) so that they can be applied to CUs with characteristics very different from each other.

The Plateau (P) curve is not defined here by any particular equation. A CU whose F0 values varies less than 100 Hz with respect to its median is considered a Plateau.

The symmetric curve (S) is described by a second order polynomial equation:

y = a + b * x + c * x. ^ 2;

whose coefficients are: a= -1.12; b= 3.786; c= -0.379.

The Falling (F) curve is characterized by a rapid rise in the first part of the interval and a slow descent until the end. Its maximum value is located in the first third of the total length of the signal. In this case the equation is of the fourth degree:

y = a + b * x + c * x. ^ 2 + d * x. ^ 3 + e * x. ^ 4;

with the following coefficients: a= -0.85; b= 9.457; c= -2.643; d= 2.7E-1; e= -1E-2.

The Rising (R) curve has a pattern symmetrical to F. In this case, therefore, it has a slow climb up to the last part of the interval, where there is a steep slope that ends with the end of the episode. The maximum value is located at its end (last third). Similarly to the previous case the shape of this curve is described by a fourth degree polynomial:

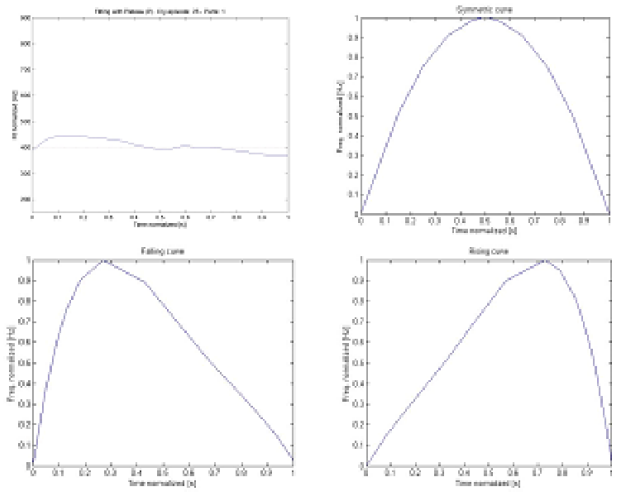
y = a + b * x + c * x. ^ 2 + d * x. ^ 3 + e * x. ^ 4;

With coefficients a= -0.01; b= 2.078; c= -0.396; d= 10E-2; e= -8.1E-3.

Figure 4.11 shows the three test curves and an example of plateau.

Afterwards, the F_0 shape of each cry-episode is compared to the 3 ideal curves according to the following tests.

The first test is made with the plateau curve. This shape is represented by the median of the CU. The CU is defined as plateau if the SD of the F_0 points around the median is lower than 100 Hz. When F_0 varies significantly (SD> 100Hz), the CU undergoes the fitting procedure to the three ideal curves: Symmetric, Rising and Falling, in order to find the one that best approximates the distribution of values of F_0.



4.11 Melody test curves

The curve that best fits the points is determined by comparing the values of the following statistical parameters:

• "SSE" (Sum of Squared Errors) is the sum of the squares of the differences between each data and the average of its group. When all the values of the CU are identical to those representing the ideal curve, SSE is equal to 0. Therefore the lower the SSE the best the fitting curve.

• "RSQUARE" is the coefficient of determination, i.e. a proportion of the variability in the data and the ideal curve. RSQUARE varies between 0 and 1: 0 corresponds to worse fitting and 1 to the best fitting.

• "RMSE" (Root Mean Square Error): is a measure of the difference between the data of the crying episode and those of the ideal curve. The lower the RMSE, the better the fit.

The fitting which corresponds to the smallest value of SSE and RMSE and for which the value of RSQUARE is closer to one is defined as the best fit.

If from all three statistics the same result is obtained, then that curve is regarded as the one that best fits the distribution of points. If only two of the three statistics are relative to the same curve, to confirm whether this curve is actually the one that best fits the data an additional test is applied, based on the position of the maximum value of the signal:

- If the maximum is located in the first part of the interval, namely that up to 35% of the total length of the CU, the best fit is Falling.
- If the maximum is located in the central part of the interval, that is, the one from 35% to 65% of the total length of the episode, the best fit is Symmetric.
- If the maximum is located in the final part of the interval, that is, the one from 65% of the total length of the episode until the end, the best fit is Rising.

If the result obtained by this method is equal to that obtained from the previous two statistical parameters, then it will be considered as the final result of the fitting analysis. When the three statistical parameters return three different curves, the CU is defined as undefined and marked with the symbol "U".

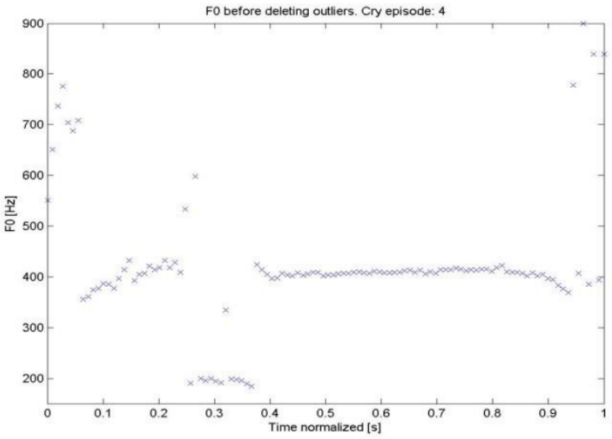
Non-compliant statistical results are obtained when the statistical parameters used to determine the best fitting are far from their optimal values. In this work we have chosen to indicate as non-compliant statistical parameters for which the following conditions are obtained:

• SSE> 10,

• RSQUARE <-1.

The method was applied to the F_0 estimated on some CUs of a LR subject not include in the GR3 database. 33 CUs of 5 different children were analysed with this method. A comparison with perceptual analysis was carried out and results show a correct classification of: 20 on 21 falling CUs classified with perceptual analysis; 3 on 3 rising CUs; 4 on 4 symmetric CUs; 2 on 1 plateau CUs.

These preliminary results point out the need for a robust method for the removal of outliers. This could be a relevant topic of future investigation. An example of real plateau shape is shown in Fig. 4.12.



4.12 An example of real plateau shape

4.4. Cry Classification Methods

As reported in Sect II, several studies show that the newborn cry contains specific features that enable the classification of various diseases and conditions by automatic techniques. In this PhD work the aim of classification was to find cry features capable to differentiate the newborns (that is, the development of their phonatory apparatus) during the first six months of life for the definition of normative ranges, which is one of the main objectives of the GR3 project.

All experiments were conducted using the Weka [249] machine learning software package with its standard settings.

For reasons of clarity the order in which they are described corresponds to that of their use in the present work.

Then, their use through the WEKA tool is described. Results are reported in Chapter 7.

In Fig. 4.13 is reported the procedure used for the classification with WEKA. It is made up by the following steps: dataset, features extraction, features selection, classification and evaluation.

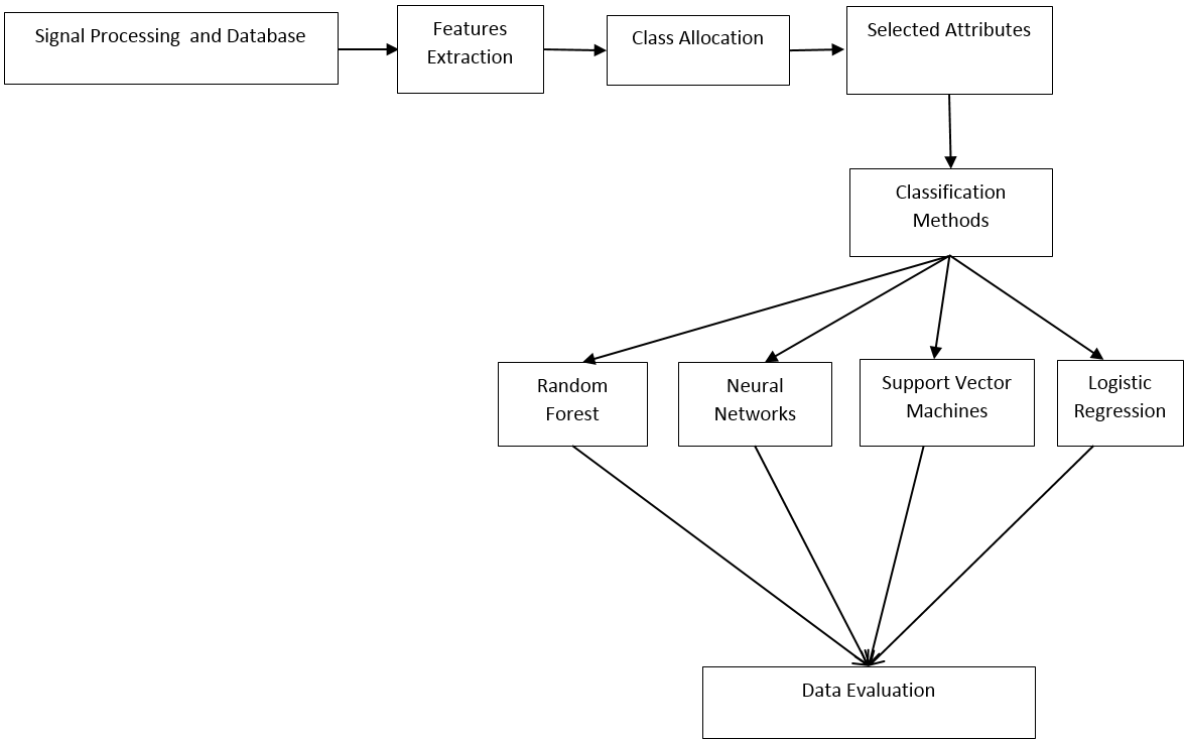


Figure 4.12 – Flow-chart of the data mining process.

4.4.1 Pattern Recognition Methods

Generally, the process of Automatic Cry Recognition is done in two steps. The first step is the acoustic processing, or features extraction, while the second is known as pattern recognition or classification. In some systems, an extra step between them is performed, this is the process called feature selection. In the acoustical analysis, the cry signal is analysed to extract the most important characteristics. The set of obtained characteristics is represented by a vector, which, for the process purposes, represents a pattern. The set of all vectors is then used to train the classifier. Later on, a set of unknown feature vectors is compared with the previous one to measure the classification output efficiency.

Genetic Algorithms

Genetic Algorithms, proposed by John Holland are a family of computational models inspired by biological evolution, which encode a potential solution to a specific problem on a simple chromosome-like data structure and apply recombination operators to the structures to preserve relevant information. Two main problems to be solved when designing a genetic algorithm are the encoding problem and the evaluation function, also called fitness function. Encoding of chromosomes is fundamental to determine the best

recombination and mutation. The evaluation function can be mathematically formulated or obtained through a simulation process. Alternatively the fitness function can be based on the performance of the system under evaluation. In this type of genetic algorithm, a population of possible solutions (chromosomes or individuals) evolves in order to optimize the solution. Each candidate solution has a number of properties that can change or be altered. The best individuals are randomly selected from the population, and each individual genome is modified (recombined and possibly randomly mutated) to generate a new generation. This method will be used for the best attributes selection.

Random Forest

Random forests are built by combining the predictions of several decision trees, each of which is trained in isolation, and then the predictions of the trees are combined through averaging [249]. A Random Forest ensemble uses a large number of individual, un-pruned decision trees. The individual trees are constructed using a simple algorithm which represents a top-down decision tree induction algorithm in which the decision tree is not pruned and at each node the inducer randomly samples *N* of the attributes and chooses the best split from among those variables. The classification of an unlabeled instance is performed using majority vote. There are three main choices to be made when constructing a random tree. These are (1) the method for splitting the leaves, (2) the type of predictor to use in each leaf, and (3) the method for injecting randomness into the trees. Another method for randomization of the decision tree is through histograms. The use of histograms has long been suggested as a way of making the features discrete, while reducing the time to handle very large datasets. Typically, a histogram is created for each feature, and the bin boundaries used as potential split points. The randomization in this process is expressed by selecting the split point randomly in an interval around the best bin boundary. One important advantage of the random forest method is its ability to handle a very large number of input attributes. Another important feature of the random forest is that it is fast.

Neural Networks (Multilayer Perceptron)

Neural network methods construct a model using a network of interconnected units called neurons [Rokach (2010)]. The multilayer feedforward neural network is the most widely studied neural network, because it is suitable for representing functional relationships between a set of input attributes and one or more target attributes. In order to construct a classifier from a neural network inducer, a training step must be employed. The training step calculates the connection weights which optimize a given evaluation function of the training data. Various search methods can be used to train these networks, of which the most widely applied one is back propagation.-Most neural networks are based on a unit called perceptron.

Support Vector Machines (SVM or SMO)

The SVM are a set of supervised learning methods used for classification. For each group of objects divided into two classes a SVM identifies the hyperplane having the maximum margin of separation,

Support Vector Machines are based on the Structural Risk Minimization principle from the computational learning theory. A SVM is a binary classifier that makes its decisions by constructing a linear decision boundary or hyperplane that optimally separates two classes. The SVM can be used to separate classes that could not be separated with a linear classifier; otherwise their application to cases of real interest would not be possible. In these cases, the coordinates of the objects are mapped in an area called "feature space" using non-linear functions, called "feature function". The feature space is a highly multi-dimensional space in which the two classes can be separated with a linear classifier.

Logistic Regression

This is a regression model applied to cases where the dependent variable is of binary type.

Linear regression can easily be used for classification in domains with numeric attributes. Indeed, we can use *any* regression technique, whether linear or nonlinear, for classification. The trick is to perform a regression for each class, setting the output equal to 1 for training instances that belong to the class and 0 for those that do not. The result is a linear expression for the class. The statistical technique called *logistic regression* approximating the 0 and 1 values directly, thereby risking illegitimate probability values when the target is overshot, logistic regression builds a linear model based on a transformed target variable.

Data Evaluation

The evaluation of the performance of classifiers is an important tool in pattern recognition, because it helps to understand the quality of an algorithm and to adjust its parameters. There are several metrics for evaluating the predictive performance of classifiers. Here we briefly describe those used in our experiments.

ROC Curves

One widely used performance measure is the Receiver Operating Characteristic (ROC) curves used to calculate the tradeoff between true positive (TP) to false positive (FP) rates. Precision helps to find how many of the classified cases are correct thus giving a measure of the performance: Precision = TP/(TP+FP). Performance can be measured through the so-called F-measure that is the harmonic mean of precision and sensitivity: F-measure= 2*sensitivity*Precision/(sensitivity+Precision), where Sensitivity = TP/(TP+FN) and FN are the false negative instances. This measure is taken as an alternative measure to the area under the ROC curve.

4.4.2 WEKA

All the experiments were carried out by means of algorithms implemented in WEKA [249]. WEKA is an open source software issued under the GNU General Public License. It is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. WEKA contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

For the classification processing we have used two techniques included in WEKA, which are described below:

Select Attributes

In order to reduce the processing time it is desirable to reduce the size of the input vectors without degrading the efficiency of the classification algorithm. To make a systematic reduction and with the goal of keeping the more relevant features for the pattern recognition process the command Select Attributes is applied. This command allows to evaluate the attributes and selects those with the best performance for the input data considered. We chose as Attribute Evaluator the CfsSubsetEval operator that evaluates the predictive ability of each attribute and the degree of redundancy between them, preferring a set of attributes that are most highly correlated with the class but with low cross-correlation between them. The output returns a measure that guides the search for the most relevant features. The searching method applied by the CfsSubsetEval operator is the GeneticSearch that searches the solution space based on a Genetic Algorithm.

Classify

Classification is a machine learning task that covers any context in which a decision based on available data is made. The learned models from training data are then evaluated using a different test dataset to determine whether the models can be generalized to new cases.

4.4.3 Test and Validation Experiments

In the Experimental results sect. III the classification of infant cry in two classes and in five classes will be presented. The first classification aims at distinguishing between low risk (LR) and high risk (HR) infants with 33 parameters. The second classification concerns children clinically validated as typically developing (TD) and it was carried out with 22 parameters. The aim is to classify these newborns according to 5 classes corresponding to the recording time points: 10 days, 6,12,18 and 24 weeks after birth.

The first action with WEKA was aimed at finding the more relevant features for a proper classification. This is performed running the Select Attributes command with the option CfsSubsetEval as Attribute Evaluator and GeneticSearch as search method. In the Experimental Results will be reported the selection of best parameters obtained with this function.

The classification algorithms used for our experiments are: Random Forest, SMO (Support Vector Machines), Multilayer Perceptron and Logistic (Logistic Regression). For all these methods 10 folds Cross-Validation was used.

The analyses presented in result section were performed on 22 starting attributes in an iterative way (mean, median, SD, min, max of F0, F1, F2, F3, CU length and number of CUs), varying the number of attributes using the SelectAttributes command, which allows to figure out which are the more efficient ones. For testing the following options were used: full training set, percentage split at 66% and 10 folds cross validation. The best results were obtained with all parameters.

5. General Movements Analysis

In the first section of this chapter, a semi-automatic method for GMs analysis was developed. It implements an approach similar to that proposed by Coluccini et al. [149] to extract movements' features such as speed and acceleration. The user manually selects the model to be used: from lower body (LB), based on three points only, to full body (FB) that takes into account 8 points. The system collects x and y coordinates values of each point on the image plane and saves it on a ".csv file" that can be handled with other software such as Matlab or Excel to extract movements features, such as speed and acceleration.
In the second section a new tool for automatic analysis of infant movements is presented based on background subtraction, skin color model and optical flow.

5.1 Semi-automatic method

This tool provides a support to the semi-automatic analysis of movements from video clips. It implements an approach similar to that proposed by Coluccini et al. [149] to extract movements' features such as speed and acceleration.
The purpose of tracking is to extract from the image the 2D position of the body segments to help the study of the movements according to amplitude, average speed and acceleration. Through the list box of the interface the user selects the type of model to be used: from lower body (LB), based on three points only, to full body (FB) that takes into account 8 points. Figure 5.1 and Figure 5.2 show an example with the FB model. Specifically, Figure 5.1 shows the selection of the first point on the body (the right hand), and Figure 5.2 shows the final FB model.

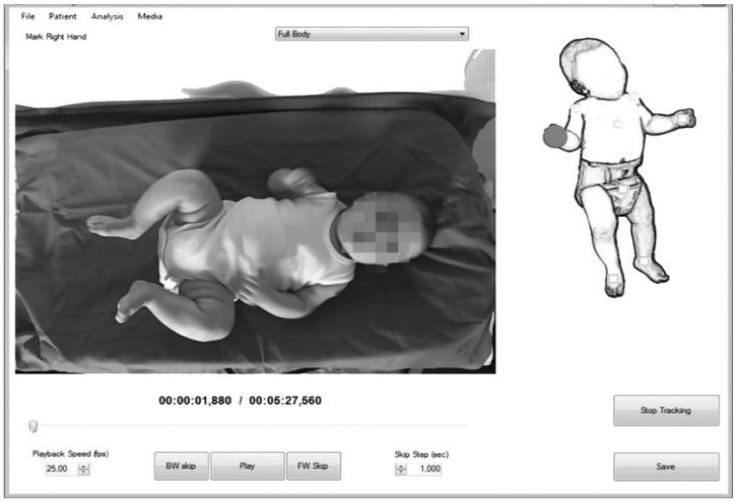


Figure 5.1. Full body (FB) model. Selection of the first point on the right hand / wrist (grey circle on the model).

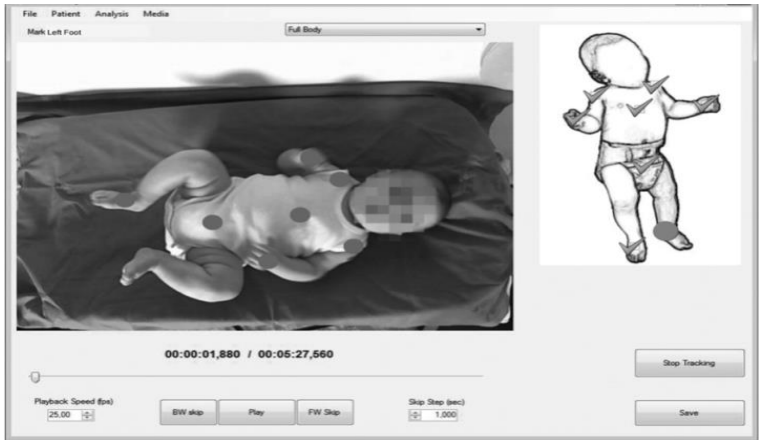


Figure 5.2 Selection of points for the FB model on the first frame. The checkmarks on the model on the right indicate that those points were selected properly.

According to Figure 5.1, the user selects on the video frame the points corresponding to the grey circles that progressively appear on the model (Figure 5.2). After selecting more points, highlighted in grey on the image (Figure 5.2), the user can move to the next frame or to the desired frame using the function "skip step" that allows skipping frames. If the circles do not identify the position of a limb an error message is displayed and the user must repeat the selection of points and apply the model again. FB limbs are listed in Table 5.1.

The interface offers two tracking options: one for the analysis frame by frame and the other every 25 frames. The user can change the frame rate for analysis. The two approaches meet different needs: the frame by frame tracking allows for higher accuracy at the expense of a larger amount of time required for the analysis, while the second option allows a faster but possibly less accurate tracking. This tool could thus support the clinician during the perceptive analysis of GMs by displaying data and trajectories.

Table 5.1. References of the limbs. The user can select from 3 up to 8 points in this list to build a body model.

RH	Right hand/wrist
RS	Right shoulder (shoulder-humeral joint)
LS	Left shoulder (shoulder-humeral joint)
LH	Left hand/wrist
St	Base of the sternum
Pb	Pubis/genitals
RF	Right foot/ ankle
LF	Left foot/ankle

The tool implements a new algorithm for recognition and tracking: in each frame it automatically picks up x and y coordinates and the corresponding time instant of each point on the image plane and saves them in a Comma Separated Values (.csv) file. With this format data can be processed through a simple Excel spread sheet or with commonly used software tools such as Matlab®.

It also automatically provides the time instant of the beginning and end of the sequences of motion (SM) and their number. A sequence of motion ends when all limbs values are lower than a fixed threshold (0.5 pixel/frame).

The tracking procedure returns a matrix of dimension NxM where N is the number of frames and the first column corresponds to time (in ms). In the other M-1 columns, taken two at a time, the x and y coordinates of each point are saved. Thus for each point a N×2 matrix is obtained. From this "position matrix" the speed relative to the origin of the image plane (left upper corner) of the corresponding point is computed as:

$$v_i = \frac{p_i - p_{i-1}}{h} \qquad i = 2 \dots N \qquad (5)$$

where: v_i: speed at the i-th frame, p_i: i-th position in the frame, h: sampling interval and N: number of frames.

From the M-1 arrays containing the x and y coordinates of the speed vectors (size: (N-1) x2) the magnitude of the velocity vector of each point is evaluated. Speeds are normalized with respect to the maximum speed value among all vectors. From v_i values acceleration is computed accordingly.

Useful plots are obtained showing trajectories of the main body segments, along with their speed and acceleration.

To make the trend of a sequence of movements more apparent, a threshold value (0.5 pixel/frame) is chosen given by the average value of each magnitude plot. Those values under the threshold are set equal to 0 while those above the threshold are set equal to 1. To smooth fluctuations speed values are filtered with a moving window of 5 samples. This step is applied to all the velocity vectors thus getting for each of them a vector of binary values indicating the presence or absence of movement of that part of the body. The vectors thus obtained are then reassembled in a matrix Nx(M-1) and represented as an image where black and white areas correspond to the absence or presence of movement, respectively.

Further, the following parameters are computed:

1. Number/start/end of sequences of motion (SM).
2. Average speed of selected points.
3. Average acceleration of selected points.
4. Number of movement units (MU), defined as an act of acceleration or deceleration [224]. A movement unit is recorded when increasing velocity exceeds the threshold value of 0.5 pixel/frame and when decreasing velocity falls below the same threshold value. This parameter tests whether the frequency of movements of individual limbs varies in infants.
5. Skewness (Sk) of the velocity of the feet and hands, that reflects the distribution of movement velocity [153]. In case of a normal distribution the skewness is zero.
6. Kurtosis of acceleration (K) of each limb, to evaluate movement data distribution. The kurtosis quantifies the shape of the probability distribution of acceleration. A distribution with a large K, showing a more sharply peaked and heavy-tailed form, indicates intermittent occurrences of limb movements and deviates from a Gaussian distribution (K=3.0) [250].

Finally, based on the cramped synchronised movements observed by Prechtl et al. [147, 256] the cross-correlation (CC) of acceleration between hands and feet is computed. The aim is to determine whether the movement of selected points proceeds in the same direction at the same time. Thus the cross-correlation of acceleration between the left and right foot, related to the similarity of the spontaneous movements of both

the feet [250], is computed, as well as that between the left and right hand, the left hand and right foot, the right hand and left foot, the right hand and left foot and the left hand and left foot.

5.2 Automatic Analysis of General Movements

Concerning GMs, two approaches were developed. The first one is based on background subtraction, skin colour model and optical flow for velocity estimation. The second one based on background subtraction, region of interest (ROI) and centroid identification. A binary mask was used to identify the infant's silhouette using morphological operators in 5 ROIs: head, right and left arm, left and right leg. For each ROI the centroid was selected and motor parameters were extracted. After locating the baby's head in the upper half-part of the image, the image was divided into four sections corresponding to the four limbs. Parameter's tracking is made searching the centroids in these quadrants. speed and acceleration of each centroid were computed through the first and second derivative of its coordinates and used to identify the indicators of motion (skewness, kurtosis, correlation …).

Video recordings assume a green background to apply a background subtraction for extracting the infant's silhouette; in particular, a colour-space conversion from RGB to YCBCR (luminance, blue chrominance, red chrominance) is performed. Green pixels with a negative red chrominance are easily identified. After this step, a skin colour model is built, modelling the colour distribution in the Cb-Cr plane as a 2D Gaussian function. This step allows filtering out the "non-skin" pixels, thus detecting the limbs and the head of the infant. To compute the velocity of each limb, the image is split into 5 areas with a semi-rigid structure (Fig.5.3) and optical flow is computed. In the second approach, a different structure was applied. After the application of this semi-rigid structure (Fig. 5.4a) and of a series of morphological operation the 5 ROI were identified. The centroids of 4 limbs and the centroid of the head were identified as centroids of these ROI (Fig. 5.4b).

For each area, the following parameters are extracted: Number/start/end of sequences of motion (SM); Average speed; Maximum speed; Average acceleration; Maximum acceleration; Number of movement units (MU) with a threshold value of 0.5 pixel/frame; Skewness (Sk) of the velocity of the feet and hands, Kurtosis of acceleration (K). To validate the algorithm a comparison between marker and marker-less methods is performed. In figure 5.5a a frame of a video recording is shown with 9 markers positioned on the body. The coordinates of the markers are extracted by a template matching method based on cross-correlation and the velocity is computed. The video is analysed with the automatic method explained above. The good agreement between the average velocity of two regions corresponding to the upper limbs right and left (v2, v3 markerless) and that of the corresponding marker-based ones is shown in figure 5.5b. In the Experimental Results are reported only analyses performed with the second approach because with this method was possible to analyse more video recordings. The first semi-rigid structure had a lot of limitation due to the cropping of image on the first frame.

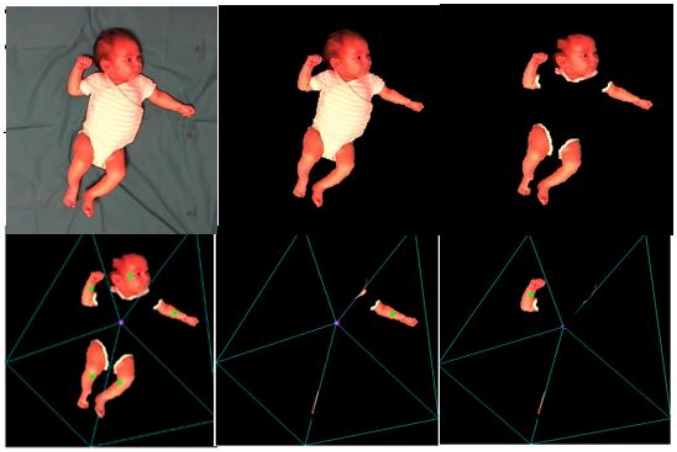


Figure 5.3. Automatic General Movement Analysis: first method. Background subtraction, skin colour model and optical flow for velocity estimation.

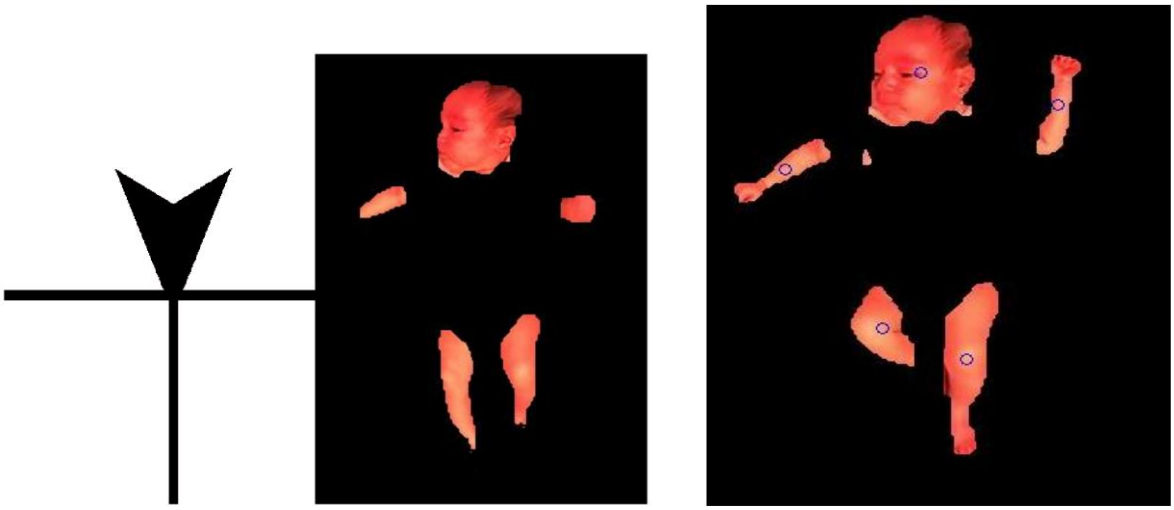


Figure 5.4a and 5.4b. Automatic General Movement Analysis: second method. background subtraction, region of interest (ROI) and centroid identification.

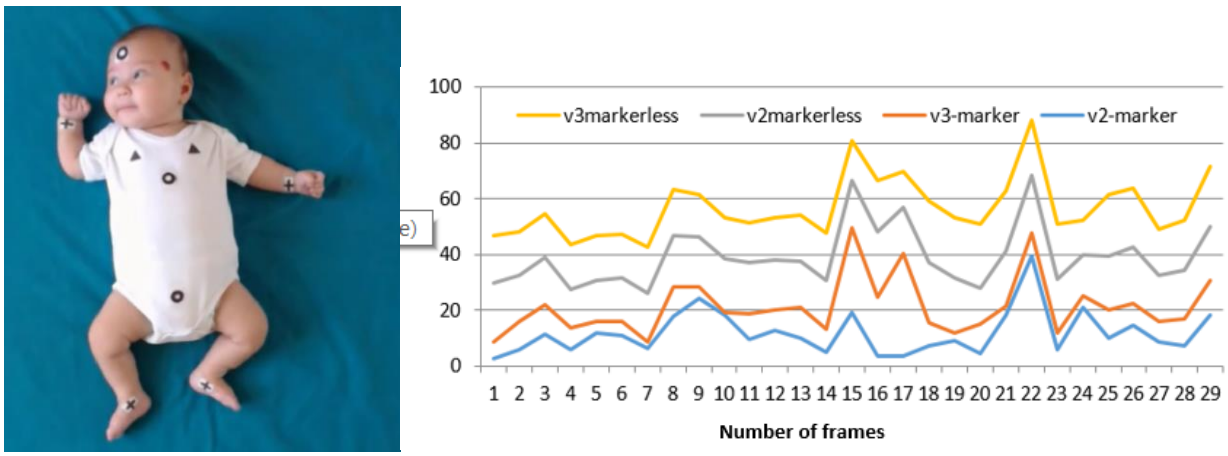


Figure 5.5a and 5.5b. Example of comparison between marker-less and marker-based analysis on 30 frames.

5.3 Parameters Extraction

Indexes of movement of clinical interest

From the velocity measurements obtained from the optical flow and the coordinates of the centroid, all other parameters were obtained for the study of the movement of each region. Firstly the vectors of the trajectories were obtained through integration of the time average of velocity vectors for each region and the acceleration vectors were derived by the velocity vectors. From this information (trajectory, velocity and acceleration) the indices of motion described below were calculated, for both markerless [209] (see Chapter 2) and marker-based analysis [153, 224, 197].

Motion sequence

A motion sequence is defined as a time interval in which at least one of the body segments is considered in motion, that is, it moves at speed greater than 0.5 pixels / frame [197]. This sequence is completed when all the limbs are still (or moving with velocity <0.5pixel / frame) for a duration of at least 0.32sec (8 consecutive frames). The calculation of this index was made according to the perceptual evaluation with which clinicians identify different sequences of GMs in a recording of several minutes. The threshold values for the speed and for the duration of the pause between two sequences were obtained empirically, based on the assessments made by myself, certified observer of the General Movements. The procedure returns the number sequences of motion detected, the instants of beginning and ending for each sequence (in frames and in seconds), their duration and the time break between two consecutive sequences. This allowed the evaluation of all other indices of motion for both the individual sequences (knowing the instants of beginning and end of the sequence), and for the whole movie.

In this study, as in [197], the number of units of motion / minute is considered as an index of motion. The units of motion were calculated for all body segments of interest (head and limbs).

Average speed

For each region was calculated the average value of the speed for the duration of the recording and for each sequence of motion. average acceleration For each region was calculated the time average of the acceleration.

Skewness

The skewness (defined in Chapter 4.3) measures the velocity distribution in the limbs. Spontaneous movements in healthy infants are fluid movements of moderate speed that should be reflected in negative skewness values but not too large in magnitude, while in infants with neurological disorders such movements are spastic and cramping [153, 224]. The skewness of the speed was evaluated not only for the legs as in [153, 224], but also for the arms. Fig. 5.6 shows the distribution of the speed of the right arm for a LR

subject analysed. The long tail to the left reflects the negative skewness of the speed, and therefore fluid movements of moderate speed.

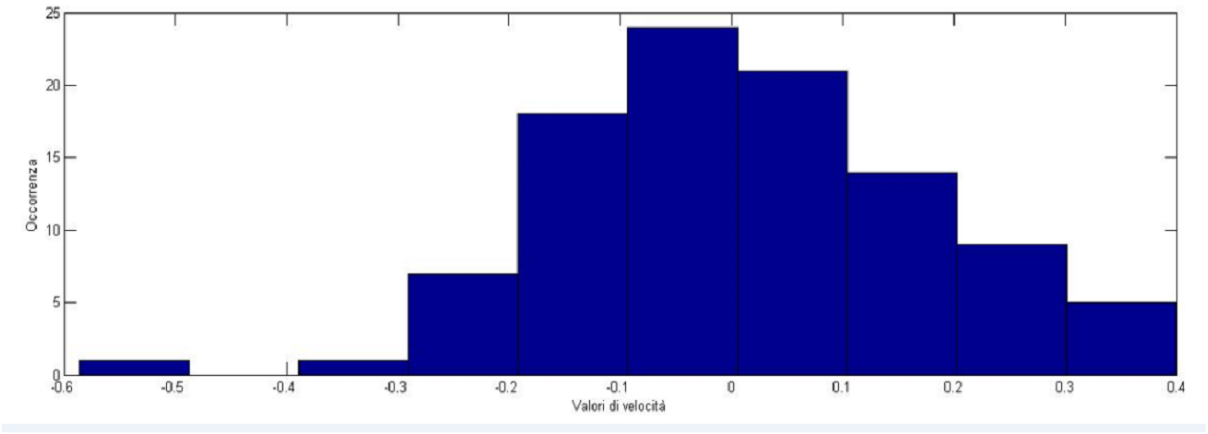


Figure 5.6. Speed distribution for the right arm of a LR newborn at 6 weeks of age.

Kurtosis

As reported by Kanemaru [197, 198], the kurtosis quantifies the shape of the distribution of the acceleration. A high value of kurtosis indicates a distinct peak in the distribution and therefore the presence of intermittent jerky motion [197]. Conversely, data with low values of kurtosis indicate a flatter distribution around the average and a slower decline to the sides. Figure 5.7 shows the distribution of the acceleration of the right arm for the same case study. The sharp peak in correspondence of the average value reflects the high value of kurtosis of the acceleration, which corresponds to a jerky motion.

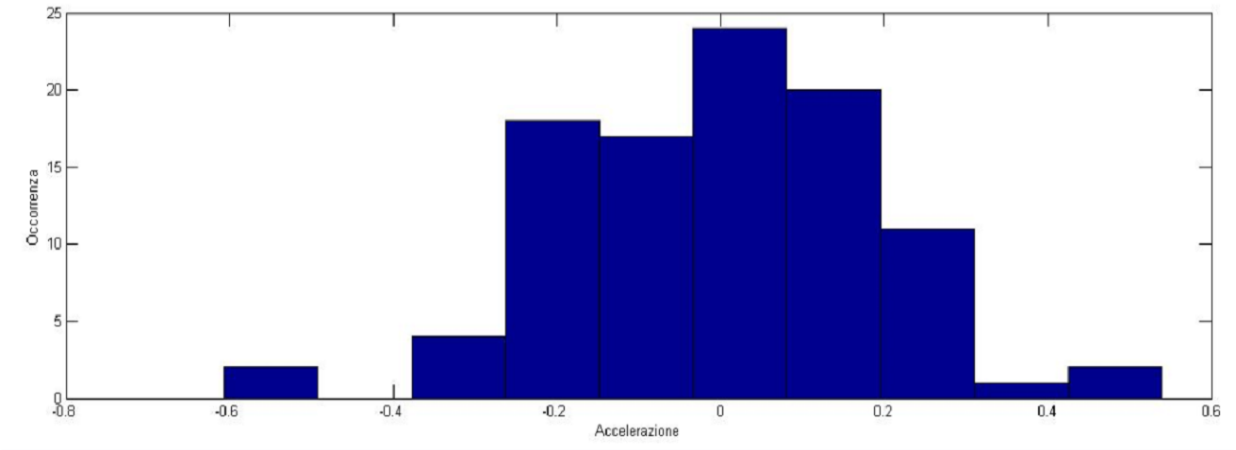


Figure 5.7 distribution of the acceleration of the right arm of a LR newborn at 6 weeks of age

Cross-correlation of speed and acceleration

To investigate the coordination of movements of the limbs and their symmetry, the cross-correlation of speeds and accelerations is calculated for all possible combinations of the limbs (right arm-left arm, right leg-left leg, right arm-right leg, left arm-left leg, right arm-left leg, left arm-right leg) as in [197] and the cross-correlation of speed, as in [197, 224]. The cross-correlation coefficient was calculated as the mean value of the correlation function between the two vectors.

Area outside the standard deviation of the moving average of trajectories [153, 224] and [209]. For the trajectories of the regions of interest, the area outside the standard deviation of the moving average of 25 points of the trajectory was computed as the difference of integrals, in the intervals in which: $m.a - \sigma < T < m.a + \sigma$, where T is the trajectory, m.a is its moving average and σ the standard deviation of the mean. In fig. 5.8 the selection of these areas is represented graphically (green) on a sequence of the trajectory of the right arm of a subject analysed

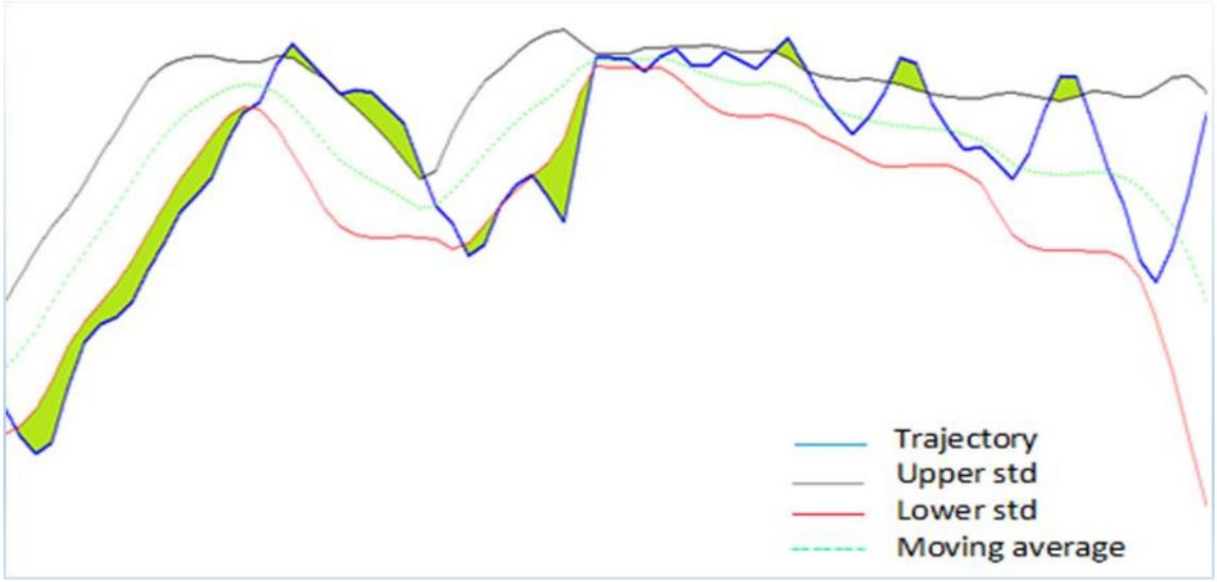


Figure 5.8 Area outside the standard deviation of the moving average of the trajectory of the right arm (colored areas in green) for a LR newborn at 6 weeks of age

Periodicity

To compute the periodicity of trajectories and velocities, a method already applied in other studies both marker-based (Meinecke et al [224]) and markerless (Rahmati et al [209]) is used. The vector of the trajectory (and of speed) was divided into three parts of equal length and each part was averaged. For each part the intersections of the vector with its average were computed and the average distance between two consecutive intersections was measured along with the standard deviation of this distance. Finally the following index of periodicity of the signal was computed:

$$P = \frac{1}{dmean + \sigma}$$

Where dmean is the average of the time intervals (in seconds) between two consecutive intersections and σ is the their standard deviation. In Fig. 5.8 is graphically shown the comparison between the trajectory and the local averages computed for each of the three parts for the right arm of a LR newborn at 6 weeks of age. As shown in Fig. 5.9 the movement is greater in the second segment than in the other two.

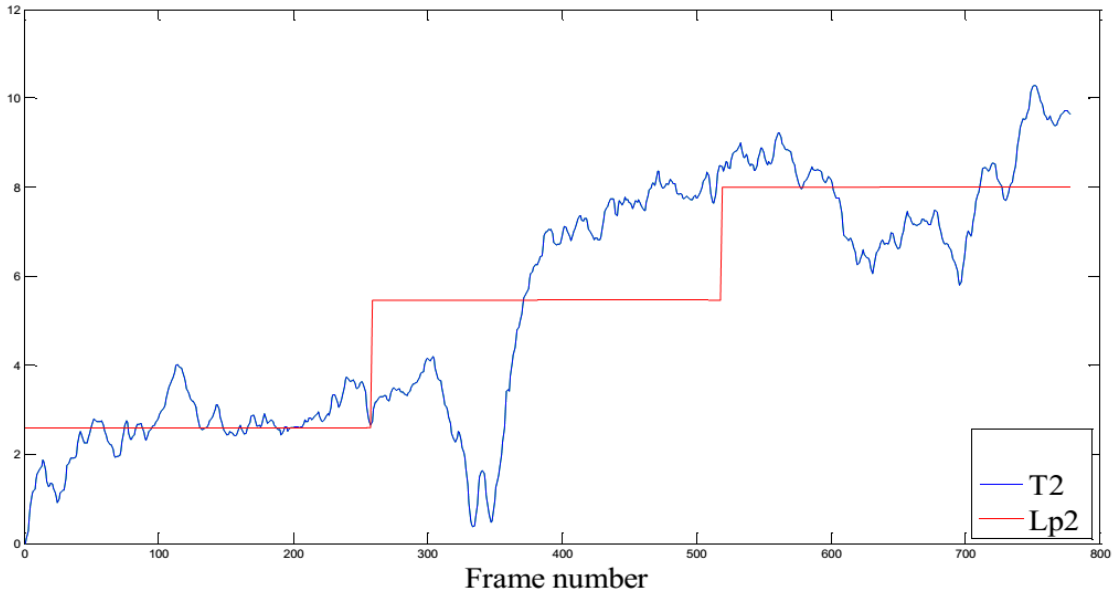


Figure 5.9. Plot of the trajectory and of the local average for the computation of the periodicity. T2 is the trajectory along the x direction of the right arm of a LR newborn at 6 weeks of age. Lp2 depicts the three local averages.

Jerk index

The jerk index, defined in [197] is an index that describes how is jerky the movement. It is defined as the change of the acceleration in time (thus through the derivative of the acceleration). It is computed by the following equation:

$$C = \frac{1}{2} \frac{\overline{(\dddot{x}^2 + \dddot{y}^2)}}{\bar{V}}$$

Where x and y are the (time-varying) coordinates of each limb and V is the average speed for the motion sequence considered. A high value of C corresponds to very jerky movements, while for low values of C the movements are more fluid. The jerk index was calculated only during active periods of motion, that is, for the time intervals during which the speed of at least one of the limbs was above the threshold of 0.5 pixels / frame.

Lateral mobility index

The lateral mobility index E is defined by Kanemaru et al [197] as the ratio between the overall mobility along the horizontal direction and that along the vertical direction. It is computed as the ratio of the root mean square of the speed of the arms and legs along the x direction compared to that along the y direction, according to the following equation:

$$E = \frac{\sqrt{(\dot{x}_{RA}+\dot{x}_{RL}+\dot{x}_{LA}+\dot{x}_{LL})^2}}{\sqrt{(\dot{y}_{RA}+\dot{y}_{RL}+\dot{y}_{LA}+\dot{y}_{LL})^2}}$$

Where $\dot{x}$ is the speed (derivative of the position) for each limb along x and y respectively. RA=right arm, RL=right leg, LL=left leg and LA =left arm . A low value of E indicates reduced lateral mobility.

6. Design of a Tool for Data Acquisition and Analysis

The development of reliable software tools to enhance early diagnosis, especially in home environments, is highly desirable particularly for the neurobehavioral assessment of the newborn. Such tools should provide objective measures to complement clinicians' qualitative analysis that is based on subjective skills. Acoustical analysis of infant's cry and automatic methods for movements analysis may provide objective parameters indicative of neurological pathologies.

The software AVIR described in Chapter 3 was structured to become a module of a larger system named Audio-Video Infant Monitoring (AVIM) whose architecture is shown in figure 6.1

During the PhD period, the analysis of such integrated system was developed reported in a publication recently submitted to a journal.

AVIM is conceived for managing data of a large set of patients. Its most innovative aspect is the ability of merging in a single tool the management of medical records and reports, audio/video data acquisition, handling and analysis, editing and filling out customized tests. Moreover, AVIM supports the clinical perceptual evaluation of cry and spontaneous movements of newborns and infants. The user can manage through a single system all the steps, from the recording to clinical data management and the results of analysis with a significant simplification of the whole procedure.

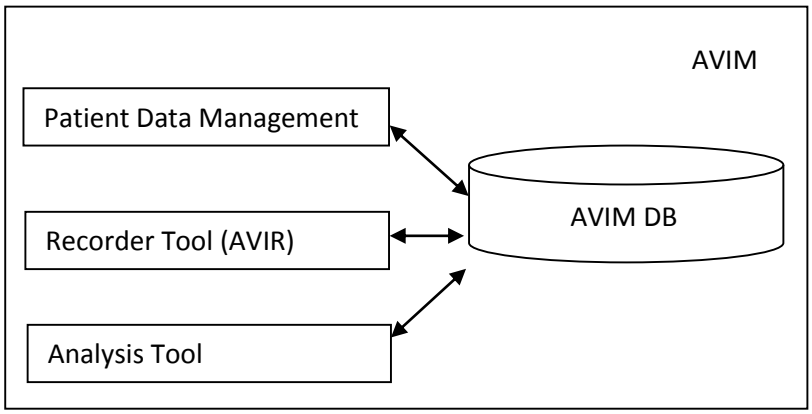


Figure 6.1 AVIM system structure. The system consists of three main sections: "Patient Data Management", "Recorder Tool" and "Analysis Tool".

Besides becoming a computer-aided diagnosis tool, AVIM was developed with the aim of guiding the clinician throughout the process of contact-less assessment and early diagnosis of neurological disorders through a user-friendly environment. Personal and clinical data, markers and notes, as well as acoustical and kinematical parameters of clinical interest are provided and stored with a minimum effort required to the operator. To highlight its capabilities AVIM is applied to the management of data of a newborn recorded five

times from 10 days after birth until the 24th week of age according to the protocol described in Chapter 3. Results are reported in Sect. III devoted to Experimental Results.

The AVIM system is a flexible monitoring system equipped with webcam and microphone that allows setting up specific tests and adding notes during the recording. It is structured so as to eventually make use of modules devoted to the automatic analysis of audio and video recordings obtained according to the protocol described in Chapter 4 and 5.

Thanks to its user-friendly structure and the simple devices used, the system is well suited for performing audio and video recordings both in the hospital and at patient's home. To this aim, AVIM is equipped with two different access levels: a restricted one for nurses or paramedics for data acquisition and an unrestricted one for experienced clinicians to manage audio and video recordings and analysis. It can be arranged for transferring data to a server provided with a centralized database.

The system is developed in C# language using the OpenCv image processing library for video acquisition and recording. The first release of AVIM is designed to run under Windows OS (Windows 7 and Windows 8).

The four AVIM tools are described in the following sections.

Patient Data Management

An integrated system to manage the patient clinical storyboard and share data and medical history through a devoted interface. It allows importing and exporting signals and images from external sources: clinicians can automatically import data from already existing medical records depending on the system used in the hospital. The user can enter patient's data for a specific analysis and then proceed using the tool devoted to audio/video recording. The storage of patient's data is managed through a centralized database structured in order to guarantee privacy and personal data protection.

Moreover this tool allows loading patient's data and patient analysis and tests. The user can eventually view the patient's report and carry out the diagnosis through the storyboard and the medical data.

Recorder Tool

The user interface includes a main window for selecting the recording devices and an acquisition window. The system allows recording from up to two microphones and two webcams at the same time. The acquisition window is equipped with single- or double-window to display video streams. The use of both pieces of equipment and windows may be useful to display both the subject and the surrounding environment (i.e. to assess mother-child interaction, or analyse the subject from two different viewpoints; see Figure 6.2). Dual microphones play a similar task, allowing to record both infant cry and e.g. the mother's voice. This module is based on the AVIR recording tool described in Chapter 3.

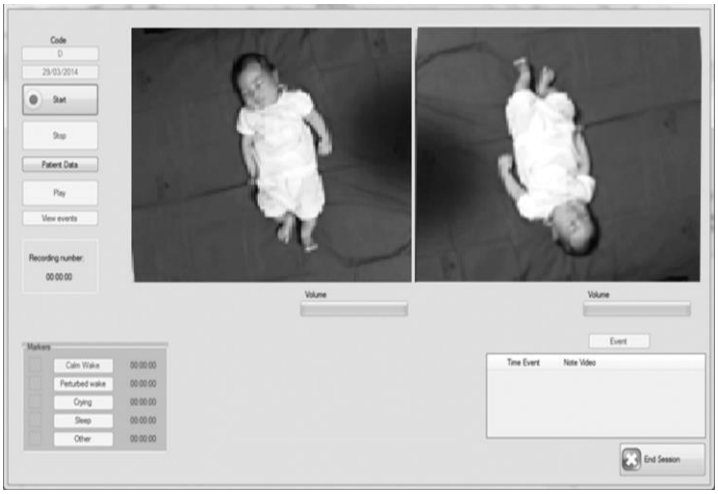


Figure 6.2. AVIM Recorder tool: acquisition window with a double view of the newborn. The user can select markers to point out the status of the patient (such as sleep, restlessness, etc.), ambient noise, interfering events and enter notes during the recording.

Analysis Tool

This tool–allows playing/cutting/copying/assessing sequences of interest using markers placed during the recordings process. The tool also allows entering clinical scores, notes and reference markers (e.g. point out a relevant event) on single crying or motion sequences without the need to resort to the use of other software or to pen and paper. Audio and video editing is managed through specific interfaces. Within each interface the selection and extraction of relevant signal segments is obtained positioning a cursor at their starting and ending points: the software automatically merges into a single file all the selected segments one after the other and saves it pressing the button "Extract". This option proved to be particularly useful in the assessment of GMs from video recordings. Figure 6.3 shows the interface for video recordings and figure 6.4 the corresponding one for audio recordings.

Figure 6.6 shows an example concerning a customized test for GMs set up through button "Test Maker". The clinical tests (fig. 6.6) for the infant GMs assessment were set up by Stella Maris Institute (Pisa, Italy).

Data sheets (active Excel sheets) are provided that allow saving the results of the test as well as personal and clinical data: name, birth, recording time, postmenstrual age, birth weight, recording date, age. All sheets allow the automatic computation of the global score.

The "Start Tracking" button allows the analysis of movements and cry which will be described in a next section.

Figure 6.3. The AVIM Analyser tool: interface for video analysis.

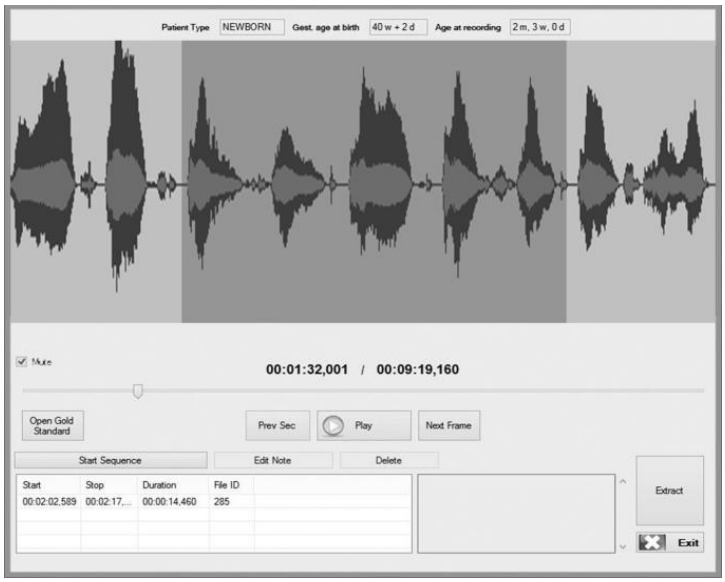


Figure 6.4. The AVIM Analyzer tool: interface for audio (newborn cry) analysis.

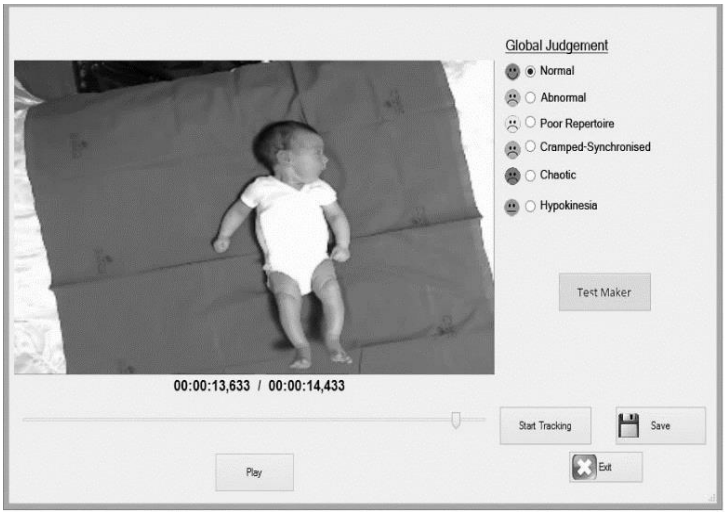


Figure 6.5 The AVIM interface for GMs analysis.

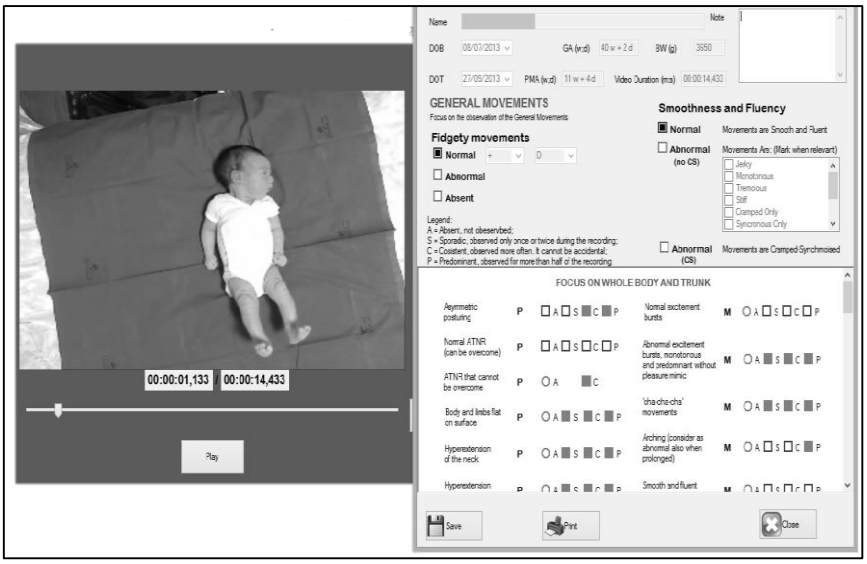


Figure 6.6. Customized test. AVIM allows recalling a specific test for the analysis of GMs

Audio Analysis

AVIM allows adding modules for the analysis of crying (NeCA) to extract the following acoustic parameters (and related statistics): fundamental frequency (F_0), first two resonance frequencies of the vocal tract (F_1 and F_2), curtosis, skewness and time duration of each cry unit (CU). Details about NeCA are reported in Sect.IV. If the analysis module is available, the acoustical analysis can be performed for the 5 recording sessions defined in the protocol.

Video Analysis

This tool provides a support to the semi-automatic analysis of movements from video clips. It implements the method described in Chapter 5 to extract movements' features such as speed and acceleration. The tracking of movements is launched with the "Start Tracking" button (figure 6).

All results obtained by the infant crying and GMs analysis are saved as a patient report. Thus the clinician can easily get the storyboard for the neuro-behavioural assessment of the newborn. Specifically, the clinician gets a report with a clinical evaluation for each session as shown in figure 6.6.

Section III -Experimental Results

7. Results

7.1 Study population

The recruitment of control infants (low-risk infants - LR) was carried out by the staff of the neonatology unit of Santa Chiara Hospital of Pisa, the Lucca Hospital and the Versilia Hospital, Lido di Camaiore, Lucca.
The recruitment of high-risk children (HR) was carried out by the staff of the Bambino Gesù Hospital in Roma both locally and elsewhere in Italy thanks to links on the website with the Italian ASD parents' associations, the Autism and ADHD (Attention-Deficit/Hyperactivity Disorder) group of the Institute of Health in Roma and through a specific brochure that was delivered to pregnant women in the above mentioned hospitals.
The aim of the GR3 project was to set normative ranges for acoustical and motion parameters in a population of about 200 control infants, both male and female (low-risk newborns, LR) and 15-20 "high-risk" (HR) infants, i.e. siblings of children already diagnosed with ASD.
However at the end of the GR3 project (31/12/2014) the total number of recruited children was equal to 72 LR and 21HR. Considering 5 registration steps for each newborn the total number of records would amount to 465. However, for various reasons, for some children not all the 5 recordings (namely at 10 days, 6^{th}, 12^{th}, 18^{th} and 24^{th} week of life) were made. In particular 18 registrations are lacking for LR (concerning 12 infants) and 16 for the HR (concerning 8 infants) at various time points. The total number of records available is therefore equal to 431. Specifically to comply with the protocol, that required all the 5 recordings,8 LR recordings and 3 HR recordings were excluded because the infants did not cry during the recording session. 6 LR and 8 HR were excluded because recording at one or more time point was missing..
Te final number of infants used in this study is therefore equal to 60 LR and 13 HR. The total of available recordings is thus equal to 365.
However to date only 39 LR and HR 10 of these infants were clinically validated because the remaining do not yet have an age greater than or equal to a minimum of 24 months required for the diagnosis of ASD. Therefore, the results that will be shown are related to a total of 250 recordings of these subjects that were listened one by one (average duration of the original recordings: 5 minutes), to verify the feasibility of the analysis.
Sequence of cry
Cry analysis is preceded by a careful selection from the entire audio track of the sequence of interest considered most significant, that is sporadic crying or whining are excluded. These sequences can last from 5 to 90 sec. The selected sequences are of variable duration according to the definition that sequences of crying must be interspersed with 2-10 seconds of break. This definition is consistent with [182] where a pause between sequences lasting more than 5 s was consider.

Specifically the sequences of crying to be analyzed were defined on the basis of experimental tests: after the pre-emphasis step usually characterized by short and repetitive sobs for breath, the beginning of the sequence, made up of a variable number of CUs, is identified. The sequence, of variable duration (5-90 s), usually ends spontaneously as it is followed by a period of silence of the minimum duration of 5 seconds. If the registration does not start with "sobs", the selection of the sequence coincides with the beginning of the audio track. In the event that this sequence is affected by background noise (voices, operator, family, environmental noises, etc.) a next sequence is selected. Finally, if this sequence is not present the less noisy part of the first sequece is selected. This event is indicated in the table containing the notes (as "presence of noise") specifying the type of perceived noise. Should not happen spontaneously, the end of the sequence may have been imposed abruptly by the operator (by pressing the stop button before the phase of silence), or stopped if the cry was "hoarse".

To detect the cause of crying (hunger, nuisance, anger, boredom / recall, pain, moan / sleep, vocalization, etc) I added a marker to be compared with that reported by operators. Finally, information was added about the distance of the microphone from the newborn, noise during recording, modes and quality of audio recording.

Clinical evaluation

As already said the analysis was carried out on a subset of 39 LR infants diagnosed with typical development (TD) and 10 HR at 24 and 36 months of life. At the time of writing this PhD thesis the remaining infants (45 LR and 15 HR) are still waiting for a clinical diagnosis as under the age of two years and therefore these results will not reported. The HR infants include 1 child with diagnosis of Autism Spectrum Disorder (ASD), 1 child with X-fragile syndrome and 2 children with language delay (LD). The clinical/behavioural assessment was performed using a set of questionnaires/interviews: Italian Questionnaire of Temperament, Bayley Scales of Infant Development, the first child vocabulary, MacArthur - Bates Communicative Development Inventory, the Modified Checklist for Autism in Toddlers (*M-CHAT*) and the Autism Diagnostic Observation Schedule (ADOS). Significant differences were found in the fidgety period (p-value 0.1).

The following Tables summarizes the data for the case studies of 39 TD and 10 HR.

Table 7.1 Apgar, gender, gestational age and childbirth for the LR and HR newborns considered (na = not available)

	LR		HR	
	mean	SD	mean	SD
APGAR1	9	0,8	8,9	0,6
APGAR5	9,2	0,5	9,7	0,5
Gender	17Female/22Male		3Female/7Male	
GA	39wks+/-3 days		38 wks+/-3days	
Childbirth	23N/15C/2na		6N/4C	

Table 7.2 Weight, head circumference and length of the LR newborns considered

39 LR	weight		c.head		length	
	mean	std	mean	std	mean	std
at birth	3260,00	410,00	34,20	1,20	50,30	1,80
10 days	3358,10	722,7	34,60	6,10	50,00	8,80
6 wks	4373,20	928,5	36,80	6,10	54,00	9,20
12 wks	5556,60	1158,1	39,00	6,70	58,00	10,20
18 wks	6477,80	1345,4	40,80	6,80	61,70	10,50
24 wks	7303,00	1539,9	41,90	7,20	65,00	11,40

Table 7.3 Weight, head circumference and length of the HR newborns considered

10 HR	Weight		c.head		length	
	mean	std	mean	std	mean	std
at birth	3359	484,00	35	1,60	49	1,90
10 days	3456,60	682,60	35,40	1,70	52,30	2,80
6 wks	4454,40	954,90	38,20	1,40	58,10	6,70
12 wks	5694,40	640,00	40,30	1,40	62,50	5,50
18 wks	6280,00	1190,50	41,70	2,30	63,60	4,80
24 wks	7455,70	759,00	43,40	1,20	67,40	3,50

7.2 BioVoice – Infant Cry Analysis

As explained in chapter 4.1 the acoustic analysis of neonatal cry requires the preliminary selection of cry units (CUs) on which the parameters of clinical interest are then estimated. This is performed with the V/UV procedure described in Sect 4.1(LTAA) whose results are reported here.

7.2.1 Voiced/Unvoiced identification

To test the effectiveness of the proposed approach for V/UV selection, its performance is compared to that of other software tools specifically designed for voice analysis in biomedical applications. Specifically, as in the following three tools are considered: Ampex [252], PRAAT [117] and MDVP [116]. They are compared using synthetic data. Synthetic signals are obtained from an Autoregressive (AR) model obtained by applying system identification techniques to real newborn cries. We investigated the ability to extract voiced intervals of the audio recording also in the case of background noise (e.g. a non-protected environment). This aspect is relevant in clinical applications, the LTAA method, already mentioned in Sect. 2, being developed for use also in non ideal situations, most of which entail ambient noise.

To test LTAA performance against the software tools mentioned above, four different kinds of noise were added to the signal: synthetic white noise (1%), brown noise (1%) and two real noises coming from voice recording in a hospital during outpatient visit, the first one consisting of mixed voices speaking in the

background (called here background noise) and the other due to the fan of an air conditioner (called here fan noise).

To built a synthetic signal, we applied the system identification approach described below to a set of newborn cries of the same type. Specifically, 9 "falling" cries (See chapter 4.3) coming from a healthy full-term newborn (not a GR3 case study), recorded at 18 weeks of life are considered, with a comparable duration (mean about 0.65 s). The derivation of a relevant system description from observed data is termed "system identification" and the resulting system description a "model". As a first step, the procedure of identification requires choosing a model structure to restrict the complexity of modelling.

The model structure which we refer to in this work is the so called autoregressive (AR) model, the general form of which is:

$$y(t) = a1\ y(t-1) + a2y(t-2) + \cdots + any(t-n) + e(t) \qquad (2)$$

where y(t) is the model output at time instant t, ai is the i-th coefficient of the model (i = 1,. . ., n), n is the model order and e(t) is the noise component, assumed to be zero-mean white Gaussian noise. This choice was motivated by the need for a simple model to be used in our experiments. Thus, system identification requires finding the "best" model order n and parameters ai, i = 1,. . ., n for the AR model in Eq. (2). In this work, the relation n ∼= 0.5Fs (in kHz) was considered as explained in Sect.II. As in our data Fs = 44 kHz, the model order is set as n = 22.

For each "falling cry" signal, the corresponding AR(22) model was estimated using the Least Squares method. Then, we defined the following "mean" model A(z) (z^{-1} is the unit delay operator) the coefficients of which were obtained by taking the mean value of the corresponding parameters of the 9 models:

$$A(z) = 1 - 3.45z^{-1} + 5.947z^{-2} - 6.466z^{-3} + 4.286z^{-4} - 0.8448z^{-5} - 1\ .591\ z^{-6} + 1\ .911\ z^{-7} - 0.736z^{-8} - 0.5336z^{-9} + 0.8679z^{-10} - 0.303z^{-11} - 0.3078z^{-12} + 0.3339z^{-13} + 0.2089z^{-14} - 0.7652z^{-15} + 0.9464z^{-16} - 0.7904z^{-17} + 0.5265z^{-18} - 0.2049z^{-19} - 0.03373z^{-20} + 0.1015z^{-21} - 0.05022z^{-22} \qquad (3)$$

Over the 9 cries, the mean fit of the model to estimation data was 99.95%. The AR(22) model in Eq. (3) was then applied to each of the 9 real falling cries to built an almost realistic synthetic signal, made by the sequence of 9 simulated cries of lengths: 0.640 s, 0.540 s, 0.560 s, 0.660 s, 0.960 s, 0.520 s, 0.580 s, 0.800 s and 0.520 s, respectively, separated by convenient time intervals (50 ms, 60 ms, 70 ms, 80 ms, 90 ms, 100 ms, 110 ms, 120 ms and 130 ms, respectively), plus a final one made up by joining the first two synthetic cries (length: 1.180 s). A short silence (15 ms) is added at the beginning and at the end of the cry sequence. Thus, the final cry episode is made of 10 distinct cries. The length of the whole signal was 7.8 s, of which 6.96 s voiced and 0.84 s unvoiced. Fig. 7.1 shows the simulated cry signal (with no noise added) and the corresponding right LTAA V/UV selection. For this signal the following choices were made in LTAA:

- ✓ Sampling frequency: 44.100 kHz.
- ✓ Down-sampling frequency: none.
- ✓ Starting and ending time instant: whole length.
- ✓ Minimum length of the events to be detected: not specified.
- ✓ Size of analysis window: 20 ms.

- ✓ Frequency band: 150–1200 Hz.

To deal with high-pitched signals adequate manual settings were required both in Praat and MVDP before analysis. With LTAA these settings are easily made for any sound through the user-interface. Results are reported in Table 7.4 where the total length (in seconds) of the voiced (V) and unvoiced (UV) intervals is given, as well as the number of voice breaks.

For the signal without added noise, LTAA is capable to properly separate the nine cries and estimate the V and UV intervals of the signal, while all other methods overestimate them. Specifically, Praat slightly overestimates the number of breaks and the length of UV intervals, while both Ampex and MDVP give a heavy overestimation of both data. In the case of white, brown and fan noise all methods overestimate both breaks and UV length, but with a limited increase for the UV length for LTAA and Praat. Finally, only LTAA properly finds the number of breaks in the case of background noise, although it slightly overestimates the UV length that is comparable to the one found with Praat which, however, overestimates the number of breaks. In general, Ampex gives very high values for the UV intervals as well as MDVP that also finds a very high number of breaks. Thus a preprocessing step with LTAA could enhance the results with all the three methods. We notice here that, in order to make LTAA comparable with the other approaches, no threshold was imposed on the minimum length of the events to be detected. Results can thus be enhanced if this option is selected. An example is reported in Fig. 7.2, concerning a real signal of about 9.5 s containing 8 high energy "episodes", some of which are not cries. Specifically, the first part of the second cry episode (1.30–1.35 s) and the third one (2.10–2.35 s) are breathing and whimpering signals and thus should be removed before performing cry analysis. In this example, we removed frames shorter than 260 ms. Fig. 7.2 (upper plot) shows the LTAA V/UV selection with this option turned off, while in the lower plot it is turned on and the frames shorter than 260 ms are properly removed. The performance of the methods is also tested on the real signals shown in Fig. 4.2 (chapter 4). Table 7.5 refers to the newborn cry repeated two times (total length = 19.63 s). The length of voiced intervals is 15.2 s and the number of UV intervals (excluding the first and the last one) is 15. On both signals the same kinds of noise as for synthetic signals were added.

All methods underestimate the lengths of the voiced intervals (mean ranging from 8.5 for Ampex to 10.9 for Praat) and overestimate the number of UV intervals (mean ranging from 21.7 with LTAA to 52.7 with MDVP). With some tools the number of cry episodes, which is a clinically relevant feature, is heavily misestimated. Overall LTAA has the best performance. We remind here that LTAA gives the option to select the minimum length for "voiced" episodes that prevents signal splitting (as shown in Fig. 7.2). This option, that would improve results, was not applied here to make the tools comparable. The results are consistent with those obtained on simulated signals.

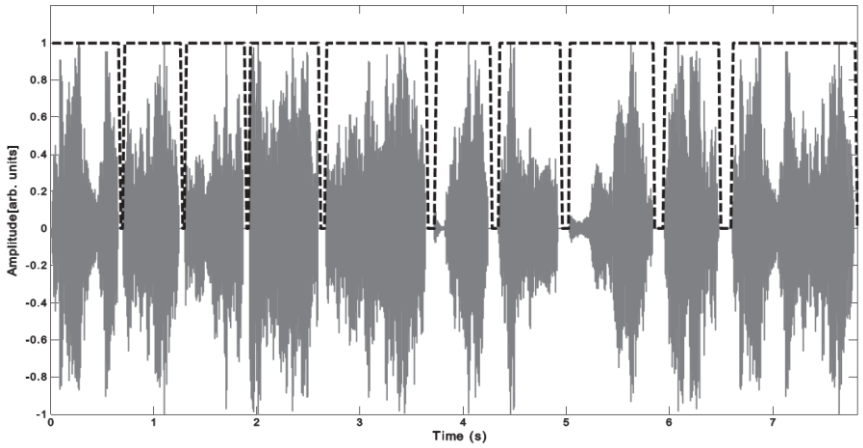


Figure 7.1 Synthetic cry signal obtained by linking together 10 single synthetic "falling" cries. Synthetic signal (grey) an LTAA V/UV selection (dashed black).

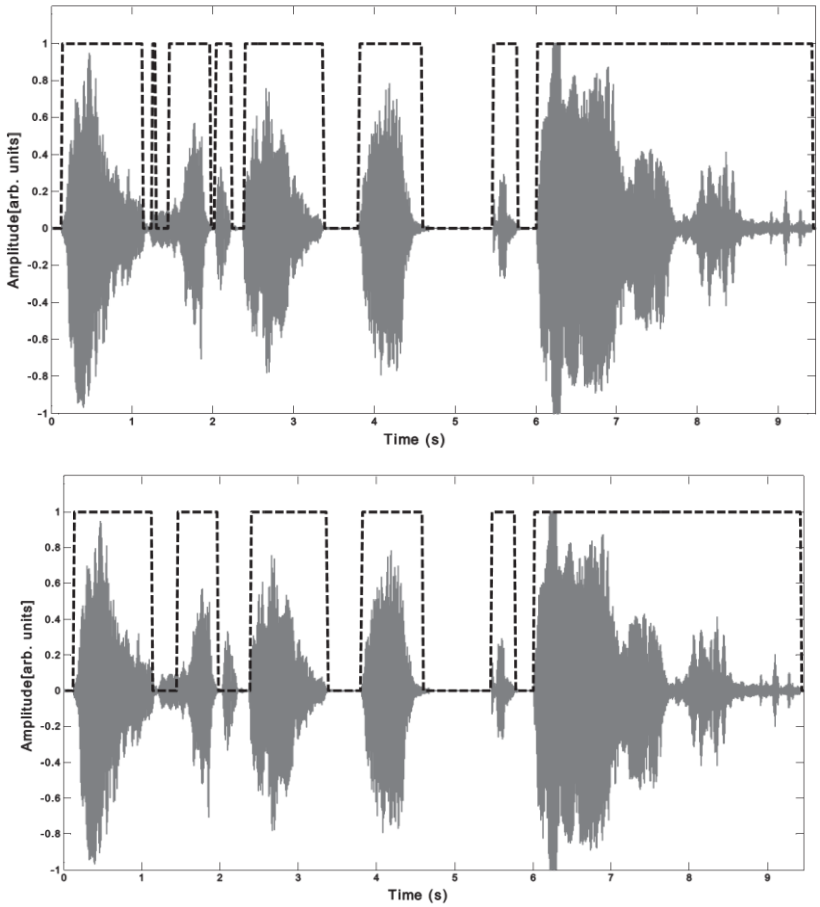


Figure. 7.2. Newborn infant cry. Upper plot: 8 voiced parts found with LTAA without any constraint on the length of the voiced parts. Lower plot: frames <260 ms are removed (1.30–1.35 s 2.15–2.35 s)

Table 7.4
Comparison of V/UV selection for a synthetic signal (newborn infant cry).

	LTAA			Praat			Ampex			MDVP		
	UV [s]	V [s]	Break n.	UV [s]	V [s]	Break n.	UV [s]	V [s]	Break n.	UV [s]	V [s]	Break n.
No noise	0.64	7.16	9	1.03	6.77	16	3.75	4.05	nd	2.6	5.2	48
White noise (1%)	1.38	6.42	11	1.47	6.33	21	6.87	0.93	nd	1.46	6.34	16
Brown noise (1%)	2.12	5.68	17	1.22	6.58	16	7.4	0.4	nd	2.48	5.32	42
Background noise	1.02	6.78	9	1.03	6.77	15	3.88	3.92	nd	2.44	5.36	39
Fan noise	1.9	5.9	16	1.09	6.71	17	6.96	0.84	nd	2.39	5.41	41

Table 7.5
Comparison of V/UV selection for a real signal (newborn infant cry, lower plot Fig. 4.2).

	LTAA			Praat			Ampex			MDVP		
	UV [s]	V [s]	Breaks	UV [s]	V [s]	Breaks	UV [s]	V [s]	Breaks	UV [s]	V [s]	Breaks
No noise	5.19	14.44	15	6.32	13.31	38	6.5	13.13	nd	7.71	11.92	66
White noise (1%)	9.99	9.64	19	9.74	9.89	26	11.4	8.23	nd	10.62	9.01	29
White noise (3%)	12.81	6.82	26	12.86	6.77	19	16.66	2.97	nd	14.84	4.79	24
Brown noise (1%)	10.55	9.08	21	8	11.63	42	11.77	7.86	nd	11.13	8.5	62
Brown noise (3%)	13.39	6.24	29	10.88	8.75	45	15.27	4.36	nd	14.86	4.77	69
Background noise	7.93	11.7	23	6.18	13.45	44	6.14	13.49	nd	7.73	11.9	62
Fan noise	10.15	9.48	19	7.11	12.52	40	10.43	9.2	nd	10.54	9.09	57

7.2.2 LR and HR comparison

Statistical comparison

In the following section the comparison between LR (then classified TD) and HR is reported. The results were obtained with BioVoice. The analysis carried out concerns the averages of the parameters computed on all the CUs of the first sequences of crying.

The results of 39 LR and 10 HR are presented. For the same infants the GMs assessment was carried out by clinicians. For these children it was possible to carry out a comparison as the clinical evaluation was available.

On each recording was computed the mean value of: number of CUs, the CU length (mean, standard deviation (SD), minimum and maximum) F_0, F_1, F_2 and F_3 (mean, median, std, minimum and maximum), the signal duration, the vocalic percentage, the voiced length, the number and length of voice breaks (pauses: mean, std, minimum and maximum) for a total of 33 parameters.

These results are shown in the following Tables.

Table 7.6

Results of 10 HR. Mean values of Fundamental frequency and Resonances frequencies.

10 HR	F0-medianM	F0 -mean M	F0 - std M	F0 – min	F0 - max
10 days	388,38	385,31	62,24	246,65	534,20
6 wks	397,47	397,57	73,00	247,91	591,16
12 wks	387,18	394,01	74,27	243,57	595,52
18 wks	427,74	424,04	75,81	244,79	617,03
24 wks	437,75	434,45	90,13	256,96	624,68

10 HR	F1-mean M	F1 - median	F1 - std	F1- min	F1-max
10 days	1138,02	1105,78	293,20	652,80	1931,37
6 wks	1124,07	1111,90	237,10	730,77	1752,07
12 wks	971,97	965,33	179,81	630,94	1475,36
18 wks	1028,27	1029,14	213,03	613,47	1602,18
24 wks	1007,33	991,73	226,49	631,98	1732,26

10 HR	F2 -mean M	F2 - median	F2 - std	F2- min	F2-max
10 days	2979,93	2822,31	722,91	2019,66	5216,30
6 wks	2907,73	2745,03	639,11	2119,57	4989,95
12 wks	2794,34	2646,63	622,29	2012,57	4913,54
18 wks	2801,54	2642,00	678,97	1930,87	4976,32
24 wks	2839,40	2666,41	724,40	1945,27	4955,23

10 HR	F3-mean M	F3 - median	F3 - std	F3- min	F3-max
10 days	5176,01	5184,39	778,89	3712,61	6868,77
6 wks	5286,18	5216,30	621,54	4075,33	6929,20
12 wks	5028,85	4981,98	513,54	4031,19	6374,41
18 wks	4890,12	4782,00	693,17	3779,68	6799,47
24 wks	4888,39	4777,39	669,62	3747,67	6805,26

10 HR	N cry ep.	Cry duration - mean	Cry duration - std	Cry duration - min	Cry duration - max
10 days	25,00	0,86	0,42	0,33	1,89
6 wks	21,91	0,79	0,42	0,26	1,79
12 wks	24,00	0,91	0,50	0,27	2,06
18 wks	22,36	0,94	0,50	0,32	2,07
24 wks	20,27	0,93	0,60	0,25	2,40

10 HR	N pauses	Pause duration - mean	Pause duration - std	Pause duration - min	Pause duration - max
10 days	24,09	1,09	1,10	0,14	4,75
6 wks	21,00	1,24	1,05	0,13	3,55
12 wks	23,09	0,60	0,46	0,07	1,93
18 wks	21,45	2,52	2,37	0,44	5,93
24 wks	19,36	0,84	0,95	0,07	3,26

Table 7.7
Results of 10 HR. Mean values of signal and voiced duration.

10 HR	signal duration	% voiced	voiced duration
10 days	46,29	42,14	20,82
6 wks	41,65	42,66	18,71
12 wks	39,74	53,60	23,58
18 wks	43,03	48,36	21,91
24 wks	38,72	47,34	20,00

Table 7.8
Results of 39 LR. Mean values of Fundamental frequency and Resonances frequencies

39 LR	F0-medianM	F0 -mean M	F0 - std M	F0 - min	F0 - max
10 days	423,9862	433,1808	113,2344	220,9375	703,7416
6 wks	427,4019	442,3586	119,0169	234,242	734,6517
12 wks	446,6962	462,4015	122,3152	242,0288	768,9389
18 wks	457,0932	474,3091	130,0464	240,8653	783,4449
24 wks	468,5792	480,3438	122,1223	248,3689	755,1639

39 LR	F1-mean M	F1 - median	F1 - std	F1- min	F1-max
10 days	1130,673	1118,808	270,6925	569,8228	1970,542
6 wks	1088,614	1076,165	272,9884	570,3435	1950,162
12 wks	1018,955	1009,184	238,8441	550,8224	1811,098
18 wks	988,0355	964,0772	260,1644	522,1184	1914,772
24 wks	979,8459	968,1575	248,8416	518,771	1834,555

39 LR	F2 -mean M	F2 - median	F2 - std	F2- min	F2-max
10 days	3362,168	3229,879	933,9822	1874,827	5795,641
6 wks	3265,028	3045,286	1024,833	1825,966	5936,525
12 wks	3238,964	3061,395	997,6816	1742,566	5821,191
18 wks	3182,481	2948,269	1035,771	1683,028	5831,517
24 wks	3125,603	2902,087	1040,631	1669,28	5852,111

39 LR	F3-mean M	F3 - median	F3 - std	F3- min	F3-max
10 days	5636,957	5593,108	939,9792	3911,854	7695,852
6 wks	5621,655	5562,828	915,8676	3851,109	7677,357
12 wks	5525,337	5422,197	817,2542	3993,449	7502,937
18 wks	5426,195	5325,549	826,5104	3888,702	7551,406
24 wks	5371,045	5273,494	808,6971	3830,749	7426,439

39 LR	N cry ep.	Cry duration - mean	Cry duration - std	Cry duration - min	Cry duration - max
10 days	22,575	0,868208	0,447879	0,3375	1,8975
6 wks	24,6	0,900703	0,464146	0,2975	1,972
12 wks	23,425	0,973987	0,521081	0,301	2,1015
18 wks	20,9	1,113317	0,613499	0,313	2,462

24 wks	22,05	1,074346	0,684809	0,316	2,542

39 LR	N pauses	Pause duration - mean	Pause duration - std	Pause duration - min	Pause duration - max
10 days	21,6	0,702857	0,515773	0,2005	1,9345
6 wks	23,625	0,625466	0,536783	0,09	2,011
12 wks	22,45	0,55755	0,460348	0,1015	1,8825
18 wks	19,925	0,518282	0,4674	0,0745	1,8195
24 wks	21,075	0,611971	0,491772	0,151	1,8825

Table 7.8
Results of 39 LR. Mean values of signal and voiced duration.

39 LR	signal duration	% voiced	voiced duration
10 days	30,95747	60,25302	18,679
6 wks	35,64808	60,40281	22,3465
12 wks	35,02633	62,90377	22,507
18 wks	34,55233	65,29558	23,2305
24 wks	34,85727	62,88825	23,0045

Significant differences were found for F_0 F_2 and F_1 mean values with the Mann-Whitney test (p-value < 0.05).

These results show that the age of 6-12 weeks is somehow "critical" in the sense that at this age the two groups tend to get closer after an initial and subsequent difference. This may be an interesting result to be investigated clinically although at present not too much reliable given the limited number of cases. However, it seems that the "first cry" (ie a few days after birth) has characteristics that differentiate healthy individuals from those at risk. Therefore, the implementation of a screening in neonatal clinics (fast, easy and cheap to perform) might be useful.

For HR subjects cry analysis has shown a different trend for F_0 as compared to the control group especially in the last two step points (18^{th} and 24^{th} week).

From the group of HR children diagnosed with ASD or developmental disorders such as those diagnosed with language disorders (LD) were identified and compared separately with the LR group subsequently validated as TD on analogy to [156].

Fig. 7.7 shows the comparison of F_0 between the TD group and 1 ASD case. There is a clear difference of the values of F_0 at 18 weeks of life. It may also be observed a slight difference at 10 days, 6 and 24 weeks. Fig. 7.8 shows the comparison of the average values of F_0 between 39 TD and 2children suffering from LD. Also in this case there is a greater difference at 18 weeks and a lower one at 12 weeks. The cases, however, are too few to draw further conclusions.

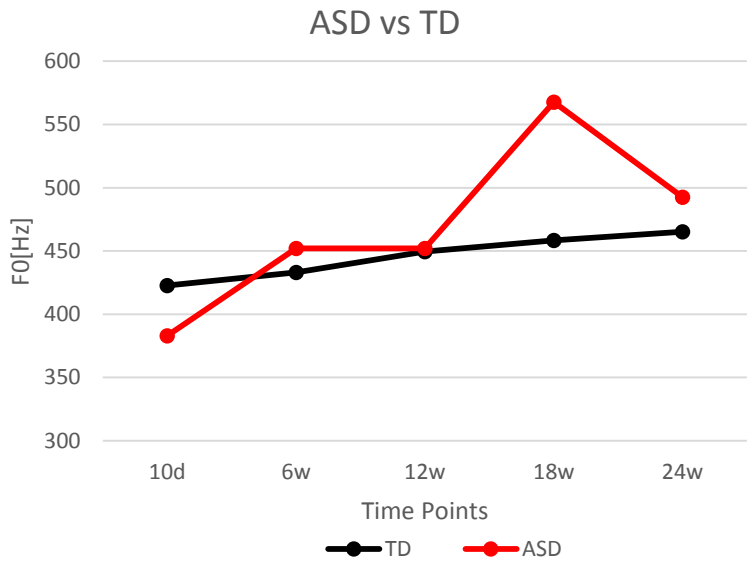


Figure 7.7 Comparison of F_0 between the TD group and 1 ASD case.

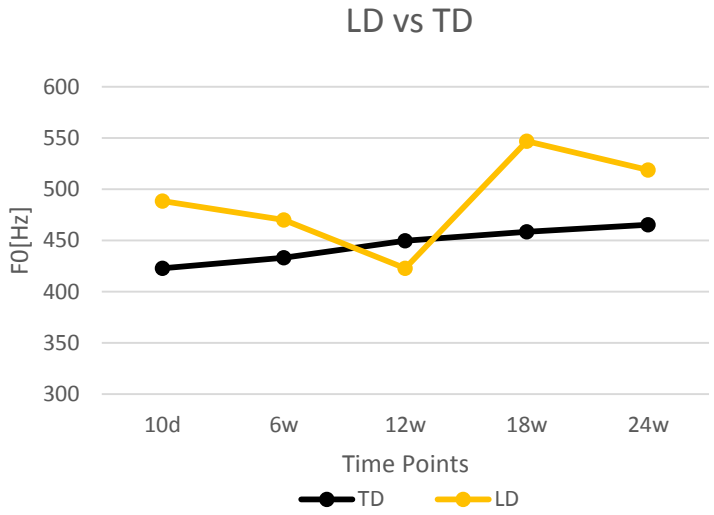


Figure 7.8 Comparison of the average values of F_0 between 39 TD and 2children suffering from LD

7.2.3 WEKA classification

To compare the two groups of control (TD) and HR subjects, a classification with WEKA software was performed applying the methods described in section II: Logistic curve, Multilayer Perceptron, Support Vector Machine and Random Forest. The best results was obtained with the Random Forest method and a 10-folds cross validation. The same result was obtained also with split option.

The mean values of the parameters were computed and the Excel spreadsheet provided by BioVoice was converted to a .csv file to be analysed with WEKA. As already said the 33 parameters provided by Biovoice are:

- CU length (mean, standard deviation, minimum and maximum)

- F0, F1, F2 and F3 (mean, median, std, minimum and maximum)
- Signal duration
- Vocalic percentage
- Voiced length
- CUs Number
- Length of voice breaks (pauses: mean, std, minimum and maximum)

Therefore, the worksheet is organised in:
-34 Columns: 33 parameters and one label to identify the class of (LR or HR)
-51 Lines: the first line contains the names of the attributes and the label of the class, the other 50 contain the analysis results obtained with Biovoice.
The analysis was carried out despite knowing that WEKA does not offer the best performance because of the limited database especially for the different number of cases between HR and LR. Nevertheless, the result was satisfactory for Random Forest method. Results, presented in Table 7.9, show the percentages of the correct instances classified in the comparison between TD and HR. The percentage is always above '80% ranging between 84,09% at 6 and 24 weeks to 86,36% in the other time points. This means that the classifier can effectively distinguish LR from HR using only the average values of the estimated parameters. The other classification methods had performances lower than 70%.

Table 7.9 – Results of classification with Weka

	10 days	6 weeks	12 weeks	18 weeks	24 weeks
39 TD vs 10 HR	86,36%	84,09%	86,36%	86,36%	84,09%

It can be seen that there is a greater difference between the time points at 10 days, 12 and 18 weeks. In particular the difference between 10 days and 18 weeks is in agreement with the result obtained for F0 through the statistical analysis.

7.3 NeCA – Infant Cry Analysis

Based on the results obtained with synthetic signals, the CWT Mexican Hat was applied for the estimation of F_0 and the CWT Complex Morlet for the estimation of F_1-F_3 in real signals. Results concern spontaneous hunger cry of 39 newborns (22 male and 17 female, gestational age at birth 39,4 ± 1,4 weeks, mean weight 3260 g ± 406 g) whose typical development (TD) was clinically assessed within the GR3project. From each recording step about 900 CUs (around 1 s of length each) were extracted with BioVoice (4551 CUs in total) and the analysis performed with NeCA. Table 7.10 summarizes the results. Notice the slightly but uniform increase of F_0 from the first time point towards the last one. RFs do not show any specific trend, though F_1 and F_2 values are higher in the first step than in the last one.

Table 7.10 – Mean and SD values for F_0, F_1, F_2 and F_3 obtained with NeCA. Details about CUs (obtained with BioVoice) are also reported

	F0 [Hz]		**F1 [Hz]**		**F2 [Hz]**		**F3 [Hz]**		**signal length [s]**		**% voiced**		**CU length[s]**		**CU tot**	**CU number**	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD		Mean	STD
10 days	457,8	92,0	1237,9	131,7	2594,4	216,2	5027,4	487,8	31,8	14,7	61,8	17,7	0,9	0,3	903	23,2	13,0
6 wks	463,5	95,8	1217,5	147,9	2491,5	203,7	5285,4	509,0	36,6	12,9	62,0	15,4	0,9	0,5	984	25,2	11,0
12 wks	474,7	90,0	1169,1	143,5	2494,0	218,8	5336,9	428,6	35,9	13,1	64,5	13,2	1,0	0,2	937	24,0	11,0
18 wks	484,9	96,2	1198,6	156,0	2518,9	231,1	5277,1	442,8	36,0	11,8	67,7	10,1	1,2	0,3	845	21,7	8,0
24 wks	490,2	87,7	1200,9	154,4	2498,6	215,1	5186,7	445,5	35,8	15,1	64,5	12,5	1,1	0,3	882	22,6	12,0

To assess whether the means are statistically different a two-tailored t-test was applied to the F_0 values. Results reported in Table 7.11 show that when the time distance between the recording sessions is "long enough" (that is more than 6 weeks) there is a statistically significant difference between the mean F_0 values. Specifically the difference is significant (0.05 confidence) between 10 days and 18 weeks and between 6 and 18 weeks, and is highly significant (0.01 confidence) for a longer time distance (10 days – 24 weeks and 6 – 24 weeks). Almost no significant difference was found for F1, while for F2 there is a (decreasing) significant difference only between the first recording step (10 days) and the other 4.

Table 7.11 – T-test results for F_0 at the five time points. The statistical significance of the differences increases with the time distance between the time points.

	10d-6w	*10d-12w*	*10d-18w*	*10d-24w*	*6w-12w*	*6w-18w*	*6w-24w*	*12w-18w*	*12w-24w*	*18w-24w*
t-test	0,53	0,09	0,01*	0,005**	0,26	0,02*	0,002**	0,30	0,14	0,55

*significant ** highly significant

7.3.1 WEKA classification based on NeCA analysis

To compare the five classes of TD subjects a classification with WEKA software was performed applying the Random Forest method and a 10-folds cross validation.

The mean values of the parameters were computed and the Excel spreadsheet provided by NeCA was converted to a .csv file to be analysed with WEKA. The 22 parameters provided by NeCA for each CU are:

- CU length
- F0, F1, F2 and F3 (mean, median, std, minimum and maximum)
- Number of F0 points used for F0 estimation

Therefore, the worksheet is organised in:

-23 Columns: 22 parameters and one label to identify the class of (LR or HR)

-51 Lines: the first line contains the names of the attributes and the label of the class, the other 50 contain the analysis results obtained with NeCA.

Significant results are shown in Table 6. The comparison between 10 days and 18 weeks and 10 days and 24 weeks are significant time points with higher percentages of correct classification above of 70 %. The other comparisons are not significant. These results could indicate the possibility to describe infant development trajectories for the acoustic parameters that it is not in the literature.

Table 7.12 – Significant time points for the description of infant development trajectories identified in the first 6 month of life.

WEKA – Random Forest NeCA	**10 days vs 18 wks**	**10 days vs 24 wks**
39 TD Trajectories Corrected instances classified	75%	76,32%

WEKA has also been applied to the classification of the 39 TD subjects in the 5 time step using the Random Forest algorithm and 10-fold cross-validation or split option. With this method the following 22 parameters were selected: length of the CU, mean, median, standard deviation, maximum and minimum of F_0, F_1, F_2 and F_3, number of points of F_0. The method has provided satisfactory results only for the steps: 10 days -18 weeks (75%) and 10 days- 24 weeks (76%).

7.4 Melody Results

The new F_0 shape descriptors: skewness, kurtosis, and percentiles might be useful for a statistical characterization of the melody of crying (F_0 shape in time).

T-test was applied to the F_0 skewness, kurtosis, and percentiles compared two time points at a time. The minimum number of CUs was 492 found at 24 weeks. Results are reported in Table 7.13, 7.14, 7.15.

Table 7.13 – Results of shape descriptor (**< 0.01 ; 0.05<*<0.01)

T-test	*10d-6w*	*10d-12w*	*10d-18w*	*10d-24w*	*6w-12w*	*6w-18w*	*6w-24w*	*12w-18w*	*12w-24w*
Skewness	**	**	0,145	*	*	0,083	**	0,582	**
Kurtosis	*	0,503	*	*	*	0,884	*	*	0,503
10th Percentile	0,584	*	**	0,584	0,081	**	0,584	*	*
20th Percentile	0,901	**	**	0,901	**	**	0,901	0,12	**
30th Percentile	0,922	**	**	0,922	**	**	0,922	0,064	**
40th Percentile	0,199	**	**	0,199	**	**	0,199	0,293	**
50th Percentile	0,416	**	**	0,416	**	**	0,416	0,453	**
60th Percentile	0,787	**	**	0,787	**	**	0,787	0,959	**
70th Percentile	0,552	**	**	0,552	**	**	0,552	0,329	**
80th Percentile	0,672	*	0,003	0,672	**	**	0,672	0,701	*
90th Percentile	0,368	**	**	0,368	*	**	0,368	0,23	**
100th Percentile	0,959	0,019	0,002	0,959	*	**	0,959	0,417	*
25th Percentile	0,803	**	**	0,803	**	**	0,803	*	**
50th Percentile	0,347	**	**	0,347	**	**	0,347	0,232	**
75th Percentile	0,545	**	**	0,545	**	**	0,545	0,831	**

Table 7.14 – Results of CU number with skewness and kurtosis positive or negative

Number of Cus with	***10d***	***6wks***	***12wks***	***18wks***	***24wks***
Skewness>0	286	284	284	308	296
Skewness<=0	206	208	208	184	196
Kurtosis>0	35	15	39	30	20
Kurtosis<=0	457	477	453	462	472
TOT CUS	492	492	492	492	492
% Skewness>0	58	58	58	63	60
% Skewness<=0	42	42	42	37	40
% Kurtosis>0	7	3	8	6	4
% Kurtosis<=0	93	97	92	94	96

As regards the values of skewness and kurtosis in the 5 time points, about 60% of positive and 40% of negative skewness values were found, that correspond to falling and rising melody of F0, respectively, with a strong prevalence (94%) of negative kurtosis values that indicate the peakedness of energy distribution. We observe, however, that the detection of symmetrical curves (skewness = 0) was not taken into account for the lack of references in the literature for the definition of suitable thresholds around 0. An interpretation of this result is thus somewhat complex and will be the subject of future studies.

Table 7.15 – Results of mean parameters of CUs calculated on 492 CUs for each time points.

Average	***10d***	***6wks***	***12wks***	***18wks***	***24wks***
Skewness>0	0,44	0,27	0,41	0,33	0,27
Skewness<0	-0,16	-0,15	-1,19	-0,19	-0,12
Kurtosis>0	4,65	3,54	2,73	1,32	2,87
Kurtosis<0	-1,16	-1,16	-1,14	-1,16	-1,20
10th Percentile	478,07	482,86	498,68	519,15	509,37
20th Percentile	457,71	458,70	483,92	496,19	506,77
30th Percentile	447,19	446,48	483,05	497,56	506,31
40th Percentile	450,56	440,97	483,41	491,29	493,45
50th Percentile	447,89	441,96	484,18	489,82	487,60
60th Percentile	446,57	444,51	484,63	484,22	487,58
70th Percentile	452,87	448,39	477,92	485,64	492,46
80th Percentile	458,16	454,82	479,21	482,26	490,71
90th Percentile	451,46	459,42	480,40	490,72	487,08
100th Percentile	462,49	462,05	483,39	491,05	492,50
25th Percentile	464,11	465,85	490,16	506,15	507,25
50th Percentile	448,55	442,73	483,70	491,27	493,63
75th Percentile	452,23	448,26	481,69	482,91	491,28
F0 mean	455,30	453,98	483,98	492,83	495,32
F0 median	445,82	441,45	469,69	470,63	486,09
F0 SD	123,58	121,38	129,10	135,98	117,46
F0 min	228,31	238,02	248,22	251,60	269,23
F0 max	740,24	755,65	799,97	814,11	766,14
F1 mean	1157,02	1138,91	1061,97	1037,70	1004,15
F1 median	1148,87	1138,18	1047,10	1011,82	995,04
F1 SD	255,12	285,47	250,11	267,85	247,45
F1 min	616,52	589,99	594,49	552,01	547,94
F1 max	1950,99	2054,94	1899,29	1944,20	1883,96
F2 mean	3453,61	3366,13	3305,10	3252,03	3187,36
F2 median	3323,51	3165,62	3137,28	3003,02	2976,77
F2 SD	935,47	1065,11	996,48	1055,76	1023,21
F2 min	1948,11	1849,86	1801,35	1748,64	1748,26
F2 max	5856,99	6097,45	5915,44	6015,42	5923,52
F3 mean	5740,16	5674,21	5611,72	5537,91	5461,30
F3 median	5662,58	5601,27	5513,94	5448,16	5389,56
F3 SD	966,55	939,39	828,43	839,81	793,65
F3 min	4034,43	3897,15	4046,82	479,55	3918,19
F3 max	7735,56	7622,61	7583,37	7718,44	7852,22

7.5 General Movement Analysis

7.5.1 Perceptual Analysis

Results of the GMs tests for writhing and fidgety period are shown in figure 7.9. The GMs assessment was carried out by clinicians on the same 39 LR and 10 HR. These results confirm that the period from the 6th to the 24th week is the best to discriminate the group LR from that HR.

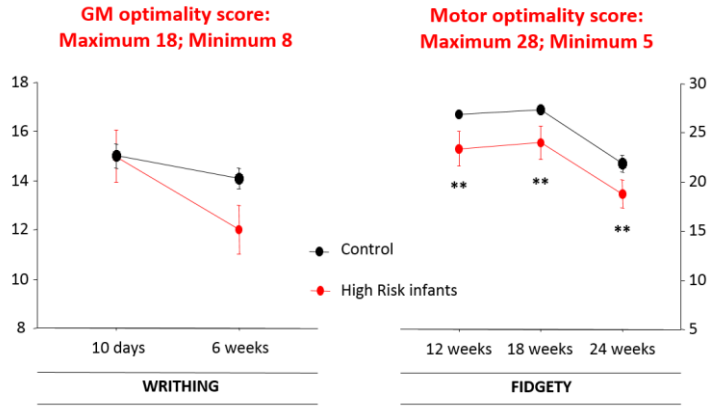


Figure 7.9 GMs evaluation

7.5.2 Automatic Analysis of GMs

The automatic method was applied on 10 TD infants for each time points for a total of 50 recordings of about 3-5 minutes of duration. The new parameters have to be validated by clinicians to set a new scale for GMs evaluation.

For the 10 newborns the clinical results concerning global scores are reported in Table 7.16 that also shows the duration of the 5 recordings in the corresponding time steps. As mentioned in the introduction writhing movements are present until the age of 8 weeks of life, while fidgety movements are commonly observed later until 20 weeks of age. Thus the Prechtl's Method on General Movement Assessment (PGM) and Preterm and Term General Movements (PTGM) tests were assessed until the 6th week and the Assessment of Motor Repertoire – 3 to 5 Months (MR) test from the 12th week on. These values assess the healthy state of the three newborns.

Table 7.16. Results of Global Scores of GMs analysis for the newborn under study at each recording session (step point). The duration of the recordings is represented here as minutes.seconds

	1st infant				**2nd infant**				**3rd infant**			
Time point	**Duration**	**PGM**	**PTGM**	**MR**	**Duration**	**PGM**	**PTGM**	**MR**	**Duration**	**PGM**	**PTGM**	**MR**
10 days	**5.11**	**18**	**40**	**-**	**6.14**	**17**	**41**	**-**	**2.10**	**10**	**22**	**-**
6 weeks	**2.28**	**18**	**41**	**-**	**3.16**	**17**	**41**	**-**	**5.05**	**13**	**30**	**-**
12 weeks	**5.27**	**-**	**-**	**28**	**5.00**	**-**	**-**	**28**	**5.00**	**-**	**-**	**24**
18 weeks	5.10	-	-	28	5.00	-	-	28	5.00	-	-	26
24 weeks	5.00	-	-	17	5.00	-	-	17	5.00	-	-	28

Table 7.17 summarizes the results concerning the movement analysis for the 10 newborns where only 4 limbs are considered: hands and feet (RH, LH, RF and LF). The table shows the number of sequences of motion (SM) and the number of movements for each limb (MU) normalized in time The values of speed for the 4 limbs is computed as the number of movements per minute. This gives a measure of the frequency of the movement of each limb. According to clinicians this definitely provides a very useful information to the clinician, both for subsequent analysis of the same movie and for a comparative analysis made by several specialists.

Furthermore for each recording, the system provides results concerning skewness and kurtosis between left and right hands and feet. Some results are reported in Table 7.18.

Table 7.18. Results of GMs analysis for the three newborns under study. Number of sequences of motion (SM) and of movements (MU) for each recording step and for the selected frames.

	1st infant					**2nd infant**					**3rd infant**				
Time point	**N. of SM**	**RH MU/min**	**LH MU/min**	**RF MU/min**	**LF MU/min**	**N. of SM**	**RH MU/min**	**LH MU/min**	**RF MU/min**	**LF MU/min**	**N. of SM**	**RH MU/min**	**LH MU/min**	**RF MU/min**	**LF MU/min**
10 days	7	100	106	48	60	3	150	120	60	90	1	15	15	15	15
6 weeks	2	14	194	13	34	2	60	120	120	120	3	50	100	100	100
12 weeks	3	28	35	29	20	5	30	60	75	30	2	3	18	24	9
18 weeks	28	45	30	65	45	3	12	12	12	12	3	60	72	36	24
24 weeks	11	45	22	38	35	7	90	90	60	60	3	210	135	45	150

Table 7.18. Results of speed and acceleration. In the table skewness (Sk) and kurtosis (K) are reported for each recording step and for the selected frames.

	1st infant								**2nd infant**								**3rd infant**							
Time point	**Sk RH**	**Sk LH**	**Sk RF**	**Sk LF**	**K RH**	**K LH**	**K RF**	**K LF**	**Sk RH**	**Sk LH**	**Sk RF**	**Sk LF**	**K RH**	**K LH**	**K RF**	**K LF**	**Sk RH**	**Sk LH**	**Sk RF**	**Sk LF**	**K RH**	**K LH**	**K RF**	**K LF**
10 days	-2,69	-1,55	-3,94	-3,90	9,59	10,76	16,11	14,45	-2,76	-3,25	-1,18	-3,27	7,25	8,98	7,32	9,67	-2,76	-3,25	-1,18	-3,27	7,25	8,98	7,32	9,67
6 weeks	-2,29	-0,05	-1,75	-1,27	18,37	8,53	16,66	13,24	-0,48	1,29	-1,29	-0,95	3,21	5,36	6,03	5,12	-0,48	1,29	-1,29	-0,95	3,21	5,36	6,03	5,12
12 weeks	-4,03	-3,06	-1,79	-2,51	20,82	18,96	12,48	17,51	-1,31	-2,61	-2,17	-0,60	6,91	16,62	9,33	3,28	-0,57	0,20	-0,80	-0,72	5,95	9,11	8,22	7,53
18 weeks	-4,00	-4,39	-2,92	-3,95	15,43	19,11	15,10	17,60	0,32	0,72	0,39	0,77	3,10	2,90	3,29	4,60	-3,63	-3,38	-3,46	-5,25	17,15	10,70	21,26	28,98
24 weeks	-4,23	-4,48	-3,42	-4,22	19,03	27,00	19,45	22,72	-0,52	-3,78	-3,39	-1,49	4,69	14,01	11,57	6,59	-1,77	-2,34	-3,18	-2,46	4,05	4,62	11,53	5,19

All results obtained through the infant crying and GMs analysis are saved as a patient report. Thus the clinician can easily get the storyboard for the neuro-behavioural assessment of the newborn. Specifically, the clinician gets a report with a clinical evaluation for each session.

It was not possible to apply the automatic analysis to all the videos because in some of them the recording protocol was not strictly complied. In fact sometimes the green sheet used as background was stirred making it impossible color filtering for the background subtraction. In other videos the presence of the parent or of the microphone provided for the voice analysis has been found for the entire duration of the video recording. Some of these cases are shown in Figure 7.10:

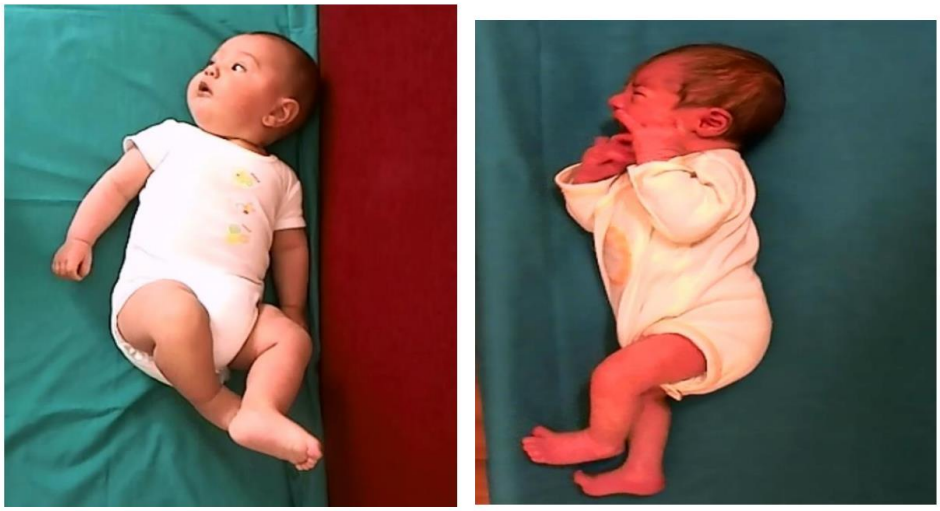


Figure 7.10. Examples of children excluded from the analysis as they are not properly positioned according to the protocol.

In the first case it was not possible to select a region of interest with the whole body of the child on a green background, in the second case the child is not in the supine position as required by the protocol.

7.6 AVIM application

In this section we present results concerning the application of AVIM to data collected from a newborn enrolled in the project Grant (GR3) "Non-invasive tools for early detection of Autism Spectrum Disorders". The newborn is a full-term female belonging to the control group of the GR3 project. The newborn data are reported in Table 7.19. where GA is the gestational age, W is the weight, L is the length and HC is the head circumference. Both the webcam and the microphone are managed through the AVIM Recorder tool. Recordings are made according to the protocol described in Chapter 3.

Questionnaires were compiled manually and saved as images in AVIM in a single folder for each patient. Patient data were entered and saved through the Patient Data Management tool. The questionnaires used to assess the GMs were built through the Clinical Test Maker tool and saved in .csv format.

The video analysis requires a careful selection of motion sequences: AVIM can simplify this step as it automatically provides the time instant of the beginning and end of the sequences of motion and their number.

Also, cry analysis is preceded by a careful selection from the entire audio track of the sequence of interest considered most significant, that is sporadic crying or whining are excluded. Acoustic parameters of clinical interest are obtained using both BioVoice and NeCA tools described in Chapter 4.2.
According to the protocol referred above the subsequent steps of the AVIM application are presented.

Table 7.19. Newborn Data. Female full-term infant belonging to the control group of the GR3 project. GA is the gestational age, W is the weight, L is the length and HC is the head circumference.

	GA(week+days)	W [g]	L [cm]	HC [cm]
At birth	37+5	2804	48	32
At 10 days	40+1	3060	50	34,5
At 6th week	43+5	4460	58	37
At 12th week	49+5	4800	60	38,5
At 18th week	55+5	5500	63	40,5
At 24th week	61+5	7760	65,5	42,3

Patient Data Management

Through the "Patient Data Management" module (figure 6.1), the user selects the type of patient and fills up a form for clinical and Patient data. Figure 7.11 shows the AVIM form for clinical data for the newborn under study at birth and figure 7.12 shows newborn's GMs data at the third recording session (12th week) This form is displayed after loading the newborn's code or surname.

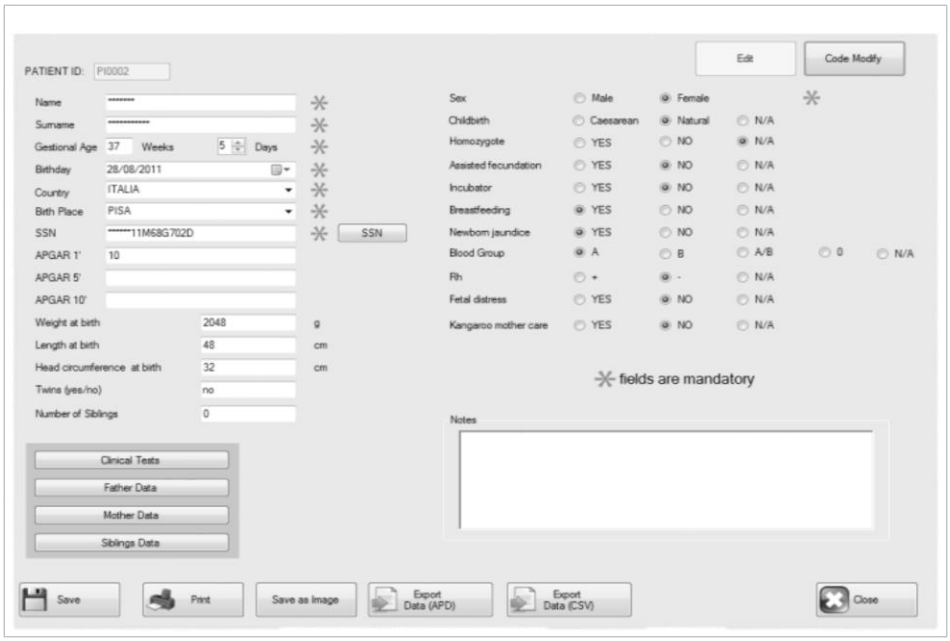


Figure 7.11 AVIM form for clinical data for the newborn under study at birth. Through the "Patient Data Management" button the user selects the type of patient (newborn) and fills up a form for clinical and personal data.

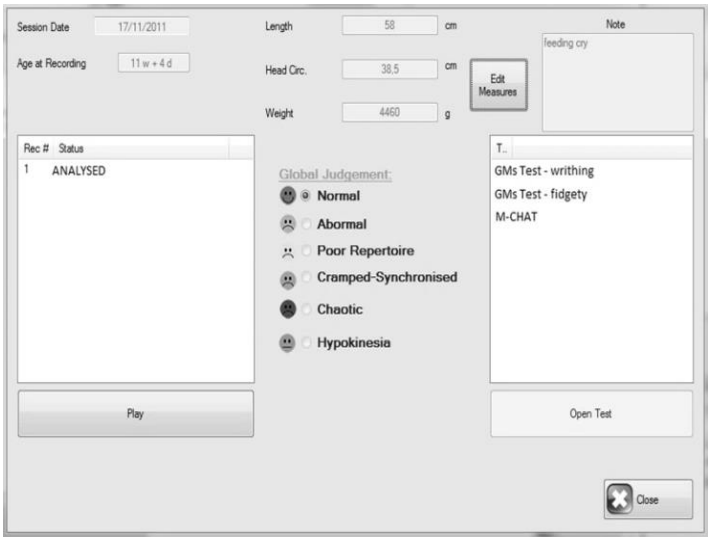


Figure 7.12 Newborn data at 12th week. The reported tests (GMs Test- writhing, GMs Test –fidgety and M-CHART)

Recorder Tool

Through the "Recorder Tool" (figure 6.1) the user sets the devices. In this example, one webcam and one microphone were used. On analogy to figure 6.2, the user starts the recording according to the protocol: the webcam is placed at one meter fixed distance from the newborn perpendicularly to the child that is placed in the supine position on a green sheet. The microphone is placed at 25 cm from the baby's head.

Analysis Tool

Through the "Analysis Tool" (figure 6.1) the user performs audio and video analysis. On analogy to figure 6.3 and figure 6.4, the user selects the audio and video frames of interest.

Audio Analysis

The protocol required spontaneous cry (therefore excluding pain-related stimulation) lasting a minimum of 2 minutes. From this recording all the parts containing interfering factors such as other voices and ambient noise were manually deleted. The selection of the crying frames of interest is made by pushing the "Extract" button (figure 6.4)The remaining useful parts have a duration between 15 s and 32 s, still more than enough for a reliable acoustical analysis.

Some results for the subject considered are reported in Table 7.20, Table 7.21 and Table 7.22 concerning mean values of F_0, F_1, F_2, CU length and pauses, respectively. The analysis is made with BioVoice.

Table 7.20 Mean values of mean, median, standard deviation (std), minimum (min) and maximum (max) of F_0, F_1 and F_2 over all CUs of each recording session (step point) for the newborn under study.

Time Point	F0 [Hz]					F1 [Hz]					F2 [Hz]				
	Median	*mean*	*std*	*min*	*max*	*median*	*mean*	*std*	*min*	*max*	*median*	*mean*	*std*	*min*	*max*
10 days	443,84	454,15	83,74	252,30	733,20	1240,48	1314,50	342,64	541,97	2445,47	3153,19	3766,88	1452,24	2113,19	7377,48
6 weeks	436,07	447,11	69,66	353,83	688,56	1362,86	1426,61	294,61	936,35	2272,20	3511,54	3969,61	1607,16	2161,16	7164,02
12 weeks	474,73	515,12	148,02	250,11	857,71	1229,40	1224,75	212,70	736,30	1854,85	3160,77	3429,20	923,60	2230,31	6379,53
18 weeks	560,74	611,33	140,78	352,12	834,27	782,13	782,06	141,96	501,48	1407,46	3555,78	3618,24	1292,43	1524,34	6377,72
24 weeks	542,05	563,13	154,32	281,17	833,47	912,71	889,19	241,12	490,15	1791,26	3494,96	3578,93	853,77	2156,30	5902,19

Table 7.21. Signal and voiced parts length, percentage of voiced parts, number of CUs and mean values of mean, standard deviation (std), minimum (min) and maximum (max) length of CUs for the newborn under study at each recording session (step point).

Time point	signal length [s]	voiced length [s]	% Voiced	CU number	mean CU length [s]	std CU length [s]	min CU length [s]	max CU length [s]
10 days	24,60	7,84	31,87	6,00	1,31	0,60	0,28	2,06
6 weeks	30,27	7,68	25,37	8,00	0,96	0,77	0,28	2,42
12 weeks	32,09	16,32	50,86	14,00	1,17	0,64	0,38	2,14
18 weeks	15,75	9,48	60,21	9,00	1,05	0,54	0,30	2,08
24 weeks	25,47	14,76	57,95	14,00	1,05	0,47	0,44	1,70

Table 7.22. Number of pauses (P) and mean values of mean, standard deviation (std), minimum (min) and maximum (max) length of pauses for the newborn under study at each recording session (step point).

Time point	P Number	mean P length [s]	std P length [s]	min P length [s]	max P length [s]
10 days	5,00	2,46	0,73	1,34	3,04
6 weeks	7,00	3,14	2,91	0,14	7,82
12 weeks	13,00	1,15	1,02	0,18	3,48
18 weeks	8,00	0,34	0,21	0,12	0,76
24 weeks	13,00	0,51	0,44	0,10	1,36

In Table 7.20, F_0 mean values show an increasing trend from step 1 to step 4, followed by a slight decrease in the last step. The first formant, F_1 shows an opposite trend, with mean values decreasing from the first to the last step while F_2 remains more or less stable.

Table 7.21 shows the percentages of the vocalic parts vocalic (cries), that of CUs and their relative length, for each recording. For the infant under study these values show about 50% of vocalic parts and CUs of duration from 15s up. Table 7.22 completes this information with the average number and duration of pauses (parts of the signal in which the baby is not crying): while the number of pauses increases slightly during the 5 steps, the average length decreases.

Video Analysis

Through the "Analysis Tool" (figure 6.1), the user selects the video recording and the video frames of interest. On these frames the clinician performs the perceptual analysis by re-loading the GMs forms, previously built according to figure 6.5 and figure 6.6. Once filled up, the forms are saved in .csv format.
For this newborn, clinical results concerning global scores are reported in Table 7.23 that shows the duration of the 5 recordings. As mentioned in the introduction writhing movements appear until the age of 8 weeks of life, while fidgety movements are commonly observed until 20 weeks of age, thus the Prechtl's Method on General Movement Assessment (PGM) and Preterm and Term General Movements (PTGM) tests were assessed until the 6 week and Assessment of Motor Repertoire – 3 to 5 Months (MR) test from the 12th week on.

Table 7.23 Results of Global Scores of GMs analysis for the newborn under study at each recording session (step point).

Step point	N. Recordings	Length [mm.ss]	PGM	PTGM	MR
10 days	1	05.11	18	40	-
6 weeks	1	03.28	18	41	-
12 weeks	1	05.27	-	-	28
18 weeks	1	05.10	-	-	28
24 weeks	1	07.23	-	-	17

Through the "Start Tracking" button (figure 6) the user launches the semi-automatic GMs analysis. The frame rate was set at 25 frames/s and the movement analysis was carried with a four-points model consisting of: right-hand (RH), left hand (LH), right foot (RF) and left foot (LF) (figure 5.2, table 5.1). The movement parameters were sampled every 10 frames (400 ms).
The tracking procedure returns a matrix of dimension Nx9 where N is the number of frames and the first column corresponds to time (in ms). In the other 8 columns, taken two at a time, the x and y coordinates are saved, respectively of: RH, LH, RF and LF. From the eight arrays containing the x and y coordinates of the speed vectors (size (N-1)x2) the magnitude of the normalized velocity vector of each point was evaluated. From these vectors the magnitude of the normalized acceleration vector of each point was evaluated.
As an example figure 7.13 shows the motiongram obtained from a sequence of GMs lasting about 180 frames at 18th weeks of age where two SM, separated by black areas for all limbs at the same frame(s), are detectable: a long one from frame 0 to frame120 and a short one from frame 130 to frame 145.

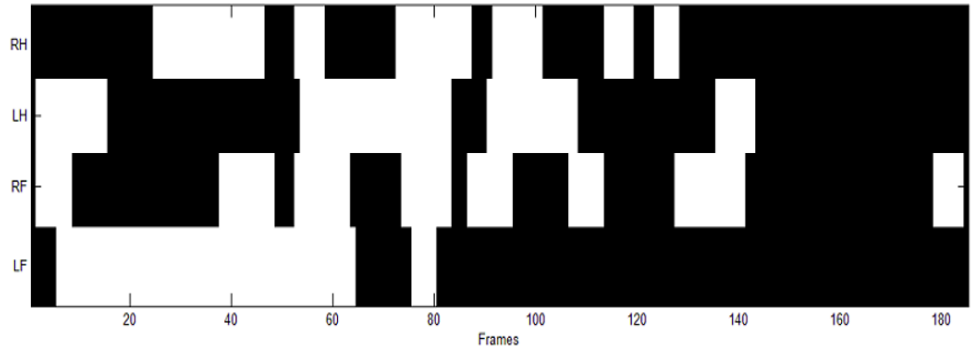


Figure 7.13 Motion diagram. Image obtained from the matrix of speed values of RH, LH, RF and LF for the newborn under study. Two SMs are visually detected: a long one from frame 0 to frame120 and a short one from frame 130 to frame 145.

Table 7.24 summarizes the results concerning the movement analysis for the newborn under study where only 4 limbs are considered: hands and feet. The starting and ending frames are manually selected within each recording. The table shows the initial and final frames, the number of sequences of motion (SM) and the number of movements for each limb (MU) normalized in time. This definitely provides a very useful information to the clinician, both for subsequent analysis of the same movie and for a comparative analysis made by several specialists.

We observe that the number of sequences of motion is variable, with an increasing trend in the last 2 steps. Instead, the number of movements decreases almost always from the first step onwards. The table also shows the values of speed for the 4 limbs as the number of movements per minute. This gives a measure of the frequency of the movement of each limb.

Table 7.24 Results of movement analysis. Number of sequences of motion (SM) and of movements (MU) for each recording step and for the selected frames.

Time step	Start [frame]	Stop [frame]	Start [s]	Stop [s]	Number of SM	RH MU/min	LH MU/min	RF MU/min	LF MU/min
10 days	1	3500	0	140	7	99,86	105,86	48,00	59,57
6 weeks	1	1700	0	68	2	14,12	194,12	13,24	33,53
12 weeks	1	2750	0	110	3	28,36	35,45	29,45	20,18
18 weeks	2750	6450	110	258	28	45,41	30,00	64,86	45,00
24 weeks	1	3500	0	140	11	45,43	22,29	37,71	35,14

Furthermore for each recording, the system provides results concerning speed, acceleration, skewness and kurtosis and correlation between left and right hands and feet. Some results are reported in Table 7.25.

Table 7.25 Results of speed and acceleration. In the table skewness (Sk) and kurtosis (K) are reported for each recording step and for the selected frames

Time step	Sk RH	Sk LH	Sk RF	Sk LF	K RH	K LH	K RF	K LF
10 days	-2,69	-1,55	-3,94	-3,90	9,59	10,76	16,11	14,45
6 weeks	-2,29	-0,05	-1,75	-1,27	18,37	8,53	16,66	13,24
12 weeks	-4,03	-3,06	-1,79	-2,51	20,82	18,96	12,48	17,51
18 weeks	-4,00	-4,39	-2,92	-3,95	15,43	19,11	15,10	17,60
24 weeks	-4,23	-4,48	-3,42	-4,22	19,03	27,00	19,45	22,72

These results could possibly support the analysis of the newborn's body symmetry in the first six month of life in a semi-automatic way.

To show the capabilities of AVIM, the audio/video data and analysis of a typically developed infant recorded in five time-points during the first six months of life are presented.

Concerning audio analysis F_0 mean values range between about 450 Hz and 600 Hz, in accordance to literature [90, 115] with an increasing trend from step 1 to step 4, followed by a slight decrease in the last step. The first formant, F_1 shows an opposite trend, with mean values decreasing from the first to the last step. The reduction in F_1 values is in agreement with the development of the vocal tract of the newborn, whose structure is modified during the first months of life, leading to the development and growth of the pharyngeal cavity [97, 98].

8. Discussion

Traditional techniques for the diagnosis of neurological disorders are recently complemented by contact-less methods. These techniques are mainly based on the assessment of parameters obtained both from automatic and perceptual analysis of audio and/or video recordings resulting in a semi-quantitative evaluation of the patient's status. Contact-less techniques provide advantages in terms of comfort and safety of the patient with respect to sensor-based/invasive methods, thus they are particularly well suited for vulnerable patients such as neonates and young infants. The development of reliable software tools to enhance early diagnosis, especially in home environments, is highly desirable particularly for the neurobehavioral assessment of the newborn. Such tools should provide objective measures to complement clinicians' qualitative analysis that is based on subjective skills. Acoustical analysis of infant's cry and automatic motor methods may provide objective parameters indicative of neurological pathologies.

In this PhD work a new tool is proposed, named AVIM, devoted to the analysis of audio and video recordings of newborn infants. Both the newborn cry and movements are indeed early indicators of neurological disorders. AVIM is made up of three sections: "Patient Data Management", "Recorder tool" and "Analysis tool".

For the automatic analysis of audio recordings AVIM may use a a new procedure based on the continuous wavelet transform named NeCA, developed during this PhD work. The clinically most relevant acoustical parameters of the newborn infant cry are provided in quasi-real time and saved in easily readable tables and plots. The ability to collect and use acoustical data organized systematically and easily accessible and editable can foster the development of a large database useful for the definition of normative ranges for the features of neonatal cry that is still missing.

Moreover, AVIM provides a semi-automatic analysis of spontaneous movements helping the clinician to quickly and better detect the presence or absence of motion and thus a faster selection of the motion sequences of interest. The proposed representation is not exhaustive with regard to aspects such as the quality and smoothness of movement, but is intended as a first step towards setting up an automatic system to assist the perceptual inspection. The distinction of the sequences of movement will be helpful to obtain an objective parameter of their variability in different stages of development of the patient and, once the method will be validated, it could be used to compare motion parameters of different subjects.

The information achieved about the phonatory capabilities of the infant, require a thorough clinical evaluation, but might nevertheless be relevant to define normative ranges for classification.

We remark that unlike available commercial and free software tools for acoustical analysis, this tool allows collecting in a single table the results of multiple recording sessions, with obvious advantages for the clinician that thus can easily and quickly track possible changes in the parameters along a time span of interest.

Concerning the video analysis, to the authors' knowledge, there are no reference values in the literature, so the results will be evaluated carefully by clinicians.

Skewness of velocity and kurtosis of acceleration carry useful information to monitor how the newborn trains his/her cognitive processes. In particular, skewness of velocity of the feet reflects possible unequal distribution of movement velocity. In [197] the skewness is computed for feet velocity only while in [250] the kurtosis is computed for both legs and arms. In our example, skewness values relative to feet are always negative, with an increasing trend is observed from the 6th week on, similarly to [224]. With the proposed tool, the same trend was found for the hands too.

The number of MU/min from the 6th week on results comparable to that found in [250] from the 12th week of age on, while it reaches much higher values until the 6th week of age. Finally, kurtosis shows a slight increasing trend, with values comparable to those found in [250].

Based on the clinical assessment and with the aid of the objective results related to crying and movement obtained with AVIM, the clinical evaluation has confirmed the normal condition for this baby in all the 5 steps of recording: movements are judged as smooth and fluent and the crying acoustical features regular in relation to the newborn’s age.

AVIM provides clinicians with a storyboard of the infant neuro-behavioural assessment allowing an easy and fast comparison of the results obtained through different tests. In addition to provide an appreciable decrease

in investigation time, costs and errors, this option is particularly relevant being conceived for simplifying and supporting the actions carried out by the clinician during the perceptual analysis.

To the authors' knowledge, to date no software tool exists integrating in one system all the items required to perform both of these marker-less analysis.

As already mentioned in Sect.1 the acoustical analysis of the infant cry is a non-invasive approach to assist the clinical specialist in the detection of abnormalities in infants with possible neurological disorders.

To date, the analysis is most often carried out with a perceptive examination based on listening to the cry and visually inspecting the signal waveform and its spectrogram. However, this approach is operator-dependent and requires a considerable amount of time often prohibitive in daily clinical practice. Thus, the scientific community is paying special attention to techniques devoted to the accurate automatic analysis of the cry.

In Chapter 1 and 2, the physiological motivations and several approaches for fundamental frequency (F_0) and resonance frequency (RFs) estimation are mentioned and described. The difficulty in the estimation of F_0 and RFs is mainly linked to the quasi-stationarity and the very high range of frequencies of interest in the newborn cry that require sophisticated adaptive numerical techniques characterized by high time-frequency resolution. The importance of this PhD thesis is the development an efficient and repeatable method for the cry analysis from the acquisition of the signal to the classification of crying.

Besides papers on cry classification, to the author's knowledge this is the first application of the wavelets to the estimation of the frequencies characteristic of the newborn crying. The method does not require manual setting of thresholds and is thus usable also by non-experts. The computing time is comparable to PRAAT: for 1 s of recording NeCA requires 0.9 seconds for the estimation of F_0 and 2.8 s for the estimation of RFs, against less than 0.5 and approximately 2s respectively with PRAAT. However, the CUs obtained with PRAAT are less reliable and a careful manual setting of ranges and thresholds is required to avoid meaningless results especially for RFs [119].

As already noted, the selection of CUs was carried out with BioVoice. This method is much slower than NeCA (it requires 3s for F_0 estimation and 4.7 s for the RFs), given the two-step estimation of F_0 on signal frames of variable length that, for the neonatal cry, may be of very limited duration (typically 4-8ms) thus providing at least twice estimated values than those obtained with NeCA. Such higher resolution could give reason of the best performance of BioVoice for RFs estimation.

NeCA is applied to a quite large real data set coming from typically developed newborns. Results are promising. The estimated values of F_0 and RFs are in the ranges reported in the literature. Finally it is important to emphasize again that the results present in the literature regarding only F_0 and the increasing trend of F_0 values during the first six months of life is in agreement with the literature.

CONCLUSION

In addition to a thorough literature review for the definition of the state of the art this PhD thesis concerns the implementation, development and use of automatic tools for collecting and analysing both audio and video data of infant crying and movements as an aid to the early diagnosis of ASDs. Specifically the PhD project has provided first results concerning the automatic analysis of newborn crying through the development of new algorithms, the improvement of existing ones and their comparison both on synthetic and real audio signals. A first approach for the automatic analysis of GMs was also developed as a support to the perceptual techniques used by clinicians. Both of these studies were carried out with the aim of showing that, as both GMs and cry features are related to the development and the integrity of the central nervous system, they could be successfully exploited through the analysis of audio and video recordings in the first months of life. Such contactless techniques are appealing for clinical diagnosis of several pathologies being easy to perform, cheap and completely non-invasive. Moreover, an automatic procedure that helps in finding correlations between general movements and cry would be of great relevance and could be exploited as an aid to early clinical diagnosis of several pathologies being easy to perform, cheap and completely non-intrusive. This is indeed a timely challenge for health informatics and for biomedical engineering.

From the methodological point of view, this work required to face many difficulties, mainly due to the lack of stationarity of newborn cry, the high variability of the frequencies involved and the need to exactly differentiate the parts of the signal of interest for the acoustical analysis that can be easily confused with components of different nature. All these problems have been addressed with a totally innovative approach as in the literature there are no effective tools for this kind of analysis.

Future studies will be devoted to the analysis of gestures such as the fluidity and elegance of movements that are difficult to be objectively quantified. With an adequate number of cases it will be possible to build a data set that would allow assessing the effectiveness of the proposed method and checking whether the above features can be discriminative for clinical diagnosis.

REFERENCES

[1] Center for Disease Control (CDC). Prevalence of autism spectrum disorder among children aged 8 years – Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2010. MMWR 2014; 63(SS01); 1-21.

[2] American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

[3] Werner E, Dawson G, Osterling J, Dinno N, (2002). Brief report: recognition of autism spectrum disorder before one year of age: retrospective study based on home videotapes, Journal of autism developmental disorders, 30:157-162.

[4] Maestro S, Casella C, Milone A, Muratori F, & Palacio-Espasa F, (1998). Study of the onset of autism through home movies. Psychopathology, 32(6):292-300.

[5] Maestro S, Muratori F, Barbieri, F., Casella, C., Cattaneo, V., Cavallaro, M. C., ... & Palacio-Espasa, F. (2000). Early behavioral development in autistic children: the first 2 years of life through home movies. Psychopathology, 34(3), 147-152.

[6] Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. International journal of developmental neuroscience, 23(2), 143-152.

[7] Bryson, S. E., & Zwaigenbaum, L. (2014). Autism Observation Scale for Infants. In Comprehensive Guide to Autism (pp. 299-310). Springer New York.

[8] Bryson SE, Zwaigenbaum L, McDermott C et al. (2008) The Autism Observation Scale for Infants: scale development and reliability data. J Autism Dev Disord 38:731-738.

[9] Black LS, deRegnier RA, Long J et al. (2004) Electrographic imaging of recognition memory in 34-38 week gestation intrauterine growth restricted newborns. Experimental Neurology 190: S72-S83.

[10] Howard JS, Sparkman CR, Cohen HG et al. (2005) A comparison of intensive behavior analytic and eclectic treatments for young children with autism. Res Dev Disabil, 26: 359-383

[11] American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). 4th ed. Washington (DC): APA (1994).

[12] Teitelbaum R, Glatman-Freedman A, Chen B et al. (1998) Movement analysis in infancy may be useful for early diagnosis of autism. PNAS, 95: 13982-13987.

[13] Michelsson K, Michelsson O. Phonation in the newborn, infant cry. Int J Pediatr Otorhinolaryngol. 1999 Oct 5;49 Suppl 1:S297-301.

[14] Bieberich, A., & Morgan, S.B. (1998). Brief report: Affective expression in children with autism or Down's syndrome. Journal of Autism and Developmental Disorders, 28, 333–338.

[15] Esposito, G., & Venuti, P. (2008). How is crying perceived in children with Autistic Spectrum Disorder? Research in Autism Spectrum Disorders, 2, 371–384.

[16] Amaral, D., Schumann, C., & Nordahl, C. (2008). Neuroanatomy of autism. Trends in Neurosciences, 31, 137–145.

[17] Schulkin, J. (2008). Autism and the amygdala: An endocrine hypothesis. Brain and Cognition, 65, 87–99.

[18] Kanner L. 1943. Autistic disturbances of affective contact. Nervous Child 2:217-250.

[19] Glazebrook C, Gonzalez D, Hansen S, Elliott D, (2011) Current Perspectives on Motor Functioning in Infants, Children, and Adults With Autism Spectrum Disorders, Journal of the American Physical Therapy Association, 91: 1116-1129.

[20] Bauman ML, Kemper TL, (2005) Neuroanatomic observations of the brain in autism: a review and future directions", International Journal of Developmental Neuroscience, 23(2–3):183-187.

[21] Baranek GT. 2002. Efficacy of sensory and motor interventions for children with autism. Journal of Autism and Developmental Disorders. 32:397-422.

[22] Dawson G , Watling R. 2000. Interventions to facilitate auditory, visual, and motor integration in autism: A review of the evidence. Journal of Autism and Developmental Disorders 30:415-421.

[23] Gevers C, Clifford P, Mager M, Boer, F. 2006. Brief report: A theory of mind-based social cognition training program for school aged children with pervasive developmental disorders. Journal of Autism and Developmental Disorders 36(4):567-571.

[24] Smith C, Goddard S, Fluck M. 2004. A scheme to promote social attention and functional language in young children with communication difficulties and autistic spectrum disorder. Educational Psychology in Practice 20:319-333.

[25] Homer R, Car E, Strain P, Todd A, Reed H. 2002. Problem behavior interventions for young children. Journal of Autism and Developmental Disorders 32: 423-426.

[26] Wing L. 1997. The autism spectrum. Lancet 350:1761-1766.

[27] American Psychiatric Association. Highlights of Changes from DSM-IV-TR to DSM-5. 2013

[28] http://www.autismspeaks.org/what-autism/diagnosis/dsm-5-diagnostic-criteria. Accessed August 16, 2014.

[29] E. C. H. C. P. Directorate-General (2005) Some elements about the prevalence of Autism Spectrum Disorder (ASD) in the European Union

[30] World Health Organization (WHO). (2013) Comprehensive and coordinated efforts for the management of autism spectrum disorders (EB133/4)

[31] Fombonne E. The prevalence of autism. JAMA. Jan 1 2003;289(1):87-9.

[32] Prevalence of autism spectrum disorders--autism and developmental disabilities monitoring network, 14 sites, United States, 2002. MMWR Surveill Summ. Feb 9 2007;56(1):12-28.

[33] Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., ... & Stone, W. L. (2011). Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. Pediatrics,128(3), e488-e495.

[34] Nazeer, A. (2011). Psychopharmacology of autistic spectrum disorders in children and adolescents. Pediatric Clinics of North America, 58(1), 85-97.

[35] Rutter, M. (2000). Genetic studies of autism: from the 1970s into the millennium. Journal of Abnormal Child Psychology, 28(1), 3-14.

[36] Vastag, B. (2004). National autism summit charts a path through a scientific, clinical wilderness. JAMA, 291(1), 29-31.

[37] Marchese, M., Conti, V., Valvo, G., Moro, F., Muratori, F., Tancredi, R., ... & Sicca, F. (2014). Autism-epilepsy phenotype with macrocephaly suggests PTEN, but not GLIALCAM, genetic screening. BMC medical genetics, 15(1), 26.

[38] Skuse, D. H. (2000). Imprinting, the X-chromosome, and the male brain: explaining sex differences in the liability to autism. Pediatric Research, 47(1), 9-9.

[39] Shriber L. 2010. Autism: A Neurological and Sensory Based Perspective. In: JH Stone, M Blouin, editors. International Encyclopedia of Rehabilitation. Available online: http://cirrie.buffalo.edu/encyclopedia/en/article/285/

[40] Auranen M, Vanhala R, Varilo T, et al. 2002. A genomewide screen for autism spectrum disorders: Evidence for a major susceptibility locus on chromosome 3q25-27. American Journal of Human Genetics 74(1):777-790.

[41] Badner JA, Gershon ES. 2002. Regional meta-analysis of published data supports linkage of autism with markers on chromosome 7. Molecular Psychiatry 7(1): 56-66.

[42] Buxbaum JD, Silverman, Smith CJ, et al. 2001. Evidence for susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. American Journal of Human Genetics 68:1514-1520.

[43] Gallagher, L., Becker, K., Kearney, G., Dunlop, A., Stallings, R., Green, A., ... & Gill, M. (2003). Brief report: A case of autism associated with del (2)(q32. 1q32. 2) or (q32. 2q32. 3). Journal of autism and developmental disorders, 33(1), 105-108.

[44] Hutcheson, H. B., Bradford, Y., Folstein, S. E., Gardiner, M. B., Santangelo, S. L., Sutcliffe, J. S., & Haines, J. L. (2003). Defining the autism minimum candidate gene region on chromosome 7. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 117(1), 90-96.

[45] Singh VK, Lin SX, Yang VC. 1998. Serological association of measles virus and human herpes virus-6 with brain autoantibodies in autism. Clinical Immunology and Immunopathology 89(1):105-108.

[46] Fombonne E, Chakrabarti S. 2001. No evidence for a new variant of measles-mumps-rubella-induced autism. Pediatrics 108(4):E58.

[47] Hviid A, Stellfeld M, Wohlfahrt J, Melbye M. 2003. Association between thimerosol-containing vaccine and autism. Journal of the American Medical Association 290(13):1763-1766.

[48] Miller E. 2003. Measles-mumps-rubella vaccine and the development of autism. Seminars in Pediatric Infectious Diseases 14(3):199-206.

[49] Taylor B, Miller E, Lingam R, Andrews N, Simmons A, Stowe J. 2002. Measles, mumps, and rubella vaccination and bowel problems or developmental regression in children with autism: Population study. British Medical Journal 324(7334): 393-396.

[50] Wilson K, Mills E, Ross C, McGowan J, Jadad A. 2003. Association of autism spectrum disorder and the measles, mumps, and rubella vaccine: A systematic review of current epidemiological evidence. Archives of Pediatric and Adolescent Medicine 157(7):628-634.

[51] Lucarelli S, Frediani T, Zigoni AM, Ferruzi F, Giardini O, Quintieri F, et al. 1995. Food allergy and infantile autism. Panminerva Medicine 37:137-141.

[52] Reichelt KL, Ekrem J, Scott H. 1990. Gluten, mild proteins and autism: Dietary intervention effects on behavior and peptide secretion. Journal of Applied Nutrition 42:1-11.

[53] Reichelt KL, Knivsberg AM. 2003. Can the pathophysiology of autism be explained by the nature of the discovered urine peptides? Nutritional Neuroscience 6:19-28.

[54] Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. Development and psychopathology,20(03), 775-803.

[55] Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., ... & Varley, J. (2010). Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. Pediatrics, 125(1), e17-e23.

[56] Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: A study of first birthday home videotapes. Journal of autism and developmental disorders, 24(3), 247-257.

[57] Yeargin-Allsopp, M., Rice, C., Karapurkar, T., Doernberg, N., Boyle, C., & Murphy, C. (2003). Prevalence of autism in a US metropolitan area. Jama,289(1), 49-55.

[58] Lester, B. M., & Boukydis, C. Z. (1992). 8. No language but a cry. Nonverbal vocal communication: Comparative and developmental approaches, 145.

[59] Adrien, J. L., Perrot, A., Sauvage, D., & Leddet, I. (1992). Early symptoms in autism from family home movies: Evaluation and comparison between 1st and 2nd year of life using IBSE Scale. Acta Paedopsychiatrica: International Journal of Child & Adolescent Psychiatry.

[60] Clifford, S., Dissanayake, C., Bui, Q. M., Huggins, R., Taylor, A. K., & Loesch, D. Z. (2007). Autism spectrum phenotype in males and females with fragile X full mutation and premutation. Journal of autism and developmental disorders,37(4), 738-747.

[61] Clifford, S. M., & Dissanayake, C. (2008). The early development of joint attention in infants with autistic disorder using home video observations and parental interview. Journal of autism and developmental disorders, 38(5), 791-805.

[62] Maestro, S., Muratori, F., Cavallaro, M. C., Pei, F., Stern, D., Golse, B., & Palacio-Espasa, F. (2002). Attentional skills during the first 6 months of age in autism spectrum disorder. Journal of the American Academy of Child & Adolescent Psychiatry, 41(10), 1239-1245.

[63] Maestro, S., Muratori, F., Cavallaro, M. C., Pecini, C., Cesari, A., Paziente, A., ... & Palacio-Espasa, F. (2005). How young children treat objects and people: an empirical study of the first year of life in autism. Child psychiatry and human development, 35(4), 383-396.

[64] Osterling, J. A., Dawson, G., & Munson, J. A. (2002). Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation.Development and psychopathology, 14(02), 239-251.

[65] Ozonoff, S., Iosif, A. M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., ... & Young, G. S. (2010). A prospective study of the emergence of early behavioral signs of autism. Journal of the American Academy of Child & Adolescent Psychiatry, 49(3), 256-266.

[66] Werner, E., Dawson, G., Osterling, J., & Dinno, N. (2000). Brief report: Recognition of autism spectrum disorder before one year of age: A retrospective study based on home videotapes. Journal of autism and developmental disorders, 30(2), 157-162.

[67] Baranek GT (1999) Autism during infancy: a retrospective video analysis of sensory-motor and social behaviors at 9-12 months of age, J Autism Dev Disord, 29(3):213-224

[68] Fernell, E., Eriksson, M. A., & Gillberg, C. (2013). Early diagnosis of autism and impact on prognosis: a narrative review. Clinical epidemiology, 5, 33.

[69] Pinborough-Zimmerman, J., Bakian, A. V., Fombonne, E., Bilder, D., Taylor, J., & McMahon, W. M. (2012). Changes in the administrative prevalence of autism spectrum disorders: Contribution of special education and health from 2002–2008. Journal of autism and developmental disorders, 42(4), 521-530.

[70] Frith, U., Morton, J., & Leslie, A. M. (1991). The cognitive basis of a biological disorder: Autism. Trends in neurosciences, 14(10), 433-438.

[71] Lord, C., Risi, S., Lambrecht, L., Cook Jr, E. H., Leventhal, B. L., DiLavore, P. C., ... & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of autism and developmental disorders,30(3), 205-223.

[72] Tager-Flusberg, H., & Caronna, E. (2007). Language disorders: autism and other pervasive developmental disorders. Pediatric Clinics of North America,54(3), 469-481.

[73] Baron-Cohen, S., Lombardo, M., Tager-Flusberg, H., & Cohen, D. (Eds.). (2013). Understanding Other Minds: Perspectives from developmental social neuroscience. Oxford University Press.

[74] Abney, D. H., Warlaumont, A. S., Haussman, A., Ross, J. M., & Wallot, S. (2014). Using nonlinear methods to quantify changes in infant limb movements and vocalizations. Frontiers in Psychology, 5, 771. doi:10.3389/fpsyg.2014.00771

[75] P. Venuti, G. Esposito, Z. Giusti, "A qualitative analysis of crying and vocal distress in children with autism", Journal of Intellectual Disability Research, vol.48, 2004, pp. 4-5.

[76] Wasz-Höckert O, Vuorenkoski V, Valanne E, Michelsson K. Sound spectrographic studies of the cry of newborn infants. Experientia 1962; 18:583-584.

[77] Wasz-Höckert O, Valanne E, Vuorenkoski V, Michelsson K, Sovijarvi A. Analysis of some types of vocalization in the newborn and in early infancy. Ann Paediatr Fenn 1963; 9:1-10

[78] Wasz-Höckert O, Partanen TJ, Vuorenkoski V, Michelsson K, Valanne E. The identification of some specific meanings in infant vocalization. Experientia 1964; 20:154

[79] Wasz-Höckert O, Partanen T, Vuorenkoski V, Valanne E, Michelsson K. Effect of training on ability to identify preverbal vocalizations. Dev Med Child Neurol 1964; 89:393-396

[80] Michelsson K. Cry analyses of symptomless low birth weight neonates and of asphyxiated newborn infants. Acta Paediatr Scand Suppl 1971; 216:1-45

[81] Michelsson K, Sirviö P, Koivisto M, Sovijarvi A, Wasz-Höckert O. Spectrographic analysis of pain cry in neonates with cleft palate. Biol Neonate 1975; 26:353-358

[82] Michelsson K, Sirviö P, Wasz-Höckert O. Pain cry in full-term asphyxiated newborn infants correlated with late findings. Acta Paediatr Scand 1977; 66:611-616

[83] Michelsson K, Sirviö P, Wasz-Höckert O. Sound spectrographic cry analysis of infants with bacterial meningitis. Dev Med Child Neurol 1977; 19:309-315

[84] Lester BM. Spectrum analysis of the cry sounds of well-nourished and malnourished infants. Child Dev 1976; 47:237-241

[85] Lester BM, Dreher M. Effects of marijuana use during pregnancy on newborn cry. Child Dev 1989; 60:765-771

[86] Lester BM, Corwin MJ, Sepkoski C, Seifer R, Peucker M, McLaughlin S, et al. Neurobehavioral syndromes in cocaine-exposed newborn infants. Child Dev 1991; 62:694-705

[87] Corwin MJ, Lester BM, Sepkoski C, McLaughlin S, Kayne H, Golub HL. Effects of in utero cocaine exposure on newborn acoustical cry characteristics. Pediatrics 1992; 89(6 Pt2):1199-1203

[88] Corwin MJ, Lester BM, Sepkoski C, Peucker M, Kayne H, Golub HL. Newborn acoustic cry characteristics of infants subsequently dying of sudden infant death syndrome. Pediatrics 1995; 96(1 Pt1):73-77

[89] Michelsson K, Eklund K, Leppanen P, Lyytinen H. Cry characteristics of 172 healthy 1-to 7-day-old infants. Folia Phoniatr Logop 2002; 5:190-200

[90] Baeck HE, de Souza MN. Longitudinal study of the fundamental frequency of hunger cries along the first 6 months of healthy babies. J Voice 2007; 5:551-559

[91] Ludington-Hoe SM, Cong X, Hashemi F. Infant crying: nature, physiologic consequences, and select interventions. Neonatal Netw 2002; 21:29-36

[92] Ohgi S, Gima H, Akiyama T. Neonatal behavioural profile and crying in premature infants at term age. Acta Paediatr 2006; 95:1375-1380

[93] Vohr BR, Lester B, Rapisardi G, O'Dea C, Brown L, Peucker M, et al. Abnormal brain-stem function (brain-stem auditory evoked response) correlates with acoustic cry features in term infants with hyperbilirubinemia. J Pediatr 1989; 115:303-308

[94] Rapisardi G, Vohr B, Cashore W, Peucker M, Lester B. Assessment of infant cry variability in high-risk infants. Int J Pediatr Otorhinolaryngol 1989; 17:19-29.

[95] Koivisto M. Cry analysis in infants with Rh haemolytic disease. Acta Paediatr Scand Suppl 1987; 335:1-73

[96] Fuller BF. Acoustic discrimination of three types of infant cries. Nurs Res 1991;40(3):156-160

[97] A. Fort, A. Ismaelli, C. Manfredi, P. Bruscaglioni, "Parametric and non-parametric estimation of speech formants: application to infant cry", Med. Eng. Phys., vol. 18, no. 8, pp. 677–91, 1996

[98] A. Fort, C. Manfredi, "Acoustic analysis of newborn infant cry signals", Med. Eng. Phys, vol. 20, pp. 432-442, 1998

[99] Chuang SM, Wang TR, Jean HH, Lee FY. The cat cry (cri du chat) syndrome: report of a case with review of 10 cases at the National Taiwan University Hospital. Taiwan Yi Xue Hui Za Zhi 1989;88(6): 628-629, 635-638

[100] Wasz-Hockert O, Koivisto M, Vuorenkoski V, Partanen TJ, Lind J. Spectrographic analysis of pain cry in hyperbilirubinemia. Biol Neonate 1971;17(3):260-271

[101] Vats-Hekkert O, Mikhelson K, Sirvio P. Cry analysis and brain lesions in newborn children. Zh Nevropatol Psikhiatr Im S S Korsakova 1978;78(10):1482-1485

[102] Goberman AM, Robb MP. Acoustic examination of preterm and full-term infant cries—the long-time average spectrum. J SpeechLang HearRes 1999;42:850–61

[103] Michelsson K, Järvenpää AL, Rinne A. Sound spectrographic analysis of pain cry in preterm infants. Early Hum Dev 1983;8(2):141–9

[104] Lester BM. Developmental outcome prediction from acoustic cry analysis in term and preterm infants. Paediatrics 1987;80(4):529–34

[105] Bellieni CV, Sisto R, Cordelli DM, BuonocoreG. Cry features reflect pain intensity in term newborns: an alarm threshold. Pediatr Res 2004;55(1):142–6

[106] Robb MP, Goberman AM, Cacace AT. An acoustic template of newborn infant crying. Folia Phoniatr Logop 1997;49(1):35–41

[107] C. Manfredi, L. Bocchi, S. Orlandi, L. Spaccaterra, G.P. Donzelli, High-resolution cry analysis in preterm newborn infants, Medical Engineering & Physics 31 (2009) 528–532

[108] M. Petroni, A.S. Malowany, C.C. Johnston, B.J. Stevens, "Identification of Pain from Infant Cry Vocalizations Using Artificial Neural Networks," Proc. of SPIE, vol. 2492, pp. 729-738, 1995

[109] J.R. Deller, J.H. Hansen, J.G. Proakis, Discrete-Time Processing of Speech Signal, McMillan Publishing Company, New York, 1993, pp. 724–728.

[110] Amaro-Camargo, E., & Reyes-García, C. A. (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 (pp. 1078-1085). Springer Berlin Heidelberg

[111] Reyes-Galaviz, O. F., Tirado, E. A., & Reyes-Garcia, C. A. (2004).Classification of infant crying to identify pathologies in recently born babies with ANFIS (pp. 408-415). Springer Berlin Heidelberg. Computers Helping People with Special Needs, edited by Klaus Miesenberger et. al., Springer, Berlin, 2004, pp 408-415.

[112] K. Wermke, W. Mende, C. Manfredi, P. Bruscaglioni, Developmental aspects of infant's cry melody and formants, Med. Eng. Phys. 24(2002) 501-514.

[113] H. R. Gilbert, M. P. Robb, Vocal Fundamental frequency characteristics of infant hunger cries': birth to 12 months, Int. J. Pediatr. Otorhinolaryngol. 34 (1996) 231-243.

[114] K. Lind, K. Wermke, Development of the vocal fundamental frequency of spontaneous cries during the first 3, Int. J. Pediatr. Otorhinolaryngol. 64 (2002) 97-104.

[115] Rothganger H. Analysis of the sounds of the child in the first year of age and a comparison to the language. Early Hum Dev 2003; 75:55-69.

[116] http://www.kayelemetrics.com

[117] http://www.fon.hum.uva.nl/praat

[118] P. Boersma, (2002) Praat, a system for doing phonetics by computer. Glot Internat. 5-9/10 (2002) 341-345.

[119] Boersma P, Acoustic analysis, in Research methods in linguistics. R. Podesva and D. Sharma (eds.), Cambridge University Press, 2014.

[120] Wermke, K., & Mende, W. (2009). Musical elements in human infants' cries: In the beginning is the melody. Musicae Scientiae, 13(2 suppl), 151-175.

[121] Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. Current biology, 19(23), 1994-1997.

[122] Wermke, K., et al., Cry melody in 2-month-old infants with and without clefts. Cleft Palate Craniofac J, 2011. 48(3): p. 321 -30.

[123] Varallyay, G., Jr., The melody of crying. Int J Pediatr Otorhinolaryngol, 2007. 71(11): p. 1699-708.

[124] Schönweiler, R., Kaese, S., Möller, S., Rinscheid, A., & Ptok, M. (1996). Neuronal networks and self-organizing maps: new computer techniques in the acoustic evaluation of the infant cry. International journal of pediatric otorhinolaryngology, 38(1), 1-11.

[125] Varallyay, G., & Benyó, Z. (2007, August). 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 (pp. 4119-4122). IEEE.

[126] Sirviö P, Michelsson K. Sound-spectrographic cry analysis of normal and abnormal newborn infants. Folia Phoniatrica et Logopaedica 1976, 28(3):161-173.

[127] Laufer MZ, Horii Y. Fundamental frequency characteristics of infant non-distress vocalization during the first twenty-four weeks. J Child Lang 1977; 4(02):171-184.

[128] Michelsson K, Michelsson O. Phonation in the newborn, infant cry. Int. J Pediatr Otorhinolaryngol 1999; 49:S297-S301.

[129] Robb MP, Crowell DH, Dunn-Rankin P. Sudden Infant Death Syndrome: Cry characteristics. Int J Pediatr Otorhinolaryngol 2013; 77(8):1263-1267.

[130] Reggiannini B, Sheinkopf SJ, Silverman HF, Li X, Lester BM. A Flexible Analysis Tool for the Quantitative Acoustic Assessment of Infant Cry. J Speech Lang Hear Res 2013; 56(5):1416-1428.

[131] Kheddache, Y., & Tadj, C. (2015). Resonance Frequencies Behavior in Pathologic Cries of Newborns. Journal of Voice, 29(1), 1-12. A. Fort, C. Manfredi, Acoustic analysis of newborn infant cry signals, Med. Eng. Phys 20 (1998) 432-442.

[132] Robb MP, Cacace AT. Estimation of formant frequencies in infant cry. Int J Pediatr Otorhinolaryng 1995; 32(1): 57-67.

[133] Poel M., Ekkel T. Analyzing infant cries using a committee of neural networks in order to detect hypoxia related disorder 2006 International Journal on Artificial Intelligence Tools 15 3 397-410.

[134] Lederman, E. Zmora, S. Hauschildt, A. Stellzing-Eisenhauer, K. Wermke, Classification of cries of infants with cleft-palate using parallel hidden Markov models. Med Biol Eng Comput 46-10 (2008) 965-75

[135] B. Mijovic; M, Silva; B. R. H. Van den Bergh; K. Allegaert; J. M. Aerts; D. Berckmans; S. Van Huffel Assessment of Pain Expression in Infant Cry Signals Using Empirical Mode Decomposition, Methods Inf Med 2010; 49: 448–452

[136] R. Sahak, W. Mansor, Y. K. Lee, A. I. M. Yassin, A. Zabidi, Performance of Combined Support Vector Machine and Principal Component Analysis in Recognizing Infant Cry with Asphyxia, 32nd Annual International Conference of the IEEE EMBS Buenos Aires, Argentina, August 31 - September 4, 2010 p.6292-6295.

[137] Zabidi, A., Khuan, L. Y., Mansor, W., Yassin, I. M., & Sahak, R. (2010, August). Optimization of MFCC parameters using Particle Swarm Optimization for diagnosis of infant hypothyroidism using Multi-Layer Perceptron. In Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE (pp. 1417-1420). IEEE.

[138] Jean-Julien Yulri Nonaka, Kentaro Katahira, and Kazuo Okanoya, Segmentation of expiratory and inspiratory sounds in baby cry audio recordings using hidden Markov models, J. Acoust. Soc. Am. 130 (5), November 2011 p.2969-2977.

[139] M. Hariharan & Lim Sin Chee & Sazali Yaacob, Analysis of Infant Cry Through Weighted Linear Prediction Cepstral Coefficients and Probabilistic Neural Network J Med Syst (2012) 36:1309–1315.

[140] M. Hariharan , R. Sindhu, Sazali Yaacob, Normal and hypoacoustic infant cry signal classification using time–frequency analysis and general regression neural network, Computer Methods and Programs in Biomedicine Volume 108, Issue 2, November 2012, Pages 559–569

[141] Tanja Etz Henning Reetz Carla Wegener, A Classification Model for Infant Cries with Hearing Impairment and Unilateral Cleft Lip and Palate, Folia Phoniatr Logop 2012;64:254–261

[142] Alaie, H. F., & Tadj, C. (2012). Cry-based classification of healthy and sick infants using adapted boosting mixture learning method for gaussian mixture models. Modelling and Simulation in Engineering, 2012, 55.

[143] Singh A.K., Mukhopadhyay J., Rao K.S., Viswanath K., Classification of infant cries using dynamics of epoch features, Journal of Intelligent Systems 2013 22 3 351-364

[144] Witten, I. H., & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

[145] H. Phagava, F. Muratori, C. Einspieler, S. Maestro, F. Apicella, G. Guzzetta, H. F. R. Prechtl, G. Cioni, General movements in infants with autism spectrum disorders, Georgian Med. News, 3-156 (2008) 100-105.

[146] C. Einspieler, H. F. R. Prechtl, Prechtl's assessment of general movements: A diagnostic tool for the functional assessment of the young nervous system. Ment. Retard. Dev. Disabil. Res. Rev. 11-1 (2005) 61-67.

[147] C. Einspieler, H. F. R. Prechtl, A. F. Bos, F. Ferrari, G. Cioni, Prechtl's Method on the Qualitative Assessment of General Movements in Preterm, Term and Young Infants, New York: Cambridge University Press (2004)

[148] B. Hopkins H. F. R. Prechtl, A qualitative approach to the development of movements during early infancy, in Continuity of neural functions from prenatal to postnatal life, Ed. H. F. R. Prechtl, London, Spastic International Medical Publications, (1984) 179–197.

[149] M. Coluccini, S. Maini, A. Sabatani, H. F. R. Prechtl, G. Cioni, Kinematic analysis of general movements in early infancy. Dev. Med. Child Neurol. 44 (2002) 14.

[150] L. Adde, J. L. Helbostad, A. R. Jensenius, G. Taraldsen, K. H. Grunewaldt, R. Støen, ENIGMA - Enhanced interactive general movement assessment, Expert Syst. Appl. 34 (2008) 2664–2672.

[151] L. Meinecke, N. Breitbach-Faller, C. Bartz, R. Damen, G. Rau, C. Disselhorst-Klug, Movements analysis in early detection of newborns at risk of developing spasticity due to infantile celerebral palsy, Hum. Mov. Sci. 25 (2006) 125-144.

[152] N. Kanemaru, H. Watanabe, G. Taga, Increasing selectivity of interlimb coordination during spontaneous movements in 2- to 4-month-old infants. Exp. Brain Res. 218-1(2012) 49-61.

[153] C. Disselhorst-Klug, F. Heinze, N. Breitbach-Faller, T. Schmitz-Rode, G. Rau, Introduction of a method for quantitative evaluation of spontaneous motor activity development with age in infants. Exp Brain Res, 218-2 (2012)305-313.

[154] L. Adde, J. L. Helbostad, A. R. Jensenium, G. Taraldsen, R. Støen, Using computer-based video analysis in the study of fidgety movements, Early Hum. Dev. 85 (2009) 541-547.

[155] G. Esposito, P. Venuti, "Developmental changes in the fundamental frequency (f0) of infants' cries: A study of children with Autism Spectrum Disorder", Early Child Dev. Care, vol. 180, no. 8, pp. 1093-1102, 2010.

[156] S. J. Sheinkopf, J. M. Iverson, B. M. Lester, "Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder", Autism Res., vol. 5 no. 5, pp. 331-339, 2012.

[157]Bolton, P. F., Golding, J., Emond, A., & Steer, C. D. (2012). Autism spectrum disorder and autistic traits in the Avon Longitudinal Study of Parents and Children: precursors and early signs. J Am Acad Child Adolesc Psychiatry, 51(3), 249-260 e225.

[158] Accardo, P. J. (2013). 50 years ago in The Journal of Pediatrics: the cry latencies of normal infants and those with brain damage. J Pediatr, 162(5), 917.

[159] Ban, H., Higuchi, R., Noda, E., Koike, M., Kusumoto, S., & Keiichirou, S. (1994). A case of a cholelithiasis in a very low birth weight infant. Acta Neonatologica Japonica, 30(1), 193-198.

[160] Stern, J. M. (1997). Offspring-induced nurturance: animal-human parallels. Dev Psychobiol, 31(1), 19-37.

[161] Zimmerman, R. K., Middleton, D. B., Burns, I. T., Clover, R. D., & Kimmel, S. R. (2007). Routine vaccines across the life span, 2007. Journal of Family Practice, 56(SUPPL. 2), S18-S37.

[162] Simon, E. N. (2004). The lessons of MMR [2]. Lancet, 363(9419), 1473-1474.

[163] Senator, D. (2006). Signs of autism in the first year of life. Pediatria Polska, 81(2), 126-133.

[164] Rümke, H. C., & Visser, H. K. A. (2004). Childhood vaccinations anno 2004. II. The real and presumed side effects of vaccination. Nederlands Tijdschrift voor Geneeskunde, 148(8), 364-371.

[165] Phadke, M. A., Patki, P. S., Kulkarni, P. S., Jadhav, S. S., & Kapre, S. V. (2004). Pharmacovigilance on MMR vaccine containing L-Zagreb mumps strain. Vaccine, 22(31-32), 4135-4136.

[166] Leslie, A. M., Mallon, R., & DiCorcia, J. A. (2006). Transgressors, victims, and cry babies: is basic moral judgment spared in autism? Social neuroscience, 1(3-4), 270-283.

[167] Kurian, M. A., Ryan, S., Besley, G. T. N., Wanders, R. J. A., & King, M. D. (2004). Straight-chain acyl-CoA oxidase deficiency presenting with dysmorphia, neurodevelopmental autistic-type regression and a selective pattern of leukodystrophy. Journal of Inherited Metabolic Disease, 27(1), 105-108,

[168] Kimmel, S. R., Burns, I. T., Wolfe, R. M., & Zimmerman, R. K. (2007). Addressing immunization barriers, benefits, and risks. Journal of Family Practice, 56(SUPPL. 2), S61-S69.

[169] Kimmel, S. R., & Wolfe, R. M. (2005). Communicating the benefits and risks of vaccines. Journal of Family Practice, 54(SUPPL. 1), S51-S57.

[170] Giffin, R., Stratton, K., & Chalk, R. (2004). Childhood vaccine finance and safety issues. Health Affairs, 23(5), 98-111.

[171] Duffner, P. K., Jalal, K., & Carter, R. L. (2009). The Hunter's Hope Krabbe Family Database. Pediatric Neurology, 40(1), 13-18.

[172] Tomas, L. (2009). Best breastfeeding and formulas. Journal of Complementary Medicine, 8(2), 32-40.

[173] Sigman, M., Dissanayake, C., Corona, R., & Espinosa, M. (2003). Social and cardiac responses of young children with autism. Autism, 7(2), 205-216.

[174] Örtqvist, Ä., Blennow, M., Carlsson, R. M., Hanson, L., Lindberg, A., Lindqvist, L., et al. (2010). Vaccination of children - A systematic review. Acta Paediatrica, International Journal of Paediatrics, 99(SUPPL. 461), 1-192.

[175] Mamidala, M. P., Polinedi A, P. T. V. P., Rajesh, N., Vallamkonda, O. R., Udani, V., & Singhal, N., Rajesh, V. (2013). Prenatal, perinatal and neonatal risk factors of Autism Spectrum Disorder: A comprehensive epidemiological assessment from India. Research in Developmental Disabilities, 34(9), 3004-3013.

[176] Katzen-Luchenta, J. (2007). The declaration of nutrition, health, and intelligence for the child-to-be. Nutrition and Health, 19 (1-2), 85-102.

[177] Bradley, E., & Finlay, F. (2009). Is cranio-sacral therapy useful in the management of crying babies? Commentary. Archives of Disease in Childhood, 94(7), 555-556.

[178] Betancur, C. (2011). Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. Brain Research, 1380, 42-77.

[179] Abd Elhameed, M. A., Abd Elbaky, A. E. O., & Kamel, E. A. (2011). A controlled study of the risk factors and clinical picture of children with Autism in an Egyptian sample. Egyptian Journal of Neurology, Psychiatry and Neurosurgery, 48(3), 271-276.

[180] Esposito, G., & Venuti, P. (2009). Comparative analysis of crying in children with autism, developmental delays, and typical development. Focus on Autism and Other Developmental Disabilities, 24(4), 240-247,

[181] Esposito, G., & Venuti, P. (2010). Understanding early communication signals in autism: A study of the perception of infants' cry. Journal of Intellectual Disability Research, 54(3), 216-223.

[182] Esposito, G., del Carmen Rostagno, M., Venuti, P., Haltigan, J. D., & Messinger, D. S. (2013). Brief Report: Atypical Expression of Distress During the Separation Phase of the Strange Situation Procedure in Infant Siblings at High Risk for ASD. Journal of Autism and Developmental Disorders, 1-6.

[183] Esposito, G., Nakazawa, J., Venuti, P., & Bornstein, M. H. (2012). Perceptions of distress in young children with autism compared to typically developing children: A cultural comparison between Japan and Italy. Research in Developmental Disabilities, 33(4), 1059-1067.

[184] Esposito, G., Nakazawa, J., Venuti, P., & Bornstein, M. H. (2013b). Componential deconstruction of infant distress vocalizations via tree-based models: A study of cry in autism spectrum disorder and typical development. Research in Developmental Disabilities, 34(9), 2717-2724.

[185] Esposito, G., Venuti, P., & Bornstein, M. H. (2011). Assessment of distress in young children: A comparison of autistic disorder, developmental delay, and typical development. Research in Autism Spectrum Disorders, 5(4), 1510-1516.

[186] Venuti, P., Caria, A., Esposito, G., De Pisapia, N., Bornstein, M. H., & de Falco, S. (2012). Differential brain responses to cries of infants with autistic disorder and typical development: An fMRI study. Research in Developmental Disabilities, 33(6), 2255-2264.

[187] American Psychological Association (2000). Diagnostic and Statistical Manual of Mental Disorders (Fourth ed.). Arlington, VA: American Psychiatric Publishing.

[188] Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). Autism diagnostic observation schedule-WPS (ADOS-WPS). Los Angeles, CA: Western Psychological Services.

[189] Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., et al. (1989). Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. J Autism Dev Disord, 19(2), 185-212.

[190] Rautava, L., Lempinen, A., Ojala, S., Parkkola, R., Rikalainen, H., Lapinleimu, H., et al. (2007). Acoustic quality of cry in very-low-birth-weight infants at the age of 1 1/2 years. Early Hum Dev, 83(1), 5-12.

[191] Lin, H., & Green, J. (2007). Effects of posture on newborn crying. Infancy, 11, 175-189.

[192] Fox, J. (1997). Applied regression analysis, linear models and related methods. Thousand Oaks, CA: Sage.

[193] Mullen, E. M. (1995). Mullen scales of early learning. Circle Pines, MN: American Guidance Service.

[194] Ainsworth, T., Jr. (1978). Ainsworth describes long term care audit. Hosp Peer Rev, 3(2), 17, 25-16.

[195]Berument, S. K., Rutter, M., Lord, C., Pickles, A., , & Bailey, A. (1999). Autism screening questionnaire: Diagnostic validity. British Journal of Psychiatry, 175, 444–451.

[196] Hollingshead, A. A. (1975). Four-factor index of social status. Unpublished manuscript, Yale University, New Haven, CT.

[197] Nao Kanemaru, Hama Watanabe, Hideki Kihara, HisakoNakano, Tomohiko Nakamura, Junji Nakano, Gentaro Taga, Yukuo Konishi, “Jerky spontaneous movements at term age in preterm infants who later developed cerebral palsy”, Early Human Development vol. 90, pp.387- 392, 2014

[198] Kanemaru N, Watanabe H, Taga G Increasing selectivity of inter-limb coordination during spontaneous movements in 2- to 4-month-old infants. Experimental Brain Research, vol. 21, no.1, pp. 49-61, 2012

[199] Karch D, Kang K, Wochner K, Philippi H, Hadders-Algra M, Pietz J, Dickhaus H, Kinematic assessment of stereotypy in spontaneous movements in infants, Gait & Posture. 36(2):307-11, 2012

[200] Dogra D P, Majumdar A K , Sural S, Mukherjee J , Mukherjee S, Singh A, "Analysis of adductors angle measurement in Hammersmith infant neurological examinations using mean shift segmentation and feature point based object tracking." Computers in Biology & Medicine, vol. 42, no. 9, pp. 25-34, 2012

[201] Berge P.R, Adde L, Espinosa G, Stavdahl O, "ENIGMA - Enhanced interactive general movement assessment", Expert Systems with Applications, vol. 34, no. 4, pp. 2664-2672, 2008

[202] Ohgi S, Morita S, Loo K.K, Mizuike C, "Time series analysis of spontaneous upper extremity movements of premature infants with brain injuries.", Physical Therapy, vol. 88, no. 9, pp.1022-1033, 2008

[203] Van Der Heide J.C, Paolicelli P.B, Boldrini A, Cioni G, "Kinematic and qualitative analysis of lower-extremity movements in preterm infants with brain lesions", Physical Therapy, vol.79, no. 6, pp. 546-557, 1999.

[204] Gotman J, "Automatic Detection of Seizures and Spikes", Journal of Clinical Neurophysiology Inpatient/Outpatient EEG Monitoring. Vol.16, no.2, pp.130-140, 1999

[205] Adde L, Helbostad J.L, Jensenius A.R, Taraldsen G, Grunewald K.H, Stoen R, "Early prediction of cerebral palsy by computer-based video analysis of general movements: a feasibility study", Developmental Medicine and Child Neurology , vol. 52, no. 8, pp. 773-778, 2010

[206] Stahl A, Schellewald C, Stavdahl O, Aamo O.M, Adde L, Kirkerod H, "An optical flow based method to predict infantile cerebral palsy", IEEE Trans Neural Syst Rehabil Eng, vol. 20, no.4, pp. 605-616, 2012

[207] Rahmati H., Aamo O.M., Stavdahl O., Adde L., Kernel-based object tracking for cerebral palsy detection, Proceedings of the 2012 International Conference on Image Processing, Computer Vision, and Pattern Recognition, IPCV 2012

[208] Adde L., Helbostad J., Jensenius A.R., Langaas M., Stoen R. "Identification of fidgety movements and prediction of CP by the use of computer-based video analysis is more accurate when based on two video recordings." , Physiotherapy Theory and Practice, vol. 29, no. 6, pp. 469-475, 2013

[209] Rahmati H, Aamo O.M, Stavdahl O, Dragon R, Adde L, Video-Based Early Cerebral Palsy Prediction Using Motion Segmentation, Proceeding of German Conference of Computer Vision and Pattern Recongnition (GCPR), September 2014

[210] Karayiannis N.B., Srinivasan S., Bhattacharya R., Wise M.S., Frost Jr. J.D., Mizrahi E.M, "Extraction of motion strength and motor activity signals from video recordings of neonatal seizures", IEEE Transactions on Medical Imaging, vol. 20, no. 9, pp. 965-980, 2001.

[211] Karayiannis NB, Varughese B, Tao G, Frost JD Jr, Wise MS, Mizrahi EM, "Quantifying Motion in video recordings of Neonatals Seizures by Regularized Optical Flow Methods", IEEE Trans Image Process, vol.14, no. 7, pp. 890-903, 2005;

[212] Karayiannis N.B, Xiong Y, Frost Jr. J.D, Wise M.S, Mizrahi E.M, "Quantifying motion in video recordings of neonatal seizures by robust motion trackers based on block motion models", IEEE Transactions on Biomedical Engineering, vol. 52, no. 6, pp. 1065-1077, 2005.

[213] Karayiannis N.B, Xiong Y, Frost Jr. J.D, Wise M.S, Mizrahi E.M, "Computerized motion analysis of videotaped neonatal seizures of epileptic origin", Epilepsia, vol. 46, no.6, pp. 901-917, 2005

[214] Karayiannis N.B, Tao G, Xiong Y, Sami A, Varughese B, Frost Jr. J.D, Wise M.S, Mizrahi E.M: Improving the accuracy and reliability of motion tracking methods used for extracting temporal motor activity signals from video recordings of neonatal seizures. IEEE Trans. Biomed. Engineering 52(4): 747-749, 2005

[215] Karayiannis N.B, Sami A, Frost Jr. J.D, Wise M.S, Mizrahi E.M, "Automated extraction of temporal motor activity signals from video recordings of neonatal seizures based on adaptive block matching", IEEE Transactions on Biomedical Engineering, vol. 52, no. 4, pp. 676-686, 2005

[216] Karayiannis N.B, Tao G, Frost Jr. J.D, Wise M.S, Hrachovy R.A, Mizrahi E.M, "Automated detection of videotaped neonatal seizures based on motion segmentation methods", Clinical Neurophysiology, vol. 117, no. 7, pp. 1582-1594, 2006

[217] Karayiannis N.B, Xiong Y, Tao G, Frost Jr. J.D, Wise M.S, Hrachovy R.A, Mizrahi E.M, "Automated detection of videotaped neonatal seizures of epileptic origin", Epilepsia, vol. 46, no.6, pp. 966-980, 2006

[218] Karayiannis N.B, Xiong Y, Frost Jr. J.D, Wise M.S, Hrachovy R.A, Mizrahi E.M, "Automated detection of videotaped neonatal seizures based on motion tracking methods", Journal of Clinical Neurophysiology, vol.23, no. 6, pp. 521 -531, 2006

[219] Osawa Y, Shima K, Bu N, Tsuji T, Tsuji T, Ishii I, Matsuda H, Orito K, Ikeda T, Noda S, "A Motion-based System to Evaluate Infant Movements Using Real-time Video Analysis", IFMBE Proceedings, vol.23, pp. 2043-2047, 2009

[220] Ntonfo K.G.M, Ferrari G, Raheli R, Pisani F, "Low-complexity image processing for real-time detection of neonatal clonic seizures", IEEE Transactions on Information Technology in Biomedicine,vol.16, no. 3,pp. 375-382, 2012

[221] Lu H, Pan Y, Mandal B, Eng H.L, Guan C, Chan D.W.S. "Quantifying Limb Movements in Epileptic Seizures Through Color-Based Video Analysis", IEEE Trans Biomed Eng. vol.60, no. 2, pp. 461-469, 2013

[222] Adde L, Rygg M, Lossius K, Oberg GK, Støen R, General movement assessment: predicting cerebral palsy in clinical practice, Early Human Development, vol. 83, no. 1, pp.13-8, 2007

[223] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H.Bischof, "Anisotropic Huber-L1 optical flow," in Proceedings of British Machine Vision Conference (BMVC), London, U.K., 2009, pp. 108.1 –108.11

[224] Meinecke L, Breitbach-Faller A, Bartz C, Damen R, Rau G, Disselhorst-Klug C, "Movement analysis in the early detection of newborns at risk for developing spasticity due to infantile cerebral palsy", Human Movement Science No. 25, pp. 125–144, 2006

[225] Zwaigenbaum, L., Bryson, S., Lord, C., Rogers, S., Carter, A., Carver, L., ... & Yirmiya, N. (2009). Clinical assessment and management of toddlers with suspected autism spectrum disorder: insights from studies of high-risk infants.Pediatrics, 123(5), 1383-1391.

[226] Zwaigenbaum, L., Thurm, A., Stone, W., Baranek, G., Bryson, S., Iverson, J., ... & Sigman, M. (2007). Studying the emergence of autism spectrum disorders in high-risk infants: methodological and practical issues. Journal of autism and developmental disorders, 37(3), 466-480.

[227] L. Bocchi, S. Orlandi, C. Manfredi, M. Puopolo, A. Guzzetta, S. Vicari and M.L. Scattoni, Early Diagnosis of Autism Spectrum Disorders – Design of the Data Acquisition and Management System, 5th European Conference of IFMBE, 14-18 September 2011, Budapest.

[228] S. Orlandi, L. Bocchi, C. Manfredi, M. Puopolo, A. Guzzetta, S.Vicari and M.L. Scattoni, Study of cry patterns in infants at high risk for autism, 7th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications. August 25-27, 2011, Firenze, Italy. Proceedings: 7-10

[229] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man, and Cybernetics 9 (1979) 62–66.

[230] P.-S. Liao, T.S. Chen, P.-C. Chung, A fast algorithm for multilevel thresholding, Journal ofInformation Science and Engineering 17 (5) (2001) 713–727.

[231] H. Chen, A de-noising method for heart sound signal using Otsu's threshold selection, in: IET Int. Communication Conf. on Wireless Mobile and Computing, 14–16 November, 2011, pp. 65–69.

[232] J. Van Santen, M. D' Imperio, Positional Effects on Stressed Vowel Duration in Standard Italian, Bell Labs – Lucent Technologies, USA/Department of Linguistics, The Ohio State University, USA, 1999.

[233] G. Varallyay Jr., Z. Benyo, A. Illenyi, Z. Farkas, L. Kovacs, Acoustic analysis of the infant cry: classical and new methods, in: Proc. 26th IEEE Annual Int. Conf., Vol. 1, 1–5 September, 2004, pp. 313–316.

[234] J.D. Markel, A.H. Gray, Linear Prediction ofSpeech, Springer-Verlag, Berlin, DE, 1982.

[235] M. A. Ruíz Díaz, C. A. Reyes García, L. C. Altamirano Robles, J. E. Xalteno Altamirano, A. Verduzco Mendoza, Automatic infant cry analysis for the identification of qualitative features to help opportune diagnosis, Biomed. Signal Proces. 7 (2012) 43-49.

[236] M. Robb, H. Sinton-White, R. Kaipa , Acoustic estimates of respiration in the pain cries of newborns, Int J Pediatr Otorhinolaryngol. 75-10, (2011) 1265-70.

[237] C. Manfredi, Adaptive Noise Energy Estimation in Pathological Speech Signals, IEEE Trans. Biomed. Eng., vol. 47, p.1538-1542, 2000.

[238] Saraswathy J, Hariharan M, Nadarajaw T, Khairunizam W, Yaacob S. Optimal selection of mother wavelet for accurate infant cry classification. Australas Phys Eng Sci Med 2014; 37(2):439-456.

[239] Saraswathy J, Hariharan M, Vijean V, Yaacob S, Khairunizam W. Performance comparison of daubechies wavelet family in Infant cry classification. In Signal Processing and its Applications (CSPA), 2012 IEEE 8th International Colloquium on; 2012: 451-455.

[240] Hariharan M, Yaacob S, Awang SA. Pathological infant cry analysis using wavelet packet transform and probabilistic neural network. Expert Systems with Applications 2011; 38(12):15377-15382.

[241] Cnockaert L, Schoentgen J, Auzou P, Ozsancak C, Defebvre L, Grenez F. Low-frequency vocal modulations in vowels produced by Parkinsonian subjects. Speech Commun 2008; 50(4): 288-300.

[242] Falek L, Amrouche A, Fergani L, Teffahi H, Djeradi A. Formantic Analysis of Speech Signal by Wavelet Transform. Conf. Proc- of the World Congress on Engineering 2011; 2:1572-1576

[243] Kronland Martinet R, Morlet J and Grossman A. Analysis of sound patterns through wavelet transforms. Int J Pattern Recogn 1987; 1(2): 273-302.

[244] Chui C. An introduction to wavelets, Academic Press,1995.

[245] Addison PS. The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance. Institute of Physics Publishing, London 2002.

[246] Daubechies I. Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics, 1992

[247] Tukey JW. Exploratory data analysis. Reading, MA: Addison-Wesley Publishing Co., 1977:44.

[248] High, R. Dealing with outliers: How to maintain your data's integrity. University of Oregon, 2000, available at http://cc.uoregon.edu/spring2000/outliers.html

[249] I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2 ed. San Francisco: Morgan Kaufmann, 2005.

[250] E. Thelen, Three-month-old infants can learn task-specific patterns of interlimb coordination, Psychological Science, 5-5 (1994) 280-285.

[251] Y. Nakajima, C. Einspieler, P.B. Marschik, A.F. Bos and H. F. R. Prechtl, Does a detailed assessment of poor repertoire general movements help to identify those infants who will develop normally?, Early Hum. Dev. 82-1 (2006) 53 -59.

[252] L.M. Van Immerseel, J.-P. Martens, Pitch and voiced/unvoiced determination with an auditory model, Journal of the Acoustical Society of America 91 (6) (1992) 3512–3526.

LIST of PUBLICATION

The following is a list of works published by the author during the course of the doctorate.

Peer-reviewed international journals

[J1] Orlandi S, Guzzetta A, Bandini A, Belmonti V, Barbagallo SD, Tealdi G, Mazzotti S, Scattoni ML and Manfredi C, AVIM - a Contactless System for Data Acquisition and Analysis: Software Architecture and First Results. Biomedical Signal Processing and Control (accepted)

[J2] Alejandro Rosales-Perez, Carlos A. Reyes-Garcia, Jesus A. Gonzalez, Orion F. Reyes-Galaviz, Hugo Jair Escalante and Silvia Orlandi, Classifying Infant Cry Patterns by the Genetic Selection of a Fuzzy Model, Biomedical Signal Processing and Control, 2014 doi:10.1016/j.bspc.2014.10.002

[J3] Rruqja, N; Dejonckere, PH; Cantarella, G; Schoentgen, J; Orlandi, S; Barbagallo, SD; Manfredi, C,Testing software tools with synthesized deviant voices for medicolegal assessment of occupational dysphonia,Biomedical Signal Processing and Control, vol.13, pp. 71-78, 2014

[J4] Orlandi S., Dejonckere P., Schoentgen J., Lebacq J., Rruqja N., & Manfredi C., Effective preprocessing of long term noisy audio recordings: An aid to clinical monitoring. Biomedical Signal Processing and Control, vol. 8 n.6, pp. 799-810, 2013

[J5] Orlandi S, Bocchi L, Donzelli GP, Manfredi C., Central blood oxygen saturation vs crying in preterm newborns. Biomed Sig Proc Control, vol. 7, pp. 88-92, 2012.

[J6] Orlandi S, Bandini A, Caratelli D, Manzo S, Schoentgen J, Manfredi C, An efficient wavelet-based method for fundamental frequency estimation: a pilot study on synthesized and real voice signals. Submitted to Frontiers in Genetics (accepted)

[J7] Manfredi C, Barbagallo SD, Baracca G, Orlandi S, Bandini A, and Dejonckere PH, Automatic Assessment of Acoustic Parameters of the Singing Voice: Application to Professional Western Operatic and Jazz Singers, Journal of Voice 2014. doi: 10.1016/j.jvoice.2014.09.014

[J8] Bandini A, Giovannelli F, Orlandi S, Barbagallo SD, Cincotta M, Vanni, P, Chiaramonti R, Borgheresi A, Zaccara G, Manfredi C, Automatic Identification of dysprosody in idiopathic Parkinson's disease. Biomedical Signal Processing and Control, 2014. On-line: http://dx.doi.org/10.1016/j.bspc.2014.07.006

[J9] Dejonckere PH, Lebacq J, Bocchi L, Orlandi S, Manfredi C, Automated tracking of quantitative parameters from single line scanning of vocal folds: A case study of the 'messa di voce' exercise. Logopedics Phoniatrics Vocology, pp.1-11, 2014. doi:10.3109/14015439.2013.861014

[J10] Ruggiano P., Larucci L., Deodati R., Giabbiani R., Bocchi L. and Orlandi S., Three-dimensional recostruction of the spatial distribution of anaesthetic during locoregional anaesthesia: 8AP2-4, European Journal of Anaesthesiology (EJA), vol. 30, pp. 120-120, 2013

[J11] Bandini A, Orlandi S, Manfredi C, Evangelisti A, Barrella M, Bevilacqua M, Bocchi L, Modelling of Thermal Hyperemia in the Skin of Type 2 Diabetic Patients. Journal of healthcare engineering, vol. 4, n. 4, pp. 541-554, 2013

[J12] Bandini A, Orlandi S, Manfredi C, Evangelisti A, Barrella M, Bevilacqua M, Bocchi L, Effect of local blood flow in thermal regulation in diabetic patient. Microvasc Res, vol. 88, pp. 42-47, 2013.

International conferences proceedings

[C1] S. Orlandi, Claudia Manfredi, A. Bandini and M.L. Scattoni, Automatic analysis of spontaneous infant cry for early diagnosis of autism spectrum disorders, XII Annual Pacific Voice Conference, Krakov, 11-13 April, 2014

[C2] S. Orlandi, A. Bandini, S.D. Barbagallo, C. Manfredi, Automatic classification of newborn cry melody, XII Annual Pacific Voice Conference, Krakov, 11-13 April, 2014

[C3] Andrea Bandini, F. Giovannelli, S. Orlandi, S.D. Barbagallo, M. Cincotta, P. Vanni, R. Chiaramonti, A. Borgheresi, G. Zaccara, C. Manfredi, Acoustic and kinematic analysis of speech in idiopathic Parkinson's disease, XII Annual Pacific Voice Conference, Krakov, 11-13 April, 2014

[C4] S. Orlandi, C. Manfredi, A. Guzzetta, V. Belmonti, S.D. Barbagallo and M.L. Scattoni, Advanced Tools for Clinical Diagnosis of Autism Spectrum Disorders, IFMBE-ICHI2013 Vilamoura, Portugal, 7-9 November, 2013

[C5] S. Orlandi, C. Manfredi, A. Guzzetta and M.L. Scattoni, Early diagnosis in autism spectrum disorders suggestions by animal models, MAVEBA 8th International Workshop , Firenze, Italy on 16-18 December 2013.

[C6] S.D Barbagallo, S. Orlandi, C. Manfredi, A new tool for audio and video analysis an aid to contact-less clinical diagnosis in newborns, MAVEBA 8th International Workshop, Firenze, Italy on 16-18 December 2013.

[C7] C. Manfredi, S. Orlandi, N. Rruqja, M.L. Scattoni and P.H. Dejonckere, An Innovative MultiPurpose Portable Voice Laboratory,10th International Advances in Quantitative Laryngology, Voice and Speech Research Conference, Cincinnati, Ohio, USA. 3-4 June 2013.
[C8] S. Orlandi, L. Bocchi, M.L. Scattoni and C. Manfredi, Early Detection of Autism Spectrum Disorders from Newborn Cry Acoustic Parameters, The Voice Foundation, 42nd Annual Symposium: Care of the Professional Voice, Philadelphia, Pennsylvania, USA, 29 May – 2 June 2013.

[C9] S. Orlandi, L. Bocchi and C. Manfredi, A General-Purpose Fast-Processing Method for Long Tern Audio Signals: Application to Newborn and Adult Recordings, The Voice Foundation, 42nd Annual Symposium: Care of the Professional Voice, Philadelphia, Pennsylvania, USA, 29 May – 2 June 2013.

[C10] S. Orlandi, C. Manfredi, L. Bocchi and M.L. Scattoni, Automatic newborn cry analysis: a noninvasive tool to help autism early diagnosis, 34th Annual International Conference of the IEEEEMBC, San Diego, California, USA, 28 August – 1 September 2012.

[C11] S. Orlandi, C. Manfredi, L. Bocchi and M.L. Scattoni, Automatic fundamental frequency estimation in early diagnosis of autism spectrum disorders, Conf. of ICVPB 2012, Erlangen, Germany., 5-7 July 2012.

National conferences proceedings

[NC1] S. Orlandi, C. Manfredi, A. Guzzetta, A. Bandini and M. L. Scattoni, An integrated system for the automatic management of audio and video recordings in newborns: first results on normative data, IV National Congress of Bioengineering (GNB2014) Pavia, June 25-27, 2014

[NC2] S. Orlandi, C. Manfredi, A. Guzzetta, G. Valeri, F. Apicella, G. Tealdi, C. Michetti, A. Caruso, S.D. Barbagallo, F. Perone, A. Bandini, F. Muratori, S. Vicari and M. L. Scattoni, Non-Invasive Tools for Early Diagnosis of Autism Spectrum Disorders, Percezione-azione e apprendimento in età evolutiva (PAED), Campus Biomedico Roma, November 22, 2013

[NC3] S. Orlandi, C. Manfredi, L. Bocchi and M.L. Scattoni, A new approach for early diagnosis of autism spectrum, Third National Congress of Bioengineering (GNB2012) Roma, June 26-29, 2012

[NC4] A. Bandini, F. Giovannelli, G. Zaccara, M. Cincotta, P. Vanni, R. Chiaramonti, A. Borgheresi, S. Orlandi, A. Bandini, Acoustic and kinematic measure of speech in idiopathic Parkinson’s disease by means of contact-less techniques, IV National Congress of Bioengineering (GNB2014) Pavia, June 25-27, 2014

[NC5] A. Bandini, F. Giovannelli, M. Cincotta, P. Vanni, R. Chiaramonti, A. Borgheresi, G. Zaccara, S.Orlandi, C. Manfredi, Automatic detection of dysprosody patterns in patients with idiopathic Parkinson’s disease, 59 th National Congress of SINC (Società Italiana di Neurofisiologia Clinica), Milano, Italy, May 14 -17, 2014.

[NC6] S. Orlandi, L.Bocchi and C.Manfredi, Voice quality parameters: application to vocal fatigue in singers, Third National Congress of Bioengineering (GNB2012) Roma, June 26-29, 2012 (poster presentation)

[NC7] A. Cosentino, S. Orlandi, C. Manfredi, M. Magherini, J. Nori and L. Bocchi, An user friendly interface for semiautomated classification of breast ultrasound video clips, Third National Congress of Bioengineering (GNB2012) Roma, June 26-29, 2012

ACKNOWLEDGEMENTS

This work was supported by the Italian Ministry of Health Grant (GR3), Young Researcher 2008, “Non-invasive tools for early detection of Autism Spectrum Disorders” and Project ECRF 2009.0772 "Interdisciplinary Laboratory of Biomedical Acoustics".