Biomedical Signal Processing and Control 45 (2018) 174–181

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Automated detection and classification of basic shapes of newborn cry melody

Claudia Manfredi [a,*], Andrea Bandini [b], Donatella Melino [a], Renaud Viellevoye [c], Masendu Kalenga [c], Silvia Orlandi [a,d]

[a] Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy
[b] University Health Network, Toronto Rehabilitation Institute, Toronto, ON, Canada
[c] Service Universitaire de Néonatologie CHR de la Citadelle, Liege, Belgium
[d] Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, ON, Canada

ARTICLE INFO

Article history:
Received 2 August 2017
Received in revised form 30 April 2018
Accepted 28 May 2018

Keywords:
Newborn infant cry melody
Automated analysis and classification
Early diagnosis of neurological impairment
Intensive care management
Autism spectrum disorders
Pre-speech development

ABSTRACT

The study of newborn cry is a promising non-intrusive and cheap approach to support the early diagnosis of neurodevelopmental disorders. Specifically, cry melody, the trend of the fundamental frequency (f0) over time, could add relevant information to the acoustical analysis of infant crying. To date, the cry analysis is mainly performed by paediatricians/neurologists through a perceptual examination based on listening to the cry and visually inspecting the f0 shape. Therefore, this approach is not widespread as the procedure is operator-dependent and requires a considerable amount of time often prohibitive in daily clinical practice.

This paper aims at providing a support to the perceptual analysis through a fully automated method for assessing the melodic shape of newborn cry. Cry units are detected within each recording, even of long duration, and their classification is performed according to five basic melodic shapes (falling, rising, symmetrical, plateau, and complex).

The method is tested on synthesized signals and applied to recordings coming from at term healthy newborns. Results are compared to the perceptual analysis performed by trained raters with up to 98% matching.

Being contact-less and cheap, this method is well suited for routinely clinical applications and could be effectively related to other clinical parameters for early detection of possible brain injuries or neurodevelopmental disorders.

1. Introduction

Infant cry, the way of communication of the baby in the first months of life, is a maturational process modulated by complex cerebral networks involving the midbrain, the cingulate gyrus and specific cortical areas. Therefore, it has been studied for decades to support the screening and diagnosis of possible disorders of the central nervous system [1–3]. Preliminary studies have linked cry analysis with neurodevelopmental disorders showing differences in fundamental frequency and duration of cry episodes between typically developing and autistic children [4–6]. Moreover, the infant cry melody (f0 trend) can discriminate babies with neurological diseases or birth defects such as cleft palate [7,8].

The study of the infant cry is particularly attractive being a method of clinical investigation totally non-intrusive and cheap, based solely on listening to the audio signal and visually inspecting its shape and spectrogram. However, the perceptual analysis requires highly trained clinicians. Moreover, listening to the crying may lead to incomplete and sometimes conflicting clinical indications as it might be performed by different clinicians, under non-optimal conditions and in a short period of time. This problem can be overcome by recording the crying which can be listened to and interpreted later. Recordings lasting from few seconds up to several minutes typically consist of dozens and dozens of "significant events", hereafter referred to as "cry units" (CUs). Typically, CUs are high-energy voiced frames of the signal, but they also include moans, breath, and background noise. The time required for the manual selection of CUs often conflicts with the frantic working schedule of clinicians. Moreover, to date there are no standards

* Corresponding author.
E-mail address: claudia.manfredi@unifi.it (C. Manfredi).

https://doi.org/10.1016/j.bspc.2018.05.033
1746-8094/

for the cry analysis, making the objective assessment of neonatal crying features very hard.

Since the early studies the need for dedicated software tools that could provide automatically the main parameters of infant cry was highlighted. However, the neonatal cry is a signal extremely difficult to analyze due to its great variability and irregularity as well as to the very high range of frequencies of interest which require sophisticated numerical techniques based on high time-frequency resolution [9–18]. To this aim, recently, some specific software tools have been developed [19–21].

To date, the most relevant clinical parameter is the fundamental frequency (f0), which reflects the regularity of the vibration of the vocal folds of the newborn, along with its melodic shape (the trend of f0 over time). Many studies defined the f0 variability range between 200 and 800 Hz (beyond 1000 Hz for pathological cries) [12,18]. Automated methods are based on the detection of CUs and the estimation of several acoustical parameters such as f0 and formant frequencies [10,11,16,17,20,30–39]. The huge number of papers dealing with the estimation of f0 show how this parameter is still the subject of extensive research, however few papers attempt to assess the melodic parameters of cry and propose methods for their automated detection and classification. They aim at highlighting linguistic differences and specific cry features in infants with diseases [22–29].

The "shape" of the melody can be classified into several categories. Schönweiler et al. [29], first introduced the four main melody shapes of f0:

- The falling (F) curve consists of a rapid rise of f0 in the first part of the time interval followed by a slow descent to the end (Fig. 1a).
- The rising (R) curve has a pattern symmetrical to F: a slow climb of f0 up to the last part of the interval, where there is a steep decrease to the end of the CU (Fig. 1b).
- The rising—falling (also defined symmetrical – S) curve is described by an increasing-decreasing trend of f0 around a maximum approximately positioned in the middle of the CU (Fig. 1c).
- A flat curve (also called plateau – P) is described by a nearly constant f0 with a variation lower than few tens of Hz around its median value (Fig. 1d).

These categories do not cover all the variations of melody shapes. Várallyay [22] found up to 77 different shapes, among which 20 represent the 95% of melodies.

The purpose of this paper is to present a fully automated method for melody classification to support the perceptual analysis of neonatal cry. For a reliable melody classification, a novel method to obtain f0 patterns sufficiently regular (i.e., devoid of outliers) is presented. The classification is performed considering 5 melodic shapes: the main shapes (R, F, S, and P) described above and the complex shape (C) [22,23], which consists in the concatenation of two or more R, F and S melodic shapes (Fig. 1e). The method is tested on synthesized signals made up by a set of basic melody shapes and applied to recordings coming from healthy newborns. To test the accuracy of the method a comparison between automated and perceptual analysis made by trained raters is performed.

2. Materials and methods

2.1. Cry units detection

The first problem addressed is the automated CUs detection in the whole recording. This is a challenging problem as the commonly used manual approach requires a long time to be performed, prohibitive for clinicians. Thus, only short parts of a whole recording are usually considered, that is those that look "most significant" from the perceptual point of view, disregarding most of the signal and thus the information inside it. To overcome this problem, the automated method implemented in BioVoice [20,30–32] is applied here. BioVoice is a user-friendly voice analysis software tool developed in Matlab. It implements parametric (AutoRegressive) techniques specifically developed to face highly irregular/non-stationary signals such as newborn infant cry. The method for CUs detection is based on the short-term energy that increases during "cry" events and decreases during "unwanted" episodes. An iterative method is applied to the varying energy histogram, based on time-varying upper and lower energy thresholds that was proven successful to avoid incorrect splitting of a single "event" into several ones. The procedure and the positive results obtained with newborn cry are described in detail in [36]. At the end of this step, all the CUs in the recorded signal are detected, and their length, starting and ending points are stored for further elaboration. Finally, CUs with duration shorter than 260 ms are discarded. In fact, too short episodes could be misclassified as CUs rather than air inhalation episodes [8,23].

2.2. f0 estimation

In the clinical practice, newborn cry is commonly recorded in a not soundproof environment, therefore noise reduction is required. In this work the recorded signal is band-pass filtered by a Kaiser Window FIR (Finite Impulse Response) filter with cut-off frequencies of 150–1050 Hz.

The f0 often shows a highly varying shape within each CU, with a range of variation of 300–400 Hz or even more between 200 Hz and 800 Hz (excluding hyper-phonated cry) [12,18]. In this work BioVoice is used for f0 estimation [20,30–32]. Its strength is in the use of parametric autoregressive models that have higher resolution with respect to the classical fast Fourier approach. In BioVoice the model order varies adaptively according to the variability of the signal: the higher the f0, the shorter the length of the analysis window [11,16,17,20,30].

2.3. f0 smoothing

Newborn CUs exhibit highly varying f0 values that may adversely affect any automated method of melody classification. Therefore, this step consists in obtaining a smoothed f0 shape, i.e., devoid of öutliers. To achieve this aim, we resort to adjustments as much as possible close to what is done by our visual system to detect a shape from a distribution of points even very irregular. To smooth f0 three steps are applied to each detected CU. The thresholds in steps 2 and 3 stem from a thorough visual analysis and tests made on a dataset of infant cries coming from 50 newborns [32,35]. Three independent raters performed the perceptual analysis of f0 shapes selecting 100 CUs (20 plateau, 20 rising, 20 falling, 20 symmetric and 20 complex) that allowed setting the thresholds for the automated analysis.

1 – Apply the Tukey method [40]. Tukey defines the lower quartile (q1) as the 25^{th} percentile of the data and the upper quartile (q3) as the 75^{th} percentile of the data. The difference between q1 and q3 is the inter-quartile difference (IQR). An outlier is defined as an f0 value below the lower threshold LT or above the upper threshold UT defined as:

$$LT = q1 - 3 * IQR$$

$$UT = q3 + 3 * IQR$$

The usefulness of this method is that it makes no assumption on the distribution of data nor depends on mean or standard deviation (STD) values.