Open access • Journal Article • DOI:10.1088/1361-6579/AB0096

Video and audio processing in paediatrics: a review. — Source link

Sandie Cabon, Fabienne Poree, Antoine Simon, O. Rosec ...+2 more authors

Institutions: University of Rennes

Published on: 26 Feb 2019 - Physiological Measurement (IOP Publishing)

Topics: Video processing and Audio signal processing

Video and audio processing in paediatrics a review

Sandie Cabon, Fabienne Porée, Antoine Simon, Olivier Rosec, Patrick Pladys, Guy Carrault

Video and audio processing in paediatrics: a review

S Cabon[1,2], F Porée[1], A Simon[1], O Rosec[2], P Pladys[1] and G Carrault[1]

[1]Univ Rennes, CHU Rennes, INSERM, LTSI - UMR 1099, F-35000 Rennes, France
[2]Voxygen, F-22560 Pleumeur-Bodou, France

E-mail: fabienne.poree@univ-rennes1.fr

Dec 2018

Abstract. Video and sound acquisition and processing technologies have known great improvements in the last decades, with many applications in the biomedical area. The motivation of this paper is to provide an overall state of the art of the advances within these topics in paediatrics, to evaluate their potential application for monitoring in Neonatal Intensive Care Unit (NICU). For this purpose, more than 150 papers dealing with video and audio processing were reviewed. For both topics, clinical applications are described according to the considered cohorts, either full-term newborns, infants and toddlers, or preterm newborns. Then, processing methods are presented, in terms of data acquisition, feature extraction and characterization. The paper is firstly focused on the exploitation of video recordings, that began to be automatically processed in the 2000s, and we show that it has been mainly used to characterize infant motion. Others applications, including respiration and heart rate estimation and face analysis, are also presented. Audio processing is then reviewed with a focus on cry analyses. We describe that this topic arose earlier, from first studies focused on induced-pain cries to newest ones dealing with spontaneous cries, and that they are mainly based on frequency features. Then, some papers dealing with non-cry signals are also discussed. Finally, we show that, even if recent improvements in digital video and signal processing allow for an increasing automation of processing, the context of NICU makes still difficult a fully-automated analysis of long recordings. Propositions to overcome some limitations are given.

1. Background

The analysis of neonatal and early childhood development is at the center of concerns of the medical community. Especially, premature babies, having several vital immature functions, receive a particular attention by the recording, in Neonatal Intensive Care Unit (NICU), of several physiological signals (Huvanandana et al., 2017), however limited by their fragility. On the other hand, video and audio acquisitions have the advantage to propose contactless and non-invasive ways to collect data for patients being cared for in hospital or at home.

Such technologies having known great improvements in the last decades, they are now used in many biomedical applications.

The use of audio and video in paediatrics found certainly its roots in sleep analysis when the Association of the Psychophysiological Study of Sleep (APSS) stated in 1969 the necessity to develop a guide for scoring sleep in infants, since the criteria of Rechtschaffen and Kales (Rechtschaffen and Kales, 1968) were "applicable only to the adult" and had "not taken into account the unique features of the developing infant" (Grigg-Damberger et al., 2007). In 1971, Anders *et al* published a manual recommending to support polysomnographic recordings by behavioral observations (Anders et al., 1971) and then carried out a study with full-term infants, where behavioral states were scored from video and audio acquisitions, according to eye state, movements and crying vocalizations (Anders and Sostek, 1976). Later on, several approaches using audio and video recordings were proposed to analyze sleep either on premature newborns (Fuller et al., 1978), or on children for the evaluation for Obstructive Sleep Apnea (OSA) (Morielli et al., 1996; Sivan et al., 1996). Finally in 2011, ASA/ASTA (Australasian Sleep Association/Australasian Sleep Technologists' Association) Paediatric Working Group recommended to record audio and video as additional information to the electrophysiological signals in the scoring of children sleep (Pamula et al., 2011).

Along with these works, in 1990, Prechtl developed a method to assess the quality of General Movement (GM) based on video observations as a diagnostic tool for early detection of brain dysfunction (Prechtl, 1990; Prechtl et al., 1997). From there, General Movement Assessment (GMA) using video has been applied in several clinical contexts. In the same way, the detection of neonatal seizures also led to a lot of works including video acquisitions (Pediaditis et al., 2012). Such analyses relied, in the oldest studies, on manual annotations of the videos that began to be automatically processed only in the 2000s, thanks to the improvements in digital video processing.

The development of automated sound processing occurred earlier since studies on newborn cries began from the 1960s with Wasz-Höckert *et al* , where it was shown, by spectrographic analysis, that four different types of cries could be distinguished as birth, pain, hunger, pleasure (see (Wasz-Höckert et al., 1985) for an historical review). From there, a huge literature arose, with the analysis of frequency features, in children and newborns. Several studies performed also detailed analyses of crying behavior in preterm newborns either solely, either in comparison with full-term newborns.

The objective of the present paper is to synthesize this abundant literature, in the context of paediatrics, and identify its clinical impacts. Specifically, the motivation of our work is to offer an overview of the existing audio and video processing methods to evaluate their potential application for monitoring in NICU. The paper is organized in two main sections: Section 2 is devoted to video processing, while audio analysis is described in Section 3. For both topics, clinical applications are first described. Since these applications differ depending on the studied population age, especially considering audio analysis, they

are presented first for full-term newborns (0 to 2 months old), infants (2 months to 1 year old) and toddlers (1 to 4 years old) and then for preterm newborns. Then, processing methods are presented, in terms of data acquisition, feature extraction and characterization. Finally, last section draws the main limitations of these studies but also gives some propositions, in the objective of developing automatic monitoring systems able to meet clinical needs in NICU.

2. Video analysis

Since the motion of a newborn is one of the most crucial information to describe his physio-pathological state, it has been the most extracted descriptor from video recordings. The estimation of other types of information has also been investigated such as respiration, heart rate and facial expression.

In this section, main clinical applications supported by video recordings are first presented. However, since, for some applications, video recordings were analyzed manually (especially for preterm), some studies with manual video analysis are included in order to present all the potential applications of video analysis. Then, data acquisition and processing methods are described.

2.1. Clinical applications

2.1.1. Full-term newborns, infants and toddlers General Movement Assessment was used to support many studies regarding different pathologies. In (Guzzetta et al., 2003), the visual assessment of motion patterns from video enabled to early detect hemiplegia (complete or partial palsy) in infants with cerebral infarction (not enough blood supply in a region of the brain). Mazzone *et al* found that abnormal GMs were early markers of motor impairment (partial or total loss of function of a body part) in infants with Down Syndrome (Mazzone et al., 2004). Nearly ten years after, automatic video processing was applied by different groups to study Cerebral Palsy (CP), which is a disorder of movements caused by an abnormal motor control center of the brain (Stahl et al., 2012; Orlandi et al., 2015).

A large part of studies using video recordings was dedicated to the analysis of neonatal seizures. Whereas first studies focused on observational classifications of seizures from video recordings (Mizrahi and Kellaway, 1987; Tharp, 2002), later, thanks to the development of video processing, several approaches were proposed to automatically detect or classify seizures from motion descriptors (Karayiannis et al., 2001; Sami et al., 2004; Karayiannis et al., 2004, 2005a,b,c,d; Karayiannis and Tao, 2006; Karayiannis et al., 2006; Ntonfo et al., 2012). For their part, Cuppens *et al* focused on the specific case of the epilepsy (Cuppens et al., 2009, 2010).

Studies regarding physiological monitoring with the support of video processing have also been conducted. Respiratory frequency has been estimated in order to prevent Sudden

Infant Death Syndrome (SIDS) (Fang et al., 2015) or to detect repeated apnea events in the context of Congenital Central Hypoventilation Syndrome (CCHS) (Cattani et al., 2014, 2017). Pulse rate has also been estimated from video acquired during Hammersmith Infant Neurological Examinations (HINE) (Sikdar et al., 2015).

Emotion and facial expression detection also received a particular attention. Face analyses were performed either to discriminate, between the behavioral states sleep, awake and cry (Hazelhoff et al., 2009), or to automatically analyze infants emotion during interactions (Zaker et al., 2013, 2014).

2.1.2. Preterm newborns Like for the previous population, video has been mainly used for General Movement Assessment regarding preterm infants. Visual general movement assessment on video recording has been proved successful to determine if infants had brain dysfunctions, either transient or persistent, and to identify infants at risk for impaired neurological outcomes (Bos et al., 1998a). Later, Bos *et al* also observed in videotape recordings the effects of the dexamethasone therapy on high-risk very preterm infant through GMA (Bos et al., 1998b). Moreover, Spittle *et al* showed that abnormal GMs directly reflect white matter injury (Spittle et al., 2008). Among these applications, no video processing methods were developed to automatize the assessment. More recently, Adde *et al* focused on the automatic prediction of CP by computer-based video analysis of general movements (Adde et al., 2009, 2010; Rahmati et al., 2014, 2015).

In the meantime, video-based analyses were performed to assess sleep quality, either visually, where video was used as a support for actigraphy measurement validation (So et al., 2005; Sung et al., 2009) or to evaluate the influence of light exposure on the sleep (Kaneshi et al., 2016), or semi-automatically, to detect the eye state in different sleep stages (Porée et al., 2015).

Other clinical investigations through video processing have been conducted. Among them, Pulled-To-Sit examination of infants was performed by the use of object tracking methods (Dogra et al., 2012). Later, with the ambition of a contact-less and non-invasive monitoring, Villarroel *et al* elaborated a monitoring solution of vital signs (respiratory rate, oxygen saturation and heart beat) through video analysis (Villarroel et al., 2014). Some authors also succeeded to estimate the infants heart rate from video imaging (Aarts et al., 2013; van Gastel et al., 2018) and Koolen *et al* estimated respiration rate by motion extraction (Koolen et al., 2015). Meanwhile, Zamzami *et al* worked on pain assessment by classifying infant's expression (Zamzami et al., 2015).

2.2. Methods for video processing

Automatic video processing in paediatrics has led to a large number of papers, with different objectives, including motion extraction and characterization, respiration and heart rate

estimation and face analysis such as depicted, in Figure 1. The next section presents these works. As a first step, a special attention is paid to video databases.

2.2.1. Video acquisition Recording settings are important quality factors and yet, only one group observed the impact of the camera set-up on motion analysis, including spatial and temporal resolutions, compression, illumination and location of moving body parts in the images (Cuppens et al., 2009, 2010). Nonetheless, all authors usually respected reasonable values in term of resolution, frame rate or compression and solutions to overcome the impact of the illumination and location variability were developed (noise reduction techniques, computation of features independent of the amplitude...).

Two database types can be identified regarding camera/patient positioning. The more common one contained video recordings performed in a controlled environment where cameras were located above a mattress and infants were placed in a supine position with no blanket and fully visible (Karayiannis et al., 2001; Sami et al., 2004; Karayiannis et al., 2004, 2005a,b,c,d; Karayiannis and Tao, 2006; Karayiannis et al., 2006; Adde et al., 2009, 2010; Stahl et al., 2012; Rahmati et al., 2014, 2015; Orlandi et al., 2015). A marginal positioning has also been used in (Zaker et al., 2013, 2014) where infants were placed in a baby seat with only the head and shoulders visible. In the second type, video data collection was integrated in the medical care routine during scheduled examinations such as HINE (Dogra et al., 2012; Sikdar et al., 2015) or directly in the bedroom (Hazelhoff et al., 2009; Cuppens et al., 2010; Ntonfo et al., 2012; Aarts et al., 2013; Villarroel et al., 2014; Cattani et al., 2014; Porée et al., 2015; Fang et al., 2015; Koolen et al., 2015; Cattani et al., 2017; van Gastel et al., 2018). This ambition to integrate the care routine has led to a higher variability in camera positioning. Nevertheless, cameras were mostly placed above one corner of the foot-end of the bed, recording the full body of the baby (Cuppens et al., 2010; Ntonfo et al., 2012; Porée et al., 2015; Cattani et al., 2014; Koolen et al., 2015; Cattani et al., 2017).

For studies focused on face analysis or heart rate, camera position may differ and be on the middle-left of the bed (Aarts et al., 2013). Zoom has been applied to focus on the head (Hazelhoff et al., 2009; Aarts et al., 2013; Zamzami et al., 2015). Dealing with closed incubators, Villarroel *et al* installed their camera above the infant, against the plastic wall (Villarroel et al., 2014). Profile point of view was adopted for the recordings performed during HINE (Dogra et al., 2012; Sikdar et al., 2015).

Most of the databases were composed by selected short video sequences from a dozen of seconds to dozens of minutes. Some exceptions were noted in (Cattani et al., 2014) with more than one hour and half of recording and in the work on continuous vital signs monitoring with initial recordings durations between 50 minutes and more than seven hours (Villarroel et al., 2014).

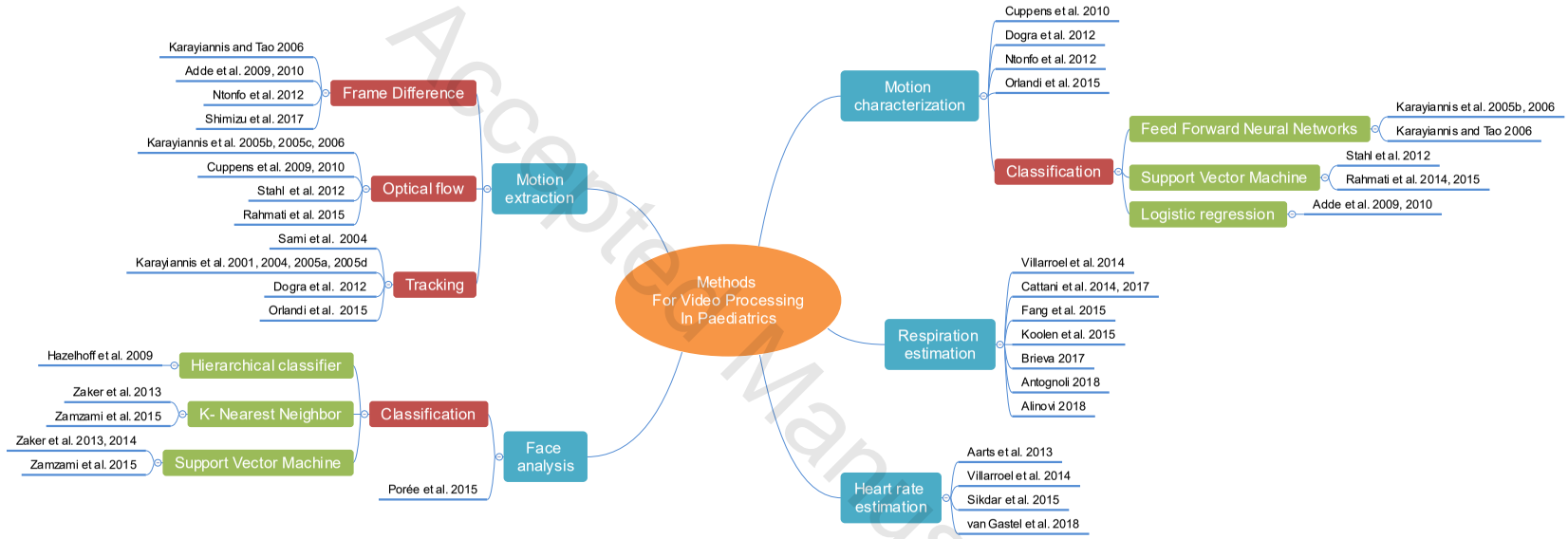

Figure 1. Overview of video processing methods in paediatrics.

The size of the studied population was very varied, from a very small number of infants (one to three) (Cattani et al., 2014; Orlandi et al., 2015) to a more consequent cohort (more than 30 patients) (Karayiannis and Tao, 2006; Adde et al., 2009).

2.2.2. Motion extraction Three popular methods have been used to automatically extract body (global) or limbs (local) movements: frame differencing, optical flow and tracking.

Frame differencing consists in computing the difference between the current and one or multiple previous frames. The resulting difference image contains the global motion information, including the infant's movements. However, this difference image is generally very noisy, since it is sensitive to very low intensity changes and to the original images noise. Moreover, since the detection is based on intensity changes, it enables to identify the contours of the moving parts, but is less sensitive to the motion of areas with homogeneous intensities. Hence, all the studies have considered post-processing to limit the influence of noise. Intensity thresholding or median filtering have been widely used to remove small intensity differences (Karayiannis and Tao, 2006; Ntonfo et al., 2012; Shimizu et al., 2017), followed by morphological operators or a threshold applied on objects size to remove the small regions (Adde et al., 2009, 2010). The use of a vector clustering was also proposed to identify the different regions of interest (Karayiannis and Tao, 2006). Finally, most authors summarized the resulting images by computing a motion signal, sometimes called "motiongram", "quantity of motion" or "motion strength signal", corresponding to the area of the moving parts (Karayiannis and Tao, 2006; Adde et al., 2009, 2010). This signal was used for further processing in order to characterize the motion (see section 2.2.3).

Optical flow is the velocity field generated by the relative motion between an object and the camera in a sequence of frames. Contrary to frame differencing which detects the moving areas but does not compute the displacements, optical flow estimation methods return a velocity or a displacement vector for each pixel. Many methods for optical flow have been proposed in the literature (Sun et al., 2014), most of them being based on the hypothesis that the intensity of one pixel remains constant between any two following frames and that the motion between two frames is small. Considering infant's motion extraction, the original method of Horn and Schunck (Horn and Schunck, 1981) has been used by different teams (Karayiannis et al., 2005b,c, 2006; Cuppens et al., 2009, 2010). This approach is based on the minimization of a metric combining two terms, a data term based on the intensity conservation constraint and a smoothness, or regularization, term modelling the spatial distribution of the flow field. The importance of formulating an appropriate smoothness term, e.g. using quadratic functions, and to tune the associated weighting factors has been demonstrated (Karayiannis et al., 2006). In the same sense, Stahl *et al* used a method based on a non-local regularization term, imposing a smoothness assumption within a specified region of the flow field (Stahl et al., 2012). In order to segment and track infant's individual body parts, Rahmati *et al* proposed, starting from some manual labels in one or more frames,

to use a multi-scale optical flow combined with a tracking of the segmented parts (Rahmati et al., 2015). As for the studies based on image differences, a temporal signal was finally extracted from the computed flow fields. It is generally based, at each frame, on the area of pixels corresponding to velocity vectors of magnitude higher than a threshold.

Tracking methods rely on the selection (either automatically or manually) of points or regions of interest and on their tracking over the image sequence. They rely on the assumption that the considered features keep a constant appearance over time and motion. This tracking is based either on local intensity differences (template matching methods) or on higher level descriptors (feature matching methods). Both approaches have been used to track either specific parts of the infants, generally selected manually (e.g. hands and feet (Orlandi et al., 2015) or center of the head (Dogra et al., 2012)) or more general points of interests (e.g. corners (Karayiannis et al., 2001)). Karayiannis *et al* especially evaluated different trackers based on block matching. Starting from the Kanade-Lucas-Tomasi (KLT) tracker (Karayiannis et al., 2001), they considered adaptive block matching (Karayiannis et al., 2005a), predictive block matching (Karayiannis et al., 2004) and a variety of other block models associated with the automatic localization of moving body parts (Sami et al., 2004). If these approaches had the advantage of estimating the motion of specific parts of the infants, and thus enabled to study precisely their motions, they were sensitive to occlusions, complex motions (e.g. rotations) and noise (Karayiannis et al., 2005d).

In short, the considered motion estimation methods have shown their ability to estimate the infant's motion. However, they most always relied on parameters which have to be tuned according to the data (e.g. noise, motion amplitude) and are sensitive to occlusions. Moreover, they generally do not differentiated the infant's motion from other people's (parents, medical staff) motion, thus needing some manual steps to identify regions of interest, making difficult the use of these approaches on long recordings.

2.2.3. Motion characterization Once the motion signal was extracted, high level features were computed to characterize movements, sometimes in order to automatically classify infants' impairments.

The first step was often to identify motion and non-motion periods (or epochs) from the motion strength signal. For this purpose, Cuppens *et al* used two methods: a) a fixed threshold determined thanks to a receiver operating characteristic curve and b) a variable threshold adapted to the noise and computed from mean and standard deviation of a selected non-movement period. They concluded that a variable threshold adapted to each video recording was the best approach with an average Predictive Positive Value (PPV) value of 94% (Cuppens et al., 2010). Considering a tracking of different limbs, Orlandi *et al* defined non-motion periods when all limbs values were lower than a fixed threshold (Orlandi et al., 2015).

Many features have been considered to describe the resulting epochs. Features

characterizing the distribution of epochs have been used. Mean, median, maximum, standard deviation of the quantity of motion have been computed (Adde et al., 2009, 2010), such as speed and acceleration, skewness of the velocity and kurtosis of the acceleration, the last two being measurements of the complexity, unintuitive repartition of the speed and acceleration values (Orlandi et al., 2015). The periodicity of the motion signal has also been investigated, e.g. using the autocorrelation function, to characterize clonic seizures (Ntonfo et al., 2012). The maximum spike duration and the number of spikes have been computed (Karayiannis et al., 2005b; Karayiannis and Tao, 2006). Positioning features of the centroid of motion (the spatial center of the positive pixels in the motion image) have also been considered (Adde et al., 2009, 2010): the mean position in the x- and y-directions, the standard deviation, velocity and acceleration. The motion of individual limbs has also been characterized, e.g. using wavelet and frequency analysis (Stahl et al., 2012), the periodicity (Rahmati et al., 2014, 2015), the correlation between trajectories (Rahmati et al., 2014, 2015; Orlandi et al., 2015), or the deviation from a smoothed version of the movement to characterize its smoothness (Rahmati et al., 2014). 2-D modeling of the camera scene had also been used to extract the angle between head and torso during HINE (Dogra et al., 2012).

The next step was to characterize or classify pathological situations based on these features. Pulled-to-sit scores have been computed by applying decision rules regarding the relative movement of the head with respect to the infant torso. The comparison with expert scores led to a sensitivity of 92% along a specificity of 96% (Dogra et al., 2012). Adde at al. aimed at determining infant with CP (Adde et al., 2009, 2010). By the use of logistic regression, they evaluated the capacity of each feature to classify the absence or presence of fidgety movements (classical circular movements, linked with the absence of CP). Best results were obtained with the standard deviation of the centroid of motion, with a sensitivity of 81.5% and specificity of 70% while other features showed weak specificities of less than 56% (Adde et al., 2009). Later, a cerebral palsy predictor was proposed, calculated as a combination of the previous descriptor with the mean and standard deviation of the quantity of motion. It reached a sensitivity of 85% and a specificity of 88% (Adde et al., 2010). These works were extended in (Stahl et al., 2012; Rahmati et al., 2014, 2015). Stahl *et al* classified infants using Support Vector Machine (SVM) with a linear kernel. They reported an accuracy of 93.7% along with a sensitivity of 85.3% (Stahl et al., 2012). Combined with an automatic segmentation of infant limb's (Rahmati et al., 2014, 2015), the same approach presented a total accuracy of 87% which was better than the results obtained with electromagnetic sensors (Rahmati et al., 2015). In (Karayiannis et al., 2005b; Karayiannis and Tao, 2006; Karayiannis et al., 2006), the goal was to distinguish neonatal seizures from random infant behaviors and to differentiate between myoclonic and focal clonic seizures. Multiple Feed Forward artificial Neuronal Networks (FFNN) were trained, using different sets of features. They reached 85% of specificity and sensibility, with an increase of 5% when a frequency feature was added (Karayiannis et al., 2006).

2.2.4. Respiration estimation Several groups aimed to characterize respiration from video recordings. Authors mostly proposed methods to estimate the frequency rate (Villarroel et al., 2014; Koolen et al., 2015; Fang et al., 2015; Alinovi et al., 2018; Antognoli et al., 2018; Brieva and Moya-Albor, 2017) and one group focused on the development of an apnea detector (Cattani et al., 2014, 2017).

Eulerian video magnification has been used to magnify the motion of low amplitude related to respiration (Koolen et al., 2015; Cattani et al., 2017; Brieva and Moya-Albor, 2017; Antognoli et al., 2018). It is based on the amplification of pixel color variation in a specified frequency band. This was followed by motion extraction, e.g. using frame differences (Cattani et al., 2014, 2017) or optical flow estimation (Koolen et al., 2015), to generate a motion signal which was further processed to characterize the periodicity of the motion. A periodic model whose parameters were estimated using a maximum-likelihood method (Cattani et al., 2014, 2017; Alinovi et al., 2018) or a short-time Fourier transform (Koolen et al., 2015) was used. If this process generally considered the whole image, one group performed the magnification in several ROIs before selecting the best ones, i.e. with the higher amplitudes of the estimated periodic variations (Alinovi et al., 2018). Respiratory rate detection was successful for most patients during quiet sleep stages (Koolen et al., 2015). Authors succeeded to identify 90 to 100% of the apneas detected by polysomnography (Cattani et al., 2014).

For their part, Fang *et al* revealed slight movements by the use of accumulative sum of difference images. Respiratory signals were then obtained from the frame by frame evolution of the intensity of automatically selected pixels in the accumulative sum. The method showed good results in 46 video sequences (Fang et al., 2015).

Villarroel *et al* based their method on four steps: a) automatic identification of stable periods (with no interaction between the infant and the medical staff) by the use of a non-parametric statistical background image estimation; b) computation of the mean intensity of Regions of Interest on each channel RGB; c) calculation of the reflectance photoplethysmogram by independent component analysis; d) identification of regular frequencies using auto regressive models e) estimation of vital signals by filtering the frequencies according to physiologically realistic ranges. Respiratory rate was obtained after the application of a band pass filter cutting of at 0.33 Hz and 1.67 Hz (i.e. between 20 and 100 breaths per minute) and oxygen saturation was calculated from respiratory signal. Respiratory rate and oxygen saturation estimation were comparable with the Philips monitor values (Villarroel et al., 2014).

2.2.5. Heart rate estimation Four studies proposed non-contact heart rate monitoring utilizing camera in preterm infant. These techniques were entirely contactless and used the principle known as pulse-oximetry. In fact, light is more absorbed by blood than by surrounding tissues so that variations in blood volume affect light transmission and

reflectance. This phenomenon can be measured by observing slight color variations in a region of interest.

First, a study showing interesting results to monitor heart rate in NICU was presented by Aarts *et al* in 2013 (Aarts et al., 2013). They proposed a method based on four steps: a) tracking of a manually selected contrasted ROI (e.g. eye, eye brow) that also contained a skin area; b) computation of the average green channel pixel values into the skin area ROI; c) computation of the joint time frequency diagrams (called plethysmograph); d) extraction of the dominant frequency, related to pulse rate. In all 19 infants, heart rates were estimated but not continuously. In fact, low ambient light level and infant motion still remained challenging conditions.

Later, as mentioned above (see section 2.2.4), Villarroel *et al* developed a non-contact vital signs monitoring, among them heart rate estimation based on the same steps as respiratory rate estimation, except that they used a different band pass filter between 1.3 Hz and 5 Hz (corresponding to 78 beats per minute and 300 beats per minute) (Villarroel et al., 2014). Resulting signals showed a great correlation with ECG-derived measurements from the Philips monitor.

Pulse rate was also extracted within a manually chosen region of the trunk of the infant by the mean of a set of different color (RGB) decomposition (Sikdar et al., 2015). Authors found that the RG and GB channel combinations were more accurate in comparison to the RGB or RB channel combination. This observation confirmed that photophlethysmography signal is strongest on the green channel.

It is important to notice that these methods, based on the analysis of slight color changes, were not adapted to acquisitions with low light levels. Recently, a method adapted to infrared illumination was thus proposed, showing a good estimation of heart rate 87% of the time (van Gastel et al., 2018).

2.2.6. Face analysis Infant's facial expression was analyzed to detect the presence of discomfort (Hazelhoff et al., 2009). Authors automatically segmented the face from the background by skin color modelling and localized the eye, eye-brow and mouth region thanks to shape assumptions. Then, they classified behavioral states by the use of a hierarchical classifier (Hazelhoff et al., 2009). However, the study showed weakness in eye-brow segmentation under lighting and viewpoint deviations.

Recent studies on facial analysis presented different features extraction methods (Zaker et al., 2013; Zamzami et al., 2015; Zaker et al., 2014; Porée et al., 2015). Zamzami *et al* detected the nose first and then expanded the mask to include the eyes and surrounding areas. A facial strain algorithm was applied on the remaining area in order to extract strain magnitude. K Nearest-Neighbors (KNN) and SVM classifier have been trained to discriminate expression between two states: pain and no-pain. KNN approach showed an higher accuracy of 96% (Zamzami et al., 2015). Facial features were also extracted by the

mean of Active Appearance Models (AAM) combined with Histogram of Oriented Gradients (HOG) to detect spontaneous facial Actions Units (AUs). Firstly, authors trained classifier for each AU and the best results were given by a SVM classifier with intra-class correlation values up to 0.73 (Zaker et al., 2013). Then, they improved the results by training classifiers to detect multi-AUs and reported F-scores between 0.58 and 0.91 (Zaker et al., 2014). Finally, Porée at al. proposed an algorithm estimating the eye state in videos of premature babies combining tracking with segmentation and characterization steps (Porée et al., 2015). The proposed approach gave more than 95% of concordance with a sensitivity and specificity, respectively ranging from 78.5% to 100.0% and from 97.69% to 100.0%.

3. Audio analysis

Acoustic analyses in paediatrics mostly relied on cry analysis. In fact, several studies have been conducted for the analysis of the cries of newborn and small infants, healthy or with various diseases, but also of premature newborns. Four research groups largely contributed on infant cry analyses: the group of Wasz-Höckert in Helsinki (Finland), the group of Lester in Providence (USA), the group of Manfredi in Firenze (Italy) and the group of Reyes-Garcia in Tonantzintla (Mexico). If first studies concerned induced pain cries, spontaneous cries have also been considered.

Following sections are organized as follows. First, we present a non-exhaustive list of studies realized in the analysis of infant cries according to their clinical context. Then, we focus on the methods of data acquisition and acoustic signal processing encountered in these studies. To a lesser extent, other sounds processing methods (pre-linguistic infant vocalizations, NICU alarms, EEG sonification and lung sound classification) have been proposed and are compiled at the end of this section.

3.1. Clinical applications

3.1.1. Full-term newborns, infants and toddlers Infant cries were largely studied for the differentiation between normal and pathological cries. In (Lester, 1976), authors compared the cries of typically developing infants and infants possibly suffering from central nervous system insult due to malnutrition. They showed that the cry of the malnourished infant had an initial longer sound, higher pitch, lower amplitude, more arrhythmia, and a longer latency to the next cry sound than the cry of the well-nourished infant. The similarity between the cry of the malnourished infant and the cry of the brain-damaged infant suggested that malnutrition might affect the regulatory function of the central nervous system. Similar conclusions were obtained by a computerized analysis in (Donzelli et al., 1994). It was shown that cries of healthy newborns with prenatal and perinatal complications have different acoustical properties than cries of low-complications newborns (Zeskind and Lester, 1978). Abnormalities were searched in the cries of newborns with multiple or severe problems during

the neonatal period, such as low birth weight, respiratory symptoms, jaundice, apnea, but also infants subsequently victims of presumed sudden infant death syndrome (Golub and Corwin, 1982). In the 2000s, normal and pathological cries began to be automatically labeled thanks to a wide variety of machine learning approaches in the context of deafness (Orozco-García and Reyes-García, 2003; Suaste-Rivas et al., 2004; Rosales-Pérez et al., 2015), hypoxia-based Central Nervous System (CNS) diseases (Ortiz et al., 2004), cleft palate (Lederman et al., 2008) and asphyxia (Suaste-Rivas et al., 2004; Hariharan et al., 2011; Rosales-Pérez et al., 2015).

Several studies relied on pain-induced cries. Fuller *et al* differentiated them from fussy and hunger cries, as well as non-cry cooing vocalizations (pre-linguistic vocalizations occurring around 3 months of age) in 30 infants ranging in age from 2 to 6 months (Fuller and Horii, 1986, 1988). This study was based on the amplitude of high-frequency components, the fundamental frequency and the spectral energy levels. Formants (vocal tract resonance frequencies) and tenseness were first added in (Fuller, 1991) and acoustic cry measures were correlated with four pain levels and four ages (between 0 and 12 months) (Fuller and Conner, 1995). Analysis of pain-induced cries was combined (Grunau et al., 1990) with a (manual) coding of the facial expressions (Neonatal Facial Coding System, NFCS) (Grunau and Craig, 1987), in order to discriminate behavioral reactions between invasive and non-invasive procedures. The objective was to test if a newborn infant's cry could be used to measure pain, after heel-prick stimulus (Runefors et al., 2000). The analysis showed that the crying sound after the painful stimulus of the heel-prick had a significantly higher fundamental frequency and lasted longer at the first than at the fifth cry. However, while the first cry was more like a cry of pain, the fifth cry more resembled crying for other reasons. The conclusion was that crying could be used to measure pain in newborn infants only when the cause of crying was known. Different pain levels were also applied (Bellieni et al., 2004) and relations between cry characteristics and a pain score on the DAN (Douleur Aiguë du Nouveau-né, newborn acute pain) scale were evaluated. Results showed that a correlation existed when the DAN score was greater than 8 (on a 0 to 10 scale), and could be used as an alarm threshold.

Analysis of cries was also developed in other contexts. Characteristics of infant cries were correlated with perception and responses of adults, parents or non-parents in (Zeskind, 1980; Green et al., 1987; Zeskind and Marshall, 1988). Acoustic characteristics of the cries of newborn of marijuana users and non-users were compared (Lester and Dreher, 1989). Differences observed between both suggested that heavy marijuana use affected the neurophysiological integrity of the infant. The separation distress call in the absence of maternal body contact was evaluated by quantifying the amount of crying during the first 90 minutes after birth when the baby was placed either skin-to-skin with the mother, in a cot, or first in a cot and then skin-to-skin (Christensson et al., 1995).

Spontaneous cries were also processed: i) in the context of profound hearing loss and/or

perinatal asphyxia (Pearce and Taylor, 1993a,b; Schönweiler et al., 1996; Suaste-Rivas et al., 2004; Reyes-Galaviz et al., 2004; Galaviz and García, 2005; Reyes-Galaviz et al., 2005; Barajas-Montiel and Reyes-García, 2006; Verduzco-Mendoza et al., 2012; Wahid et al., 2016), ii) to find possible early signs of autism (Sheinkopf et al., 2012; Orlandi et al., 2012b) iii) in the context of monitoring (Orlandi et al., 2015), iv) to better understand vocal development and early communication (Zeskind et al., 1993; Wermke and Mende, 2009; Borysiak et al., 2016). Cries of hard-of-hearing and healthy infants were compared through duration, amplitude and melody (fundamental frequency fluctuations along a cry) description (Várallyay et al., 2004; Várallyay, 2006, 2007). Recently, baby emotional cries have also been studied in order to be integrated in a robotic baby caregiver (Yamamoto et al., 2013).

3.1.2. Preterm newborns Characterization of the premature crying episodes and their differences with cries of full-term infants were also largely explored, in order to explain differences observed in their neurophysiological maturity, and the subsequent impact on their speech development.

First studies focused on the analysis of induced-pain cries. The influence of neurophysiological maturity on induced-pain cries was firstly deducted when full-term newborns cries spectral variability was found to be more complex than for preterm newborns (Tenold et al., 1974). Later, a comparison between cries of premature and full-term newborns showed that the cries of smallest premature newborns compared with the controls were shorter, with higher fundamental frequency, and included bi-phonation and glide more often (Michelsson et al., 1983; Goberman and Robb, 1999). However, the cry characteristics changed with increasing conceptual age and the older the child the more the cry pattern resembled that of the full-term (Thodén et al., 1985). Evaluation of pain from facial expression and crying was performed in premature infants, but also in full-term and 2- and 4-month-old infants, each group receiving a dedicated stimulus (Johnston et al., 1993). Results from this multivariate analysis showed that premature infants were different from older infants, that full-term newborns were different from others, but that 2- and 4-month-olds were similar. Later, Stevens *et al* included two variables, severity of illness and behavioral state (sleep or awake) in the analysis (Stevens et al., 1994). Behavioral state was found to influence the facial action variables and severity of illness modified the acoustic cry variables. As for non-premature newborns, some of these works coupled the cry analysis with the facial activity coding NFCS (Craig et al., 1993; Johnston et al., 1993; Stevens et al., 1994).

Analysis of spontaneous cries of preterm infants has been recently investigated (Wermke et al., 2002; Orlandi et al., 2012a; Shinya et al., 2014; Manfredi et al., 2009; Orlandi et al., 2016). A comparison between spontaneous cries of six premature newborns (three pairs of twins) recorded at different ages (8th-9th week, 15th-17th week and 23rd-24th week) was performed (Wermke et al., 2002). Essential changes in the cries were observed from the 8th-9th week of life up to the 23rd-24th week of life, and were interpreted as an intentional

articulatory activity. The distress during cry was also shown as correlated with central blood oxygenation (Orlandi et al., 2012a). Results indicated that a similar decrease in oxygenation level occurs in both groups of patients, but the recovery time after the crying episode was more stable and rapid in full-term newborns than in preterm ones. An acoustic analysis of cries before feeding in both healthy preterm infants at term-equivalent ages and full-term newborns was performed (Shinya et al., 2014). Effects of gestational age, body size at recording and IntraUterine Growth Retardation (IUGR) were then investigated, and showed that shorter gestational age was significantly associated with a higher fundamental frequency, although no relation was found with smaller body size at recording nor IUGR. Fundamental frequency and formants of preterm newborns were shown to be related to an increasing gestational age (Manfredi et al., 2009) and generally higher for full-term newborns (Orlandi et al., 2016).

3.2. Methods for acoustic signal processing

Acoustic signal processing in paediatrics dealt with different objectives: extraction of features, cry segmentation, cry classification and assessment of other sounds (Figure 2). This section covers all these topics and is supplemented by a paragraph dedicated to feature definition. But first, as for the video analysis section, a particular attention has been paid to audio acquisition methods.

3.2.1. Audio acquisition Most of the studies have been conducted on real audio signals where microphones were placed from 10 to 30 centimeters of the infant's mouth. In case of pain-induced cry analysis, recordings have been performed from the stimulus to the end of the cry event (Golub and Corwin, 1982; Grunau et al., 1990; Michelsson and Michelsson, 1999; Várallyay et al., 2004) whereas in other clinical contexts, no further details about recording period were usually reported. In fact, recordings were annotated by experts resulting into small cry signals (few seconds) database. Thus, all authors constructed their own cry signals database supported by specific clinical annotations. Among them, we can cite the Baby Chillanto cry signals database, property of the Instituto Nacional de Astrofisica Optica y Electronica (Mexico). It was composed by five types of annotated cries (pain, hunger, normal, asphyxia and deaf) and on which relied the work of Reyes-Garcia group (Orozco-García and Reyes-García, 2003; Ortiz et al., 2004; Suaste-Rivas et al., 2004; Reyes-Galaviz et al., 2004; Galaviz and García, 2005; Reyes-Galaviz et al., 2005; Barajas-Montiel and Reyes-García, 2006; Hariharan et al., 2011; Verduzco-Mendoza et al., 2012; Díaz et al., 2012; Rosales-Pérez et al., 2015; Wahid et al., 2016). Cohort size varied from few infants (Lederman et al., 2008; Yamamoto et al., 2013) to more than three hundred babies (Várallyay, 2007) or more than 120 premature babies in (Stevens et al., 1994).

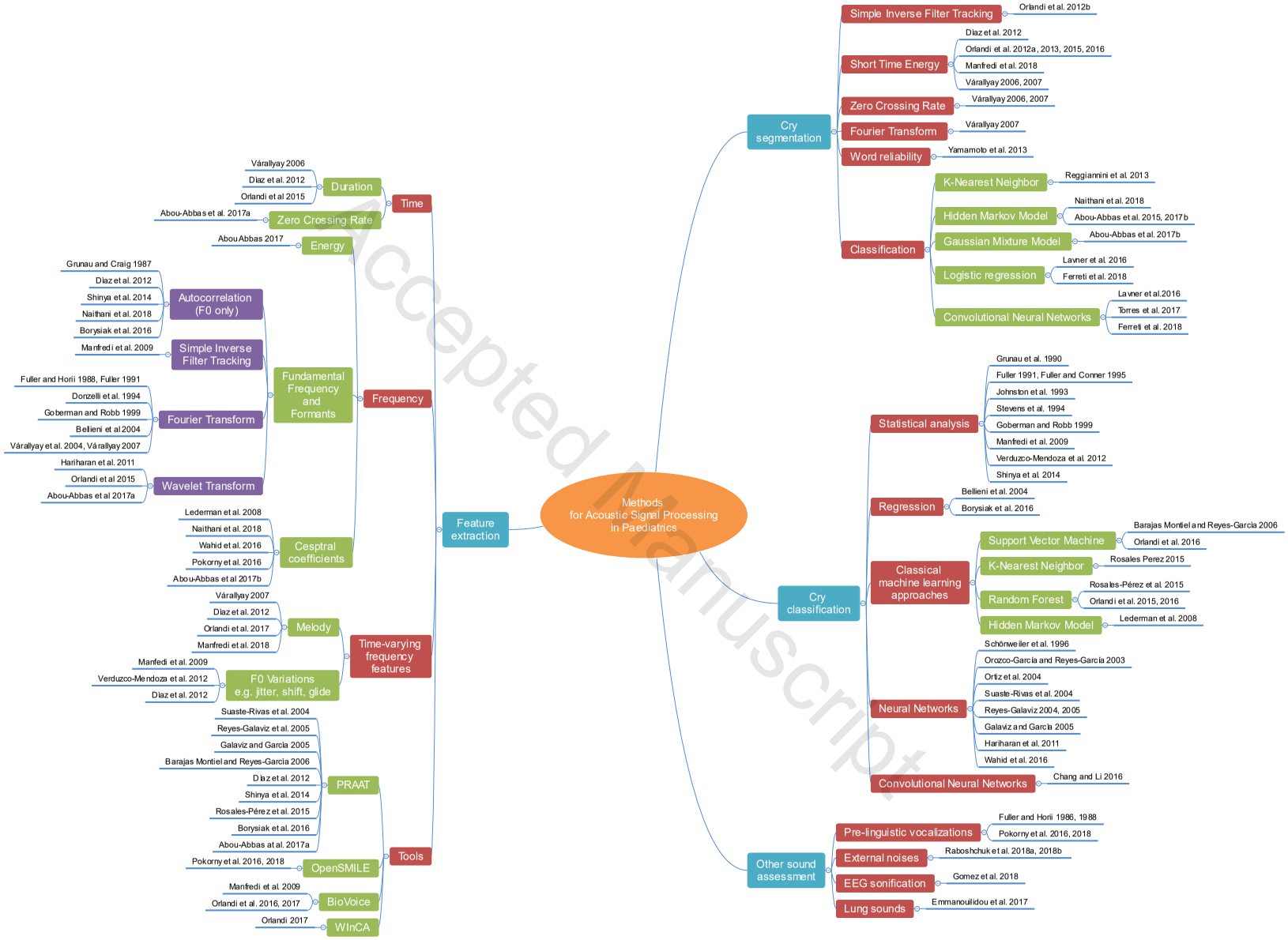

Figure 2. Overview of acoustic signal processing methods in paediatrics.

Several types of noise can affect the audio recordings (e.g. alarms, voice). Their occurrence directly depends on the acquisition environment. Authors usually did not mention where the recording took place except when it was performed at home (Várallyay, 2007; Yamamoto et al., 2013; Pokorny et al., 2016). However, we reported two types of acoustic environments: i) with low energy noise (e.g. background noise) (Várallyay, 2006, 2007; Orlandi et al., 2012a; Díaz et al., 2012; Orlandi et al., 2012b; Reggiannini et al., 2013; Orlandi et al., 2013, 2015, 2016; Manfredi et al., 2018) and ii) with high energy noise (e.g. medical device sounds, adults' voices) (Yamamoto et al., 2013; Abou-Abbas et al., 2015; Lavner et al., 2016; Abou-Abbas et al., 2017b; Naithani et al., 2018).

Only Manfredi's group dealt with synthetic crying signals (Orlandi et al., 2013), and recently, synthetic signals of the melody of cries have been also constructed (Orlandi et al., 2017; Manfredi et al., 2018). It was also proposed to create a simulated database to reproduce realistic acoustic scenarios (Ferretti et al., 2018). Noisy conditions have been created by synthetically adding four noises (human speech, interfering cries, beep sounds and background noise) to a clean set created by convolving 64 infant cry recordings with synthetic impulse responses.

3.2.2. Feature definition The analysis of cries mainly relied on the computation of features in two domains: time and frequency (see (LaGasse et al., 2005) for a review). In addition, several studies investigated the variations of the fundamental frequency along a cry event.

Time features The most common time features were naturally computed to characterize the duration of cries (Golub and Corwin, 1982; Lester and Dreher, 1989; Grunau et al., 1990; Donzelli et al., 1994; Stevens et al., 1994; Fuller and Conner, 1995; Christensson et al., 1995; Michelsson and Michelsson, 1999; Goberman and Robb, 1999; Runefors et al., 2000; Várallyay, 2006; Díaz et al., 2012). Among them, we found the total time in crying (Runefors et al., 2000; Várallyay, 2006), the ratio between duration of cry and total audio segment duration (Goberman and Robb, 1999; Várallyay, 2006), the mean (Donzelli et al., 1994; Várallyay, 2006) or the variation coefficient (Donzelli et al., 1994) of cry durations. Non-cry episodes were also examined through similar duration metrics (Várallyay, 2006). In the case of induced-pain, latency time (interval between the painful stimulus and the first cry) has been considered (Grunau et al., 1990; Stevens et al., 1994; Michelsson and Michelsson, 1999; Runefors et al., 2000).

In addition, intensity, or loudness, features such as average amplitude (Golub and Corwin, 1982) or amount of energy (Abou-Abbas et al., 2017a,b) of cries have been proposed. Zero Crossing Rate (ZCR) has also been quantified to characterize cries (Abou-Abbas et al., 2017a).

Frequency features The first way to describe the spectral composition of a signal is to compute spectral energy features. Authors proposed different approaches such as the computation of the overall spectral energy of the signal (Johnston et al., 1993; Stevens et al., 1994; Fuller and Conner, 1995; Reggiannini et al., 2013) or the energy only induced by low or high frequencies (Fuller and Horii, 1988; Fuller, 1991; Goberman and Robb, 1999).

Fundamental frequency (F0) ‡ was investigated by virtually all the cited works. In fact, it is a direct measurement of vocal development since it corresponds to the rate of glottal opening and closing in the vocal tract. F0 has been characterized through different statistical features computed from several cries such as mean (e.g. (Michelsson and Michelsson, 1999; Orlandi et al., 2015; Borysiak et al., 2016)), but also maximum and minimum (Shinya et al., 2014), standard deviation (Orlandi et al., 2016) or variation coefficient (Donzelli et al., 1994).

Formants, or resonance frequencies, are produced by the instantaneous shape of the vocal tract and have also been statically analyzed. Most of the studies that examined resonant frequencies were focused on the first two formants F1 and F2 (Wermke et al., 2002; Fuller, 1991; Donzelli et al., 1994; Orlandi et al., 2015), but some authors also proposed to assess F3 (Donzelli et al., 1994; Manfredi et al., 2009; Orlandi et al., 2016).

However, the difficulties met to extract resonance frequencies led to the computation of other coefficients to describe the spectral envelop such as Mel Frequency Cepstral Coefficients (MFCCs) (Lederman et al., 2008; Rosales-Pérez et al., 2015; Pokorny et al., 2016; Abou-Abbas et al., 2017b) or Linear Prediction Cepstral Coefficients (LPCCs) (Suaste-Rivas et al., 2004; Wahid et al., 2016). MFCCs are obtained through the projection of the signal on a mel-scale inspired by psychoacoustic model of human auditory perception whereas LPCCs are computed from the modelling of the vocal tract.

Time-varying frequency features The most common time-varying frequency descriptor is the pattern of F0 over a cry, or melody shape. Four main melodic shapes have been defined by Schönweiler *et al* (Schönweiler et al., 1996): falling, rising, falling-rising (or rising-falling) and flat. They were then reduced to three fundamental units (falling, rising and flat) and have been shown as the basis of 77 melodic shapes (Várallyay, 2007). A fifth melody shape, called "complex shape", was also considered to cover all melodic patterns composed by more than two fundamental units (Wermke and Mende, 2009; Orlandi et al., 2017; Manfredi et al., 2018).

Several other features have been defined to assess the F0 variations along a cry unit or during a cry event (succession of cry units). Among them, we can cite jitter (cycle-to-cycle variations of F0) (Fuller and Horii, 1986), shift (sudden change in F0) (Michelsson and Michelsson, 1999; Díaz et al., 2012; Runefors et al., 2000), glide (rapid variations in

‡ The term "pitch" is also employed by some authors. However, in speech production, fundamental frequency and pitch are not identical since pitch strictly refers to the perceptual quality of the frequency i.e. how our auditory system perceives different frequencies.

F0) (Michelsson and Michelsson, 1999; Runefors et al., 2000; Verduzco-Mendoza et al., 2012; Díaz et al., 2012), vibrato (series of waves with remarkable frequency variations) (Michelsson et al., 1983; Verduzco-Mendoza et al., 2012) or glottal roll (phonation of weak intensity and low F0, below the normal average measurement of the tone) (Michelsson et al., 1983; Verduzco-Mendoza et al., 2012).

3.2.3. Feature extraction This section provides an overview of the extraction methods used to estimate acoustics features. They have been organized regarding the three types of feature targeted: time, frequency and time-varying frequency. In addition, a paragraph has been dedicated to tools that were developed and/or used by the different authors.

Time features Computation of duration features relied on a segmentation step in order to extract cry and non-cry epochs. It has been performed either manually (Golub and Corwin, 1982; Lester and Dreher, 1989; Grunau et al., 1990; Donzelli et al., 1994; Stevens et al., 1994; Fuller and Conner, 1995; Christensson et al., 1995; Michelsson and Michelsson, 1999; Goberman and Robb, 1999; Runefors et al., 2000) or automatically (Várallyay, 2006; Díaz et al., 2012; Orlandi et al., 2015) (cf. section 3.2.4). For their part, ZCR or energy were computed directly from the signal by windowing (Abou-Abbas et al., 2017a).

Frequency features Three types of methods have been employed to estimate frequency features: temporal, spectral and cepstral.

Regarding time domain methods, autocorrelation function has been largely used to approximate the fundamental frequency (Grunau and Craig, 1987; Díaz et al., 2012; Shinya et al., 2014; Borysiak et al., 2016; Naithani et al., 2018). In practice, a window, encompassing several pitch periods (F0 was usually searched between 150 Hz and 1000 Hz), was used to calculate short-term autocorrelation sequence and the fundamental frequency was obtained at the maximum of this sequence. Sometimes, it has been supported by correction of the F0 estimation tracking errors (Borysiak et al., 2016). Manfredi *et al* also proposed a tuned method of the Simple Inverse Filter Tracking (SIFT) algorithm (Markel, 1972) which gave better F0 estimation (Manfredi et al., 2009).

In early studies, frequency feature extraction was generally based on a spectrographic analysis (Grunau et al., 1990; Michelsson and Michelsson, 1999; Runefors et al., 2000), but more recent ones developed automatized estimations methods using Fourier Transform (Fuller and Horii, 1988; Fuller, 1991; Donzelli et al., 1994; Goberman and Robb, 1999; Bellieni et al., 2004; Várallyay et al., 2004; Várallyay, 2007) or Wavelet Transform (Hariharan et al., 2011; Orlandi et al., 2015; Abou-Abbas et al., 2017a). Most of the time, energy features were directly computed from spectrum and peak-picking procedures were implemented to extract F0 or resonance frequencies (Fuller, 1991; Bellieni et al., 2004). The main limit of this approach was the presence of noise parts due to silence during crying episodes. To

overcome this issue, several authors worked with Long Time Average Spectrum (LTAS) that was the average spectrum computed from all selected cry periods of interest (e.g. without pauses or silences) (Fuller and Horii, 1988; Donzelli et al., 1994; Goberman and Robb, 1999). For their part, Varallyay *et al* used Smoothed Spectrum Method (SSM). In fact, resulting spectrum only contained cry components thanks to an initial statistical processing removing noise parts induced by silence (Várallyay et al., 2004; Várallyay, 2007). Recently, Continuous Wavelet Transform (CWT) approaches, known for their robustness to noise, have also been considered to estimate F0 and formants (Orlandi et al., 2015) or to extract energy feature from different component levels (Hariharan et al., 2011; Abou-Abbas et al., 2017a).

As mentioned in Section 3.2.2, cepstral coefficients (MFCCs and LPCCs) have also been extracted to describe audio signals (Lederman et al., 2008; Wahid et al., 2016; Pokorny et al., 2016; Abou-Abbas et al., 2017b; Naithani et al., 2018). MFCCs were traditionally computed in six steps: a) partitioning the signal into short frames, b) computing the power spectrum density, c) projecting the power spectra on mel-filter bank (simulation of human auditory perception) and sum the energy in each filter, d) taking the logarithm of all filter bank energies e) calculating the Discrete Cosinus Transform (DCT) of the log filter bank energies f) keeping DCT coefficients as MFCCs. However, Abou-Abbas *et al* recently proposed to compute similar MFCCs features after applying Energy Mode Decomposition (EMD) (Abou-Abbas et al., 2017a,b) or by using Discrete Wavelet Transform (DWT) instead of DCT (Abou-Abbas et al., 2017b). For their part, LPCCs were computed from the Smoothed Auto Regressive (AR) power spectrum where AR coefficients were estimated by Levinson-Durbin algorithm.

Time-varying frequency features Melody shapes have been identified by two methods based on the projection of F0 pattern into a grid: Five-Line Method (FLN) (fixed 5-lines grid from 330 Hz to 700 Hz) (Várallyay, 2007) and dodecagram (variable grid depending on the first F0 estimation of the cry) (Díaz et al., 2012). The dodecagram method showed better results due to its adaptability. Melody shapes have also been synthetically constructed by the mean of rules on F0 variations (e.g. the falling shape was obtained setting a maximum frequency of 650 Hz at 0.4 second, followed by a slow decrease towards 450 Hz) in order to compare tool abilities to estimate F0 patterns (Orlandi et al., 2017). Recently, an approximation of melody shapes by quadratic and fourth order polynomial functions was added (Manfredi et al., 2018).

The other F0 variations features (e.g. jitter, shift, glide) were mainly defined by decision rules on the F0 contour, either visually noticed (Michelsson et al., 1983; Michelsson and Michelsson, 1999) or automatically computed (Manfredi et al., 2009; Verduzco-Mendoza et al., 2012; Díaz et al., 2012). As an example, in (Díaz et al., 2012), a shift was detected when a sudden change of ascends or descends in the F0 between 100 Hz and 600 Hz within 0.1 second occurred.

Tools Some software tools for the acoustic analysis of infant cries were developed and/or used. The most popular one is PRAAT, initially proposed for adult's voice by Boesrma in 2002 (Boersma, 2002). It was used in (Suaste-Rivas et al., 2004; Reyes-Galaviz et al., 2005; Galaviz and García, 2005; Barajas-Montiel and Reyes-García, 2006; Díaz et al., 2012; Shinya et al., 2014; Rosales-Pérez et al., 2015; Borysiak et al., 2016; Abou-Abbas et al., 2017a). Acoustic parameters have also been extracted by the mean of the openSMILE toolkit (Eyben et al., 2013; Pokorny et al., 2016, 2018). Both softwares provided automatic computation of a wide variety of features (e.g. F0, formants, MFCCs, LPCCs, jitter, shimmer) but had to be manually tuned to give relevant analysis of infant cries.

For their part, Manfredi *et al* developed BioVoice (Manfredi et al., 2009; Orlandi et al., 2016) and WInCA (Orlandi et al., 2017), two softwares adapted to infant cry analysis, where different estimation methods of F0 (respectively, SIFT and wavelet) and resonance frequencies (respectively, peak picking in the power spectral density and wavelet) were implemented. When comparing the two approaches with PRAAT, where F0 was computed by autocorrelation method and formants were approached by LPCCs, authors found out that PRAAT gave better results to approximate fundamental frequency whereas BioVoice was more accurate for formant estimation (Orlandi et al., 2017).

3.2.4. Automatic cry segmentation In the analysis of sound, one of the problems relies on the segmentation of the recordings into Voiced/UnVoiced (V/UV) parts, also called detection of Cry Units (CU), in order to extract relevant acoustic parts. If in a large majority of papers, the cry segments were manually selected, some recent studies, described below, proposed solutions to perform this segmentation automatically.

A V/UV detection procedure, based on SIFT, where an interval was selected as voiced if the maximum of the autocorrelation function is greater than a fixed threshold was proposed (Orlandi et al., 2012b). Methods based on thresholding the Short-Term Energy (STE) function were also investigated (Díaz et al., 2012; Orlandi et al., 2012a, 2013, 2015, 2016; Manfredi et al., 2018). In (Orlandi et al., 2013), authors proposed to automatically compute the threshold using Otsu method (Otsu, 1979). Cry segmentation was also performed by combining two short-time methods: STE and ZCR (Várallyay, 2006). In fact, STE provided a distinction between audible sounds and silence while ZCR permitted to detect V/UV parts. Then, a threshold was applied to extract CU. Later, authors added a third step to distinguish harmonic and non-harmonic audio segments (Várallyay, 2007). It was based on the hypothesis that spectral structure of a crying segment is harmonic since it only contains the fundamental frequency and its harmonics. Nevertheless, no quantitative evaluation was performed. A marginal method has been proposed by Yamamoto *et al* (Yamamoto et al., 2013) where a baby cry was detected if at least one of the two following conditions was respected: the word reliability computed by the adult word detection tool "Julius" (Lee et al., 2001) was under a threshold or the change of the fundamental frequency for a time

period was superior to another threshold. A detection accuracy of only 69.4% was obtained.

A few recent approaches considered the cry segmentation as a classification problem. Reggiannini *et al* proposed an automatized classification in three classes: voiced part, unvoiced part and silence. By the use of KNN, they achieved to discriminate the three states with an Area Under Curve (AUC) of 0.88 (Reggiannini et al., 2013). Expiratory and inspiratory phases of the cries were separated in some studies. Hidden Markov Model (HMM) based approaches led to 83.79% of accuracy (Abou-Abbas et al., 2015) when six classes (expiratory phases, inspiratory phases, noise, adult speech, silence and beeps) were considered. Later, these results were improved by formulating the problem with three classes: expiratory phases, inspiratory phases and others (Abou-Abbas et al., 2017b) or residuals (Naithani et al., 2018) (a class regrouping all previous mentioned noisy sounds). HMM and Gaussian Mixture Model (GMM) methods were compared and GMM gave the best results with a classification error rate of 8.9% (Abou-Abbas et al., 2017b). A total accuracy of 89.2% was also reached with HMM in (Naithani et al., 2018).

Newly, deep learning approaches have been considered to detect cries in a domestic environment (Lavner et al., 2016; Torres et al., 2017) or in NICU (Ferretti et al., 2018). In all studies, log mel-frequency features were computed on windows and combined to construct the Convolution Neural Network (CNN) input layer. Such methods requiring large dataset, authors proposed to introduce normalization and regularization to adapt CNN to modest dataset (Torres et al., 2017) or to enhance dataset by adding simulated data (Ferretti et al., 2018). CNN slightly outperformed classification methods in order to detect cries. In fact, CNN gave lower false-positive rate than logistic regression (Lavner et al., 2016) and an AUC upper than 90% was reached (Torres et al., 2017). In NICU, an average accuracy of 86.58% was obtained (Ferretti et al., 2018).

3.2.5. Automatic cry classification Automated cry classification have been performed after a manual (Grunau et al., 1990; Fuller, 1991; Johnston et al., 1993; Stevens et al., 1994; Fuller and Conner, 1995; Goberman and Robb, 1999; Orozco-García and Reyes-García, 2003; Bellieni et al., 2004; Suaste-Rivas et al., 2004; Ortiz et al., 2004; Reyes-Galaviz et al., 2004, 2005; Galaviz and García, 2005; Barajas-Montiel and Reyes-García, 2006; Lederman et al., 2008; Hariharan et al., 2011; Verduzco-Mendoza et al., 2012; Shinya et al., 2014; Rosales-Pérez et al., 2015; Borysiak et al., 2016; Wahid et al., 2016) or an automated (Manfredi et al., 2009; Orlandi et al., 2015, 2016) segmentation step. First studies used classical statistical approaches such as Student T-test (Manfredi et al., 2009; Verduzco-Mendoza et al., 2012), ANOVA (Grunau et al., 1990; Fuller, 1991; Goberman and Robb, 1999; Shinya et al., 2014) and MANOVA (Johnston et al., 1993; Stevens et al., 1994; Fuller and Conner, 1995) or regression (Bellieni et al., 2004; Borysiak et al., 2016). It was applied either to compare infant ages (Johnston et al., 1993; Goberman and Robb, 1999; Manfredi et al., 2009; Shinya et al., 2014) and gender (Fuller, 1991; Borysiak et al., 2016), to evaluate pain (Fuller and

Conner, 1995; Stevens et al., 1994; Bellieni et al., 2004) or to recognize pathologies (Grunau et al., 1990; Verduzco-Mendoza et al., 2012).

More recent papers investigated classification approaches using a high number of features. Different numbers of cry classes were considered according to the clinical target: two classes (normal vs abnormal (Orozco-García and Reyes-García, 2003; Ortiz et al., 2004; Lederman et al., 2008; Hariharan et al., 2011; Rosales-Pérez et al., 2015) or preterm vs full-term (Orlandi et al., 2015, 2016)), three classes (normal, hypo acoustic and asphyxia (Reyes-Galaviz et al., 2004; Suaste-Rivas et al., 2004; Galaviz and García, 2005; Reyes-Galaviz et al., 2005; Barajas-Montiel and Reyes-García, 2006), hunger, pain and sleep (Chang and Li, 2016) or hunger, pain and no-pain-no-hunger (Barajas-Montiel and Reyes-García, 2006)) and five classes (pain, asphyxia, hunger, deaf and normal (Wahid et al., 2016)). A wide variety of machine learning approaches has been evaluated, regrouping classical methods such as SVM (Barajas-Montiel and Reyes-García, 2006; Orlandi et al., 2016), KNN (Rosales-Pérez et al., 2015), Random Forest (RF) (Rosales-Pérez et al., 2015; Orlandi et al., 2015, 2016), HMM (Lederman et al., 2008) or Neural Networks (Schönweiler et al., 1996; Orozco-García and Reyes-García, 2003; Suaste-Rivas et al., 2004; Ortiz et al., 2004; Reyes-Galaviz et al., 2004, 2005; Galaviz and García, 2005; Hariharan et al., 2011; Wahid et al., 2016). Classification results were efficient since some studies reached results above 95% (Orozco-García and Reyes-García, 2003; Reyes-Galaviz et al., 2004, 2005; Galaviz and García, 2005; Barajas-Montiel and Reyes-García, 2006; Hariharan et al., 2011).

Deep learning was also investigated to classify cries into three categories: hungry, pain and sleep (Chang and Li, 2016). Spectrogram of cries were computed by Fast Fourier Transform (FFT) and used as input layer of a CNN. The method showed promising results with 78.5% of accuracy.

3.2.6. Other sound assessment Several recent audio processing methods have been proposed regarding non-cry signals and concerning either pre-linguistic vocalizations (including cooing) (Fuller and Horii, 1986, 1988; Pokorny et al., 2016, 2018). Non-voice analyses were also proposed in different contexts such as external noise detection (Raboshchuk et al., 2018a,b), EEG sonification (Gomez et al., 2018) or lung sound assessment (Emmanouilidou et al., 2017).

Cooing, manually selected, were analyzed by Fuller *et al* . in 30 infants ranging in age from 2 to 6 months, where significant differences were found in F0 and Mean Spectral Energy (MSE) with classical cries (fussy, hungry and pain) (Fuller and Horii, 1986, 1988). Pre-linguistic vocalizations have also been studied in 7- to 12-month-old infants having received the diagnosis of a neuro-developmental disorder (autism, Rett syndrome, fragile X syndrome) by Pokorny *et al* . They processed retrospective home video recordings provided by the family and made during family events, before the disorder was diagnosed. In (Pokorny et al., 2016), a comparison between manual and automated segmentation of vocalizations, using machine

learning (HMM, SVM and RF), led to only 38.0%, where errors came from confusions with parental voices or voices from television. Later, they evaluated more than 6000 features to differentiate typical and atypical early speech language of one infant with Rett syndrome. Main differences were observed in auditory attributes such as timbre and pitch (Pokorny et al., 2018).

For their part, Raboshchuk *et al* focused on the automatic detection of alarms and external vocalizations (e.g. nurses, parents) in NICU. In (Raboshchuk et al., 2018a), alarms were detected thanks to the knowledge of alarm characteristics (e.g. frequencies) integrated to a GMM classifier. In (Raboshchuk et al., 2018b), several pre-processing approaches were tested: spectral subtraction, non-negative matrix factorization and combination of both. Best results were obtained with non-negative matrix factorization followed by spectral subtraction.

Recently, marginal purposes were investigated through audio processing for example to detect neonatal seizures from EEG or to detect lung sounds abnormalities. EEG signal was converted into an audible audio signal (process is called sonification) in order to hear relative frequency change when a seizure occurred (Gomez et al., 2018). It was shown that sonification methods perform similarly well, with a smaller inter-observer variability in comparison with visual interpretation. Lung audio recordings of 1000 children were also studied (Emmanouilidou et al., 2017). First, noise suppression techniques were applied to discard ambient sounds, sensors artifacts or crying. Notably, crying episodes were discarded by the use of SVM classifier trained with spectrotemporal features. Finally, normal and pathological lung sounds were classified through SVM classifier with an accuracy of 86.7%.

4. Discussion and Conclusion

This review showed that a lot of works have been published since several decades in the domains of video and audio processing in paediatrics.

The review of video processing showed that video recordings have been mainly exploited for motion analyses, in two major clinical contexts, general movement assessment and neonatal seizures detection and characterization. These studies have shown that the quality and quantity of movements are markers of the infant's neurological health.

If recent improvements in digital video processing allowed an increasing automation, most of the above-mentioned studies need to be manually initialized in order to select the considered region (whole baby or limbs). Promising results were recently obtained by CNN but the method was only applied on a controlled environment setup (Dossso et al., 2018). Furthermore, most of the proposed methods extracted global motion information (i.e. without identifying each limb contribution). This may be improved by exploiting methods developed in adults using more precise body descriptors, such as kinematic or shape models (Poppe, 2007).

Video-based respiration and heart rate estimation techniques showed interesting results. However, these techniques only work when the baby is not moving. Nowadays, the most valuable application in term of infant monitoring may be the automatic detection of apneas that generally do not occur when the baby is moving. Similarly, face analysis can't yet be integrated into a monitoring system since none of the proposed techniques are robust to occlusions that can happen in NICU, either from the baby itself or from external adult manipulations.

Another limitation is related to the video recordings duration and the constrained acquisition setups. Indeed, for most of the studies, recordings only contained periods of interest and the infants were placed in some specific conditions in order to ensure an appropriate acquisition. Furthermore, infants were generally not covered, with appropriate lightening conditions and no external interventions. On the other hand, long video recordings may include medical staff or parents' presence in the image that had to be detected and eliminated to discard non-suitable periods. This problem has been recently addressed in (Cabon et al., 2017). Similarly, absences of the newborn in the bed will have to be detected to avoid the analysis of irrelevant periods. A recent study, based on motion analysis, showed encouraging results in that way (Long et al., 2018).

The review of acoustic analyses shown that most of the studies was devoted to cries. Initially focused on pain-induced cries, more recent studies considered also spontaneous cries. Processing, most of the time based on frequency features, allowed to distinguish normal and pathological cries but also to classify different types of cries. They were also explored for premature newborns to identify differences in their neurophysiological maturity. A few papers dealing with the processing of pre-linguistic vocalizations, NICU alarms, EEG sonification and lung sound classification were also identified.

As for video, long audio recordings are parasitized by different sources such as nurse or parents' voices, alarms of monitoring devices, ventilation noise, etc. Although some authors worked on audio recordings performed in such environment, the automatic recognition of pathologic cries in NICU still remains difficult. In fact, only a few recent studies showed around 90% of accuracy in cry segmentation but no classification was proposed from there (Abou-Abbas et al., 2017b; Naithani et al., 2018; Ferretti et al., 2018).

Additionally, baby sounds other than cries, like coughing but also vowel sounds, were slightly or not investigated. And yet, they are concomitant with many diseases (Hirschberg, 1980) and may be an indicator of maturation (Thach, 2007) and of vocal development (Caskey et al., 2011). In fact, authors usually discarded them or included them with other sounds without making distinction (e.g. as expiratory sounds).

It is also important to notice that joint audio and video processing was not yet envisaged at this time. Only one study integrating audio and video processing, was, to our knowledge, published in (Orlandi et al., 2015), where a contactless system for Audio-Video Infant Monitoring (AVIM) was proposed. The analysis of movements was semi-automatic since

the user had to select points to track on the video frame whereas cry analysis was performed automatically after a manual suppression of interfering sounds. Nevertheless, audio and video were processed separately. Obviously, a combination of these two components could broaden the scope of applications in early clinical diagnosis of several pathologies. Moreover, it could be helpful in automatic sleep analysis, where both motion and baby sounds are important behavioral descriptors.

On the other hand, this review being dedicated to audio- and video-based systems, other acquisition systems have not been included. Briefly, the use of infrared thermography has been investigated to measure the skin temperature of newborns (Anderson et al., 1990; Heimann et al., 2013; Abbas and Leonhardt, 2014) or the respiratory rate and timing (Goldman, 2012). Another example is the use of depth cameras (e.g. Microsoft Kinect) to analyze infants' movements (Olsen et al., 2014; Marschik et al., 2017).

Finally, if a lot of works have dealt and continue to deal with the automated processing of video and audio in paediatrics, a fully-automated efficient system does not exist. It will have to tackle above-mentioned difficulties by integrating robust processing methods to cope with unconstrained and long-term acquisition time such as encountered with monitoring systems in NICU.

Acknowledgment

Results incorporated in this publication received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N^o 689260 (Digi-NewB project).

Conflict of interest statement

There is no conflict of interest associated with this publication.

5. References

References

Aarts, L. A. M., Jeanne, V., Cleary, J. P., Lieber, C., Nelson, J. S., Bambang Oetomo, S., and Verkruysse, W. (2013). Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit - a pilot study. Early Human Development, 89:943–948.

Abbas, A. K. and Leonhardt, S. (2014). Intelligent neonatal monitoring based on a virtual thermal sensor. BMC Medical Imaging, 14:9.

Abou-Abbas, L., Alaie, H. F., and Tadj, C. (2015). Automatic detection of the expiratory and inspiratory phases in newborn cry signals. Biomedical Signal Processing and Control, 19:35–43.

Abou-Abbas, L., Tadj, C., and Fersaie, H. A. (2017a). A fully automated approach for baby cry signal segmentation and boundary detection of expiratory and inspiratory episodes. The Journal of the Acoustical Society of America, 142(3):1318–1331.

Abou-Abbas, L., Tadj, C., Gargour, C., and Montazeri, L. (2017b). Expiratory and inspiratory cries detection using different signals' decomposition techniques. Journal of Voice, 31(2):259–e13.

Adde, L., Helbostad, J. L., Jensenius, A. R., Taraldsen, G., Grunewaldt, K. H., and Stoen, R. (2010). Early prediction of cerebral palsy by computer-based video analysis of general movements: A feasibility study. Developmental Medicine & Child Neurology, 52(8):773–8.

Adde, L., Helbostad, J. L., Jensenius, A. R., Taraldsen, G., and Stoen, R. (2009). Using computer-based video analysis in the study of fidgety movements. Early human development, 85(9):541–7.

Alinovi, D., Ferrari, G., Pisani, F., and Raheli, R. (2018). Respiratory rate monitoring by video processing using local motion magnification. In 2018 26th European Signal Processing Conference (EUSIPCO), pages 1780–1784. IEEE.

Anders, T. F., Emde, R. N., and Parmelee, A. H. (1971). A manual of standardized terminology, techniques and criteria for scoring of states of sleep and wakefulness in newborn infants. Los Angeles: UCLA Brain Information Service, NINDS Neurological Information Network.

Anders, T. F. and Sostek, A. M. (1976). The use of time lapse video recording of sleep-wake behavior in human infants. Psychophysiology, 13(2):155–8.

Anderson, E., Wailoo, M., and Petersen, S. (1990). Use of thermographic imaging to study babies sleeping at home. Archives of Disease in Childhood, 65(11):1266–1267.

Antognoli, L., Marchionni, P., Nobile, S., Carnielli, V., and Scalise, L. (2018). Assessment of cardio-respiratory rates by non-invasive measurement methods in hospitalized preterm neonates. In 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pages 1–5. IEEE.

Barajas-Montiel, S. E. and Reyes-García, C. A. (2006). Fuzzy support vector machines for automatic infant cry recognition. In Intelligent Computing in Signal Processing and Pattern Recognition, pages 876–881. Springer.

Bellieni, C. V., Sisto, R., Cordelli, D. M., and Buonocore, G. (2004). Cry features reflect pain intensity in term newborns: An alarm threshold. Pediatric research, 55(1):142–146.

Boersma, P. (2002). PRAAT, a system for doing phonetics by computer. Glot international, 5(9/10):341–345.

Borysiak, A., Hesse, V., Wermke, P., Hain, J., Robb, M., and Wermke, K. (2016). Fundamental frequency of crying in two-month-old boys and girls: Do sex hormones during mini-puberty mediate differences? Journal of Voice.

Bos, A. F., Martijn, A., Okken, A., and Prechtl, H. F. R. (1998a). Quality of general movements in preterm infants with transient periventricular echodensities. Acta Paediatrica, 87(3):328–335.

Bos, A. F., Martijn, A., van Asperen, R. M., Hadders-Algra, M., Okken, A., and Prechtl, H. F. (1998b). Qualitative assessment of general movements in high-risk preterm infants with chronic lung disease requiring dexamethasone therapy. The Journal of Pediatrics, 132(2):300–6.

Brieva, J. and Moya-Albor, E. (2017). Phase-based motion magnification video for monitoring of vital signals using the hermite transform. In 13th International Conference on Medical Information Processing and Analysis, volume 10572, page 105720M. International Society for Optics and Photonics.

Cabon, S., Poree, F., Simon, A., Ugolin, M., Rosec, O., Carrault, G., and Pladys, P. (2017). Motion estimation and characterization in premature newborns using long duration video recordings. IRBM, 38(4):207–213.

Caskey, M., Stephens, B., Tucker, R., and Vohr, B. (2011). Importance of parent talk on the development of preterm infant vocalizations. Pediatrics, pages peds–2011.

Cattani, L., Alinovi, D., Ferrari, G., Raheli, R., Pavlidis, E., Spagnoli, C., and Pisani, F. (2014). A wire-free, non-invasive, low-cost video processing-based approach to neonatal apnoea detection. BIOMS 2014 - 2014 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications, Proceedings, pages 67–73.

Cattani, L., Alinovi, D., Ferrari, G., Raheli, R., Pavlidis, E., Spagnoli, C., and Pisani, F. (2017). Monitoring infants by automatic video processing: A unified approach to motion analysis. Computers in Biology and Medicine, 80:158–165.

Chang, C.-Y. and Li, J.-J. (2016). Application of deep learning for recognizing infant cries. In Consumer Electronics-Taiwan (ICCE-TW), 2016 IEEE International Conference on, pages 1–2. IEEE.

Christensson, K., Cabrera, T., Christensson, E., Uvnas-Moberg, K., and Winberg, J. (1995). Separation distress call in the human neonate in the absence of maternal body contact. Acta Paediatrica, 84(5):468–473.

Craig, K. D., Whitfield, M. F., Grunau, R. V., Linton, J., and Hadjistavropoulos, H. D. (1993). Pain in the preterm neonate: Behavioural and physiological indices. Pain, 52(3):287–299.

Cuppens, K., Lagae, L., Ceulemans, B., Van Huffel, S., and Vanrumste, B. (2010). Automatic video detection of body movement during sleep based on optical flow in pediatric patients with epilepsy. Medical & Biological Engineering & Computing, 48(9):923–31.

Cuppens, K., Lagae, L., and Vanrumste, B. (2009). Towards automatic detection of movement during sleep in pediatric patients with epilepsy by means of video recordings and the optical flow algorithm. IFMBE Proceedings, 22:784–789.

Díaz, M. A. R., García, C. A. R., Robles, L. C. A., Altamirano, J. E. X., and Mendoza, A. V. (2012). Automatic infant cry analysis for the identification of qualitative features to help opportune diagnosis. Biomedical Signal Processing and Control, 7(1):43–49.

Dogra, D. P., Majumdar, A. K., Sural, S., Mukherjee, J., Mukherjee, S., and Singh, A. (2012). Toward automating hammersmith pulled-to-sit examination of infants using feature point based video object tracking. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 20:38–47.

Donzelli, G. P., Rapisardi, G., Moroni, M., Zani, S., Tomasini, B., Ismaelli, A., and Bruscaglioni, P. (1994). Computerized cry analysis in infants affected by severe protein energy malnutrition. Acta Paediatrica, 83(2):204–11.

Dossso, Y. S., Bekele, A., Nizami, S., Aubertin, C., Greenwood, K., Harrold, J., and Green, J. R. (2018). Segmentation of patient images in the neonatal intensive care unit. In 2018 IEEE Life Sciences Conference (LSC), pages 45–48. IEEE.

Emmanouilidou, D., McCollum, E. D., Park, D. E., and Elhilali, M. (2017). Computerized lung sound screening for pediatric auscultation in noisy field environments. IEEE Transactions on Biomedical Engineering, 65(7):1564–1574.

Eyben, F., Weninger, F., Gross, F., and Schuller, B. (2013). Recent developments in openSMILE, the munich open-source multimedia feature extractor. In Proceedings of the 21st ACM International Conference on Multimedia, MM '13, pages 835–838, New York, NY, USA. ACM.

Fang, C.-Y., Hsieh, H.-H., and Chen, S.-W. (2015). A vision-based infant respiratory frequency detection system. In Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on, pages 1–8. IEEE.

Ferretti, D., Severini, M., Principi, E., Cenci, A., and Squartini, S. (2018). Infant cry detection in adverse acoustic environments by using deep neural networks. In 26th European Signal Processing Conference, EUSIPCO 2018. European Signal Processing Conference, EUSIPCO.

Fuller, B. F. (1991). Acoustic discrimination of three types of infant cries. Nursing Research, 40(3):156–160.

Fuller, B. F. and Conner, D. A. (1995). The effect of pain on infant behaviors. Clinical Nursing Research, 4(3):253–273.

Fuller, B. F. and Horii, Y. (1986). Differences in fundamental frequency, jitter, and shimmer among four types of infant vocalizations. Journal of Communication Disorders, 19(6):441–447.

Fuller, B. F. and Horii, Y. (1988). Spectral energy distribution in four types of infant vocalizations. Journal of Communication Disorders, 21(3):251–61.

Fuller, P. W., Wenner, W. H., and Blackburn, S. (1978). Comparison between time-lapse video recordings of behavior and polygraphic state determinations in premature infants. Psychophysiology, 15(6):594–8.

Galaviz, O. F. R. and García, C. A. R. (2005). Infant cry classification to identify hypo acoustics and asphyxia comparing an evolutionary-neural system with a neural network system. In Mexican International Conference on Artificial Intelligence, pages 949–958. Springer.

Goberman, A. M. and Robb, M. P. (1999). Acoustic examination of preterm and full-term infant cries: The long-time average spectrum. Journal of Speech, Language, and Hearing Research, 42(4):850–61.

Goldman, L. J. (2012). Nasal airflow and thoracoabdominal motion in children using infrared thermographic video processing. Pediatric Pulmonology, 47(5):476–486.

Golub, H. L. and Corwin, M. J. (1982). Infant cry: A clue to diagnosis. Pediatrics, 69(2):197–201.

Gomez, S., O'Sullivan, M., Popovici, E., Mathieson, S., Boylan, G., and Temko, A. (2018). On sound-based interpretation of neonatal eeg. arXiv preprint arXiv:1806.03047.

Green, J. A., Jones, L. E., and Gustafson, G. E. (1987). Perception of cries by parents and nonparents: Relation to cry acoustics. Developmental Psychology, 23(3):370.

Grigg-Damberger, M., Gozal, D., Marcus, C. L., Quan, S. F., Rosen, C. L., Chervin, R. D., Wise, M., Picchietti, D. L., Sheldon, S. H., and Iber, C. (2007). The visual scoring of sleep and arousal in infants and children. Journal of Clinical Sleep Medicine, 3(2):201–40.

Grunau, R. V. and Craig, K. D. (1987). Pain expression in neonates: Facial action and cry. Pain, 28(3):395–410.

Grunau, R. V., Johnston, C. C., and Craig, K. D. (1990). Neonatal facial and cry responses to invasive and non-invasive procedures. Pain, 42(3):295–305.

Guzzetta, A., Mercuri, E., Rapisardi, G., Ferrari, F., Roversi, M. F., Cowan, F., Rutherford, M., Paolicelli, P. B., Einspieler, C., Boldrini, A., Dubowitz, L., Prechtl, H. F., and Cioni, G. (2003). General movements detect early signs of hemiplegia in term infants with neonatal cerebral infarction. Neuropediatrics, 34(2):61–6.

Hariharan, M., Yaacob, S., and Awang, S. A. (2011). Pathological infant cry analysis using wavelet packet transform and probabilistic neural network. Expert Systems with Applications, 38(12):15377–15382.

Hazelhoff, L., Han, J., Bambang-Oetomo, S., et al. (2009). Behavioral state detection of newborns based on facial expression analysis. In International Conference on Advanced Concepts for Intelligent Vision Systems, pages 698–709. Springer.

Heimann, K., Jergus, K., Abbas, A. K., Heussen, N., Leonhardt, S., and Orlikowsky, T. (2013). Infrared thermography for detailed registration of thermoregulation in premature infants. Journal of Perinatal Medicine, 41(5):613–620.

Hirschberg, J. (1980). Acoustic analysis of pathological cries, stridors and coughing sounds in infancy. International Journal of Pediatric Otorhinolaryngology, 2(4):287–300.

Horn, B. K. and Schunck, B. G. (1981). Determining optical flow. Artificial Intelligence, 17(1-3):185–203.

Huvanandana, J., Thamrin, C., Tracy, M., Hinder, M., Nguyen, C., and McEwan, A. (2017). Advanced analyses of physiological signals in the neonatal intensive care unit. Physiological Measurement, 38(10):R253.

Johnston, C. C., Stevens, B., Craig, K. D., and Grunau, R. V. (1993). Developmental changes in pain expression in premature, full-term, two-and four-month-old infants. Pain, 52(2):201–208.

Kaneshi, Y., Ohta, H., Morioka, K., Hayasaka, I., Uzuki, Y., Akimoto, T., Moriichi, A., Nakagawa, M., Oishi, Y., Wakamatsu, H., Honma, N., Suma, H., Sakashita, R., Tsujimura, S.-I., Higuchi, S., Shimokawara, M., Cho, K., and Minakami, H. (2016). Influence of light exposure at nighttime on sleep development and body growth of preterm infants. Scientific Reports, 6:21680.

Karayiannis, N. B., Sami, A., Frost, J., Wise, M. S., and Mizrahi, E. M. (2004). Quantifying motion in video recordings of neonatal seizures by feature trackers based on predictive block matching. In Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE, volume 1, pages 1447–1450. IEEE.

Karayiannis, N. B., Sami, A., Frost, J. D., Wise, M. S., and Mizrahi, E. M. (2005a). Automated extraction of temporal motor activity signals from video recordings of neonatal seizures based on adaptive block matching. IEEE Transactions on Biomedical Engineering, 52:676–686.

Karayiannis, N. B., Srinivasan, S., Bhattacharya, R., Wise, M. S., Frost, J. D., J., and Mizrahi, E. M. (2001). Extraction of motion strength and motor activity signals from video recordings of neonatal seizures. IEEE Transactions on Medical Imaging, 20(9):965–80.

Karayiannis, N. B. and Tao, G. (2006). An improved procedure for the extraction of temporal motion strength signals from video recordings of neonatal seizures. Image and Vision Computing, 24(1):27–40.

Karayiannis, N. B., Tao, G., Frost, J. D., J., Wise, M. S., Hrachovy, R. A., and Mizrahi, E. M. (2006). Automated detection of videotaped neonatal seizures based on motion segmentation methods. Clinical Neurophysiology, 117(7):1585–94.

Karayiannis, N. B., Tao, G., Xiong, Y., Sami, A., Varughese, B., Frost, J. D., J., Wise, M. S., and Mizrahi, E. M. (2005b). Computerized motion analysis of videotaped neonatal seizures of epileptic origin. Epilepsia, 46(6):901–17.

Karayiannis, N. B., Varughese, B., Tao, G., Frost, J. D., J., Wise, M. S., and Mizrahi, E. M. (2005c). Quantifying motion in video recordings of neonatal seizures by regularized optical flow methods. IEEE Transactions on Image Processing, 14(7):890–903.

Karayiannis, N. B., Xiong, Y., Frost, J. D., J., Wise, M. S., and Mizrahi, E. M. (2005d). Quantifying motion in video recordings of neonatal seizures by robust motion trackers based on block motion models. IEEE Transactions on Biomedical Engineering, 52(6):1065–77.

Koolen, N., Decroupet, O., Dereymaeker, A., Jansen, K., Vervisch, J., Matic, V., Vanrumste, B., Naulaers, G., Huffel, S. V., and Vos, M. D. (2015). Automated respiration detection from neonatal video data. Proceedings of the 4th International Conference on Pattern Recognition Applications and Methods, pages 164–169.

LaGasse, L. L., Neal, A. R., and Lester, B. M. (2005). Assessment of infant cry: Acoustic cry analysis and parental perception. Mental Retardation and Developmental Disabilities Research Reviews, 11(1):83–93.

Lavner, Y., Cohen, R., Ruinskiy, D., and IJzerman, H. (2016). Baby cry detection in domestic environment using deep learning. In 2016 ICSEE International Conference on the Science of Electrical Engineering, pages 1–5. IEEE.

Lederman, D., Zmora, E., Hauschildt, S., Stellzig-Eisenhauer, A., and Wermke, K. (2008). Classification of cries of infants with cleft-palate using parallel hidden markov models. Medical & Biological Engineering & Computing, 46(10):965–975.

Lee, A., Kawahara, T., and Shikano, K. (2001). Julius - an open source real-time large vocabulary recognition engine. In Proc. European Conference on Speech Communication and Technology (EUROSPEECH), pages 1691–1694.

Lester, B. M. (1976). Spectrum analysis of the cry sounds of well-nourished and malnourished infants. Child Development, pages 237–241.

Lester, B. M. and Dreher, M. (1989). Effects of marijuana use during pregnancy on newborn cry. Child Development, pages 765–771.

Long, X., van der Sanden, E., Prevoo, Y., ten Hoor, L., den Boer, S., Gelissen, J., Otte, R., and Zwartkruis-Pelgrim, E. (2018). An efficient heuristic method for infant in/out of bed detection using video-derived motion estimates. Biomedical Physics & Engineering Express, 4(3):035035.

Manfredi, C., Bandini, A., Melino, D., Viellevoye, R., Kalenga, M., and Orlandi, S. (2018). Automated detection and classification of basic shapes of newborn cry melody. Biomedical Signal Processing and Control, 45:174–181.

Manfredi, C., Bocchi, L., Orlandi, S., Spaccaterra, L., and Donzelli, G. P. (2009). High-resolution cry analysis in preterm newborn infants. Medical Engineering & Physics, 31(5):528–32.

Markel, J. (1972). The SIFT algorithm for fundamental frequency estimation. IEEE Transactions on Audio and Electroacoustics, 20(5):367–377.

Marschik, P. B., Pokorny, F. B., Peharz, R., Zhang, D., O'Muircheartaigh, J., Roeyers, H., Bölte, S., Spittle, A. J., Urlesberger, B., Schuller, B., et al. (2017). A novel way to measure and predict development: A heuristic approach to facilitate the early detection of neurodevelopmental disorders. Current Neurology and Neuroscience Reports, 17(43):1–15.

Mazzone, L., Mugno, D., and Mazzone, D. (2004). The general movements in children with down syndrome. Early Human Development, 79(2):119–30.

Michelsson, K., Järvenpää, A., and Rinne, A. (1983). Sound spectrographic analysis of pain cry in preterm infants. Early Human Development, 8(2):141–149.

Michelsson, K. and Michelsson, O. (1999). Phonation in the newborn, infant cry. International Journal of Pediatric Otorhinolaryngology, 49 Suppl 1:S297–301.

Mizrahi, E. M. and Kellaway, P. (1987). Characterization and classification of neonatal seizures. Neurology, 37(12):1837–44.

Morielli, A., Ladan, S., Ducharme, F. M., and Brouillette, R. T. (1996). Can sleep and wakefulness be distinguished in children by cardiorespiratory and videotape recordings? Chest, 109(3):680–7.

Naithani, G., Kivinummi, J., Virtanen, T., Tammela, O., Peltola, M. J., and Leppänen, J. M. (2018). Automatic segmentation of infant cry signals using hidden Markov models. EURASIP Journal on Audio, Speech, and Music Processing, 2018(1):1–14.

Ntonfo, G. M. K., Ferrari, G., Raheli, R., and Pisani, F. (2012). Low-complexity image processing for real-time detection of neonatal clonic seizures. IEEE Transactions on Information Technology in Biomedicine, 16(3):375–382.

Olsen, M. D., Herskind, A., Nielsen, J. B., and Paulsen, R. R. (2014). Model-based motion tracking of infants. In European Conference on Computer Vision, pages 673–685. Springer.

Orlandi, S., Bandini, A., Fiaschi, F., and Manfredi, C. (2017). Testing software tools for newborn cry analysis using synthetic signals. Biomedical Signal Processing and Control, 37:16–22.

Orlandi, S., Bocchi, L., Donzelli, G., and Manfredi, C. (2012a). Central blood oxygen saturation vs crying in preterm newborns. Biomedical Signal Processing and Control, 7(1):88–92.

Orlandi, S., Dejonckere, P. H., Schoentgen, J., Lebacq, J., Rruqja, N., and Manfredi, C. (2013). Effective pre-processing of long term noisy audio recordings: An aid to clinical monitoring. Biomedical Signal Processing and Control, 8(6):799–810.

Orlandi, S., Garcia, C. A. R., Bandini, A., Donzelli, G., and Manfredi, C. (2016). Application of pattern recognition techniques to the classification of full-term and preterm infant cry. Journal of Voice, 30(6):656–663.

Orlandi, S., Guzzetta, A., Bandini, A., Belmonti, V., Barbagallo, S. D., Tealdi, G., Mazzotti, S., Scattoni, M. L., and Manfredi, C. (2015). AVIM - A contactless system for infant data acquisition and analysis: Software architecture and first results. Biomedical Signal Processing and Control, 20:85–99.

Orlandi, S., Manfredi, C., Bocchi, L., and Scattoni, M. (2012b). Automatic newborn cry analysis: A non-invasive tool to help autism early diagnosis. In Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, pages 2953–2956. IEEE.

Orozco-García, J. and Reyes-García, C. A. (2003). A study on the recognition of patterns of infant cry for the identification of deafness in just born babies with neural networks. In Iberoamerican Congress on Pattern Recognition, pages 342–349. Springer.

Ortiz, S. D. C., Beceiro, D. I. E., and Ekkel, T. (2004). A radial basis function network oriented for infant cry classification. In Iberoamerican Congress on Pattern Recognition, pages 374–380. Springer.

Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1):62–66.

Pamula, Y., Campbell, A., Coussens, S., Davey, M., Griffiths, M., Martin, J., Maul, J., Nixon, G., Sayers, R., Teng, A., et al. (2011). ASTA/ASA addendum to the AASM guidelines for the recording and scoring of paediatric sleep. In Journal of Sleep Research, volume 20, pages 4–4. Wiley-Blackwell Publishing.

Pearce, S. and Taylor, B. (1993a). Energy distribution in the spectrograms of the cries of normal and birth asphyxiated infants. Physiological Measurement, 14:263–68.

Pearce, S. and Taylor, B. (1993b). Time-frequency analysis of infant cry: Measures that identify individuals. Physiological Measurement, 14:253–62.

Pediaditis, M., Tsiknakis, M., and Leitgeb, N. (2012). Vision-based motion detection, analysis and recognition of epileptic seizures–a systematic review. Computer Methods and Programs in Biomedicine, 108(3):1133–48.

Pokorny, F. B., Bartl-Pokorny, K. D., Einspieler, C., Zhang, D., Vollmann, R., Bölte, S., Gugatschka, M., Schuller, B. W., and Marschik, P. B. (2018). Typical vs. atypical: Combining auditory Gestalt perception and acoustic analysis of early vocalisations in Rett syndrome. Research in Developmental Disabilities, 82:109–119.

Pokorny, F. B., Peharz, R., Roth, W., Zöhrer, M., Pernkopf, F., Marschik, P. B., and Schuller, B. W. (2016). Manual versus automated: The challenging routine of infant vocalisation segmentation in home videos to study neuro (mal) development. In Interspeech, pages 2997–3001.

Poppe, R. (2007). Vision-based human motion analysis: An overview. Computer Vision and Image Understanding, 108:4–18.

Porée, F., Simon, A., Cabon, S., Corolleur, A., Nardi, N., Pladys, P., and Carrault, G. (2015). Traitement de vidéos de polysomnographie pour l'estimation de l'état des yeux chez le nouveau-né prématuré. In XXVe Colloque GRETSI, pages 1–4, Eyes.

Prechtl, H. F. (1990). Qualitative changes of spontaneous movements in fetus and preterm infant are a marker of neurological dysfunction. Early Human Development, 23(3):151–8.

Prechtl, H. F., Einspieler, C., Cioni, G., Bos, A. F., Ferrari, F., and Sontheimer, D. (1997). An early marker for neurological deficits after perinatal brain lesions. Lancet, 349(9062):1361–3.

Raboshchuk, G., Nadeu, C., Jančovič, P., Lilja, A. P., Köküer, M., Mahamud, B. M., and De Veciana, A. R. (2018a). A knowledge-based approach to automatic detection of equipment alarm sounds in a neonatal intensive care unit environment. IEEE journal of Translational Engineering in Health and Medicine, 6:1–10.

Raboshchuk, G., Nadeu, C., Pinto, S. V., Fornells, O. R., Mahamud, B. M., and de Veciana, A. R. (2018b). Pre-processing techniques for improved detection of vocalization sounds in a neonatal intensive care unit. Biomedical Signal Processing and Control, 39:390–395.

Rahmati, H., Aamo, O. M., Stavdahl, Ø., Dragon, R., and Adde, L. (2014). Video-based early cerebral palsy prediction using motion segmentation. In Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE, pages 3779–3783. IEEE.

Rahmati, H., Dragon, R., Aamo, O. M., Adde, L., Stavdahl, Ø., and Van Gool, L. (2015). Weakly supervised motion segmentation with particle matching. Computer Vision and Image Understanding, 140:30–42.

Rechtschaffen, A. and Kales, A. (1968). A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Los Angeles: UCLA Brain Information Service/Brain Research Institute.

Reggiannini, B., Sheinkopf, S. J., Silverman, H. F., Li, X., and Lester, B. M. (2013). A flexible analysis tool for the quantitative acoustic assessment of infant cry. Journal of Speech, Language, and Hearing Research, 56(5):1416–1428.

Reyes-Galaviz, O. F., Tirado, E. A., and Reyes-Garcia, C. A. (2004). Classification of infant crying to identify pathologies in recently born babies with anfis. In International Conference on Computers for Handicapped Persons, pages 408–415. Springer.

Reyes-Galaviz, O. F., Verduzco, A., Arch-Tirado, E., and Reyes-García, C. A. (2005). Analysis of an infant cry recognizer for the early identification of pathologies. In Nonlinear Speech Modeling and Applications, pages 404–409. Springer.

Rosales-Pérez, A., Reyes-García, C. A., Gonzalez, J. A., Reyes-Galaviz, O. F., Escalante, H. J., and Orlandi, S. (2015). Classifying infant cry patterns by the genetic selection of a fuzzy model. Biomedical Signal Processing and Control, 17:38–46.

Runefors, P., Arnbjörnsson, E., Elander, G., and Michelsson, K. (2000). Newborn infants' cry after heel-prick: Analysis with sound spectrogram. Acta Paediatrica, 89(1):68–72.

Sami, A., Karayiannis, N. B., Frost, J. D., Wise, M. S., and Mizrahi, E. M. (2004). Automated tracking of multiple body parts in video recordings of neonatal seizures. Building, pages 312–315.

Schönweiler, R., Kaese, S., Möller, S., Rinscheid, A., and 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.

Sheinkopf, S. J., Iverson, J. M., Rinaldi, M. L., and Lester, B. M. (2012). Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder. Autism Research, 5(5):331–339.

Shimizu, A., Ishii, A., and Okada, S. (2017). Monitoring preterm infants' body movement to improve developmental care for their health. In Consumer Electronics (GCCE), 2017 IEEE 6th Global Conference on, pages 1–5. IEEE.

Shinya, Y., Kawai, M., Niwa, F., and Myowa-Yamakoshi, M. (2014). Preterm birth is associated with an increased fundamental frequency of spontaneous crying in human infants at term-equivalent age. Biology Letters, 10(8).

Sikdar, A., Behera, S. K., Dogra, D. P., and Bhaskar, H. (2015). Contactless vision-based pulse rate detection of infants under neurological examinations. In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 650–653. IEEE.

Sivan, Y., Kornecki, A., and Schonfeld, T. (1996). Screening obstructive sleep apnoea syndrome by home videotape recording in children. European Respiratory Journal, 9(10):2127–31.

So, K., Buckley, P., Adamson, T. M., and Horne, R. S. C. (2005). Actigraphy correctly predicts sleep behavior in infants who are younger than six months, when compared with polysomnography. Pediatric Research, 58:761–765.

Spittle, A. J., Brown, N. C., Doyle, L. W., Boyd, R. N., Hunt, R. W., Bear, M., and Inder, T. E. (2008). Quality of general movements is related to white matter pathology in very preterm infants. Pediatrics, 121(5):e1184–9.

Stahl, A., Schellewald, C., Stavdahl, O., Aamo, O. M., Adde, L., and Kirkerod, H. (2012). An optical flow-based method to predict infantile cerebral palsy. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 20(4):605–14.

Stevens, B. J., Johnston, C. C., and Horton, L. (1994). Factors that influence the behavioral pain responses of premature infants. Pain, 59(1):101–9.

Suaste-Rivas, I., Reyes-Galaviz, O. F., Diaz-Mendez, A., and Reyes-Garcia, C. A. (2004). A fuzzy relational neural network for pattern classification. In Iberoamerican Congress on Pattern Recognition, pages 358–365. Springer.

Sun, D., Roth, S., and Black, M. J. (2014). A quantitative analysis of current practices in optical flow estimation and the principles behind them. International Journal of Computer Vision, 106(2):115–137.

Sung, M., Adamson, T. M., and Horne, R. S. C. (2009). Validation of actigraphy for determining sleep and wake in preterm infants. Acta Paediatrica, 98:52–57.

Tenold, J. L., Crowell, D. H., Jones, R. H., Daniel, T. H., McPherson, D. F., and Popper, A. N. (1974). Cepstral and stationarity analyses of full-term and premature infants' cries. The Journal of the Acoustical Society of America, 56(3):975–80.

Thach, B. T. (2007). Maturation of cough and other reflexes that protect the fetal and neonatal airway. Pulmonary Pharmacology & Therapeutics, 20(4):365–370.

Tharp, B. R. (2002). Neonatal seizures and syndromes. Epilepsia, 43 Suppl 3:2–10.

Thodén, C.-J., Järvenpää, A.-L., and Michelsson, K. (1985). Sound spectrographic cry analysis of pain cry in prematures. In Infant Crying, pages 105–117. Springer.

Torres, R., Battaglino, D., and Lepauloux, L. (2017). Baby cry sound detection: A comparison of hand crafted features and deep learning approach. In International Conference on Engineering Applications of Neural Networks, pages 168–179. Springer.

van Gastel, M., Balmaekers, B., Oetomo, S. B., and Verkruysse, W. (2018). Near-continuous non-contact cardiac pulse monitoring in a neonatal intensive care unit in near darkness. In Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics, volume 1050114, pages 1–9. International Society for Optics and Photonics.

Várallyay, G. (2006). Future prospects of the application of the infant cry in the medicine. Periodica Polytechnica Electrical Engineering, 50(1-2):47–62.

Várallyay, G. (2007). The melody of crying. International Journal of Pediatric Otorhinolaryngology, 71(11):1699–1708.

Várallyay, G., Benyó, Z., Illényi, A., Farkas, Z., and Kovács, L. (2004). Acoustic analysis of the infant cry: Classical and new methods. In Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE, volume 1, pages 313–316. IEEE.

Verduzco-Mendoza, A., Arch-Tirado, E., Reyes-García, C. A., Leybon-Ibarra, J., and Licona-Bonilla, J. (2012). Spectrographic cry analysis in newborns with profound hearing loss and perinatal high-risk newborns. Cirugia y Cirujanos, 80(1):3–10.

Villarroel, M., Guazzi, A., Jorge, J., Davis, S., Watkinson, P., Green, G., Shenvi, A., McCormick, K., and Tarassenko, L. (2014). Continuous non-contact vital sign monitoring in neonatal intensive care unit. Healthcare Technology Letters, 1(3):87–91.

Wahid, N., Saad, P., and Hariharan, M. (2016). Automatic infant cry pattern classification for a multiclass problem. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 8(9):45–52.

Wasz-Höckert, O., Michelsson, K., and Lind, J. (1985). Twenty-five years of Scandinavian cry research. In Infant Crying, pages 83–104. Springer.

Wermke, K. and Mende, W. (2009). Musical elements in human infants' cries: In the beginning is the melody. Musicae Scientiae, 13(2_suppl):151–175.

Wermke, K., Mende, W., Manfredi, C., and Bruscaglioni, P. (2002). Developmental aspects of infant's cry melody and formants. Medical Engineering & Physics, 24(7-8):501–14.

Yamamoto, S., Yoshitomi, Y., Tabuse, M., Kushida, K., and Asada, T. (2013). Recognition of a baby's emotional cry towards robotics baby caregiver. International Journal of Advanced Robotic Systems, 10(2):86.

Zaker, N., Mahoor, M. H., Mattson, W. I., Messinger, D. S., and Cohn, J. F. (2013). A comparison of alternative classifiers for detecting occurrence and intensity in spontaneous facial expression of infants with their mothers. 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013, 12.

Zaker, N., Mahoor, M. H., Messinger, D. S., and Cohn, J. F. (2014). Jointly detecting infants' multiple facial action units expressed during spontaneous face-to-face communication. 2014 IEEE International Conference on Image Processing (ICIP), 80208:1357–1360.

Zamzami, G., Ruiz, G., Goldgof, D., Kasturi, R., Sun, Y., and Ashmeade, T. (2015). Pain assessment in infants: Towards spotting pain expression based on infants' facial strain. In Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on, volume 5, pages 1–5. IEEE.

Zeskind, P. S. (1980). Adult responses to cries of low and high risk infants. Infant Behavior and Development, 3:167–177.

Zeskind, P. S. and Lester, B. M. (1978). Acoustic features and auditory perceptions of the cries of newborns with prenatal and perinatal complications. Child Development, pages 580–589.

Zeskind, P. S. and Marshall, T. R. (1988). The relation between variations in pitch and maternal perceptions of infant crying. Child Development, pages 193–196.

Zeskind, P. S., Parker-Price, S., and Barr, R. G. (1993). Rhythmic organization of the sound of infant crying. Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology, 26(6):321–333.