Biomedical Signal Processing and Control 19 (2015) 35–43

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Automatic detection of the expiratory and inspiratory phases in newborn cry signals

Lina Abou-Abbas*, Hesam Fersaie Alaie, Chakib Tadj

Electrical Engineering Department, École de Technologie Supérieure, Montreal, QC H3C 1K3, Canada

ARTICLE INFO

Article history:
Received 31 July 2014
Received in revised form 12 March 2015
Accepted 13 March 2015
Available online 3 April 2015

Keywords:
HMM
Automatic segmentation
Newborn cry signals
Mel Frequency Cepstral Coefficients
Viterbi algorithm
Baum Welch algorithm

ABSTRACT

An analysis of newborn cry signals, either for the early diagnosis of neonatal health problems or to determine the category of a cry (e.g., pain, discomfort, birth cry, and fear), requires a primary and preliminary preprocessing step to quantify the important expiratory and inspiratory parts of the audio recordings of newborn cries. Data typically contain clean cries interspersed with sections of other sounds (generally, the sounds of speech, noise, or medical equipment) or silence. The purpose of signal segmentation is to differentiate the important acoustic parts of the cry recordings from the unimportant acoustic activities that compose the audio signals. This paper reports on our research to establish an automatic segmentation system for newborn cry recordings based on Hidden Markov Models using the HTK (Hidden Markov Model Toolkit). The system presented in this report is able to detect the two basic constituents of a cry, which are the audible expiratory and inspiratory parts, using a two-stage recognition architecture. The system is trained and tested on a real database collected from normal and pathological newborns. The experimental results indicate that the system yields accuracies of up to 83.79%.

1. Introduction

With early newborn screening, a serious illness can be diagnosed such that treatment can begin before severe problems appear, and in certain cases, sudden mortality or disability can be prevented. Clearly, the presence of disease must be detected at an early stage. Systematic screening combined with better diagnostic tools is therefore required to meet future medical challenges, with the aim of supporting clinical decision-making and improving the effectiveness of treatment [1]. These tools have evolved considerably in recent years in terms of improving screening and symptom evaluation, and the newborn cry signal has been the object of strong research interest for the past three decades.

Researchers have amassed enough evidence to conclude that a cry signal contains relevant information on the psychological and physiological condition of the newborn, formal relationships have been established between the acoustic features extracted from the cries and the health problems of the child [2–5]. Various studies are currently under way to devise a tool that analyzes cries automatically, to diagnose neonatal pathologies [6–8].

We are involved in the design of an automatic system for early diagnosis, called the Newborn Cry-based Diagnostic System (NCDS), which can detect certain pathologies in newborns at an early stage. The implementation of this system requires a database containing hundreds of cry signals.

The overwhelming problem that arises when working with such a database is the diversity of acoustic activities that compose the audio recordings, such as background noise, speech, the sound of medical equipment and silence. Such diversity could harm the analysis process, as the presence of any acoustic component other than the cry itself could result in the misclassification of pathologies by reducing the NCDS system performance. This is because the NCDS would decode every segment of the recording signal, whether it is part of a cry or not. In this case, unwanted segment insertion in essential crying segments would lengthen the process of classification unnecessarily and leave the system prone to error. An important subtask of the NCDS is the manipulation of the newborn cry sound, and what is needed to perform this subtask is a segmentation system. Until now, few works have been carried out in this area. In this paper, we propose an automatic segmentation module designed to isolate the audible expiration and inspiration parts of cry sounds to serve as a preprocessing step of our NCDS.

The rest of this paper is organized as follows: Related work is presented in Section 2. The HMM and the HTK are reviewed briefly in Section 3. The training corpus and the testing corpus are described in Section 4. In Section 5, the architecture of the

* Corresponding author. Tel.: +1 5144736060.
E-mail addresses: Lina.Abou-Abbas.1@etsmtl.net (L. Abou-Abbas), Hesam.farsaie-alaie@etsmtl.net (H. Fersaie Alaie), Chakib.Taj@etsmtl.ca (C. Tadj).

http://dx.doi.org/10.1016/j.bspc.2015.03.007
1746-8094/

proposed system is presented, and details of the individual blocks are described in five subsections. Section 6 contains the implementation of the system, the obtained results, and the discussion. Finally our conclusions are presented in Section 7.

2. Related work

Several studies have been conducted in which the infant cry is analyzed (categorization of the cry, disease classification based on the cry). In 1985, for example, Corwin and Golub outlined four acoustic categories composing a cry episode, which are: (a) expiratory phonation (with F0 ranging from 250 to 750 Hz), (b) expiratory hyperphonation (with F0 ranging from 1000 to 2000 Hz), (c) expiratory dysphonation (aperiodic expiratory segment), (d) inspiratory phonation (associated with any perceptually audible sound generated by the newborn during inspiration, or high-pitched cries during inspiration) [2,9].

In most studies, the cry segmentation phase was performed manually, a human operator was asked to monitor the recorded audio signals and pick out only the important cry parts from the recordings [3,10,11]. This manual task is tiresome and too time-consuming when the volume of data is large. The cry segmentation that serves the needs of a real-time diagnostic tool should be performed automatically.

In some studies, the authors have applied various voice activity detection software approaches such as the traditional methods of ZCR (Zero Crossing Rate) and STE (Short Time Energy), with some modification of the thresholds [4,5,12,13]. In general, these methods are of limited use in this context, as speech and cry sounds have different features. With these methods, particularly in the search for the high-energy parts of the audio signals, not only are the meaningful parts of cry vocalizations found but also background noise, speech, and machine sounds. In other words, the typical voice activity detection methods alone are not suitable for segmenting a cry signal. The corpuses used to examine these methods (ZCR, STE) were composed only of cry sounds, which are sequences of expiration and inspiration, alternating with short periods of silence and background noise. The main goal of authors was to eliminate silence and background noise without affecting the audible expiration and inspiration phases.

Few studies have been conducted specifically on the automatic segmentation of cry signals [14–17]. Two novel algorithms were introduced by modifying the Harmonic Product Spectrum (HPS) method [14]. The HPS method was created to detect the fundamental frequency of an audio signal. The authors showed that it is possible to check the regularity structure of the spectrum using the HPS method and classify its content by detecting the meaningful parts of the cry sounds. Another study on the segmentation of cry signals was conducted by Cohen [16] with the purpose of labeling each successive segment as a cry/non-cry/non-activity. However, with the methods presented in [16], the inspiration parts as well as the dysphonic vocalizations of the cry spectrum that could be presented with irregular or non-harmonic structure were ignored.

Recent studies have shown that differentiated characteristics in expiratory and inspiratory vocalizations exist in adults as well in newborns [18].

Assuming that the inspiratory phase of a cry episode reflects a laryngeal contraction of the ingressive airstream, inspiratory vocalization has been proven to be useful in the identification of newborns at risk for various health conditions [9]. In fact, the amount of time the inspiratory phase lasts in newborns with respiratory disorders is greater than it is in normal newborns [19]. Indeed, recent medical evidence confirms that a relationship exists between upper airway obstruction and sudden infant death syndrome and sleep apnea. Despite this evidence, it is surprising to find acoustic data that are limited to the expiratory phase alone [9].

To create an effective diagnostic tool based on the cry signals, the involvement of both the expiratory and the inspiratory components is a prerequisite. The aim of this study is to identify and quantify both the audible inspiratory and expiratory components of a newborn cry automatically.

The work presented in this paper is based on the well-established and widely used Hidden Markov Model (HMM) statistical technique, which has been successfully applied in automatic speech recognition and segmentation systems.

To the best of our knowledge, no work has yet been carried out on the automatic segmentation of crying signals recorded in noisy environments without manually pre-processing the signals to remove at least irrelevant acoustic activities, such as speech and beep sounds around the infant.

In recent work [17], authors applied an automatic segmentation approach based on a HMM classification tool to segment the expiratory and inspiratory sounds of cry signals. The difference with this recent approach compared to our approach is not only with the limited number of infants and the limited available acoustic activities types (due to the environment in which recordings are taking place) but also the way in which they applied the HMM. The authors of [17] considered only three classes, Expiration (EX), Inspiration (IN) and Silence (SI). As a first stage, to train each class, they used different techniques such as Support Vector Machines (SVM) as well as Gaussian Mixture Models (GMM) consisting of 5 and 20 Gaussian components. To reduce errors by taking into account the arrangement in time between the three classes, the authors added a second stage using the Viterbi algorithm. The whole architecture of the approach in [17] could be taken as an HMM architecture of three states. In fact, the segmentation approach presented in [17] performed well, but its performance needs to be enhanced to segment audio signals recorded in a noisy environment (e.g., sounds of speech, medical equipment, noise, and silence).

To provide a better understanding of the context of this study, some important terms used must be predefined.

- *Inspiration* is associated with inspiratory phonation as defined by Golub and Corwin [2].
- *Expiration* is referred to the acoustic output during the expiration phase of a cry (it can be phonation, dysphonation, or hyperphonation), as well as any audible expiration sound generated by the infant outside its cry episodes. Note that we do not make a distinction here between the expiration phases that occur during or following a cry.
- *A cry sequence* consists of long periods of expiratory crying separated by short inspiratory episodes.

We have avoided using the terms voiced inspiration and voiced expiration to describe the important parts of the cry. In fact, a dysphonation vocalization is characterized in earlier studies as an unvoiced part during a cry and it is considered one of the most useful vocalizations in the detection of newborns at risk of various health conditions [20]. For this reason, we prefer using the terms audible inspiration and audible expiration.

3. Hidden Markov Model and the HTK

HMMs underlie the most modern automatic speech recognition (ASR) systems. They have many potential applications in statistical signal processing and acoustic modeling, including the segmentation of recorded signals [21]. The basic principles of any ASR system involve constructing and manipulating a series of statistical models that represent the various acoustic activities of the sounds to