Expiratory and Inspiratory Cries Detection Using Different Signals' Decomposition Techniques

*Lina Abou-Abbas, *Chakib Tadj, *Christian Gargour, and †Leila Montazeri, *Montreal, and †Canada

Summary: This paper addresses the problem of automatic cry signal segmentation for the purposes of infant cry analysis. The main goal is to automatically detect expiratory and inspiratory phases from recorded cry signals. The approach used in this paper is made up of three stages: signal decomposition, features extraction, and classification. In the first stage, short-time Fourier transform, empirical mode decomposition (EMD), and wavelet packet transform have been considered. In the second stage, various set of features have been extracted, and in the third stage, two supervised learning methods, Gaussian mixture models and hidden Markov models, with four and five states, have been discussed as well. The main goal of this work is to investigate the EMD performance and to compare it with the other standard decomposition techniques. A combination of two and three intrinsic mode functions (IMFs) that resulted from EMD has been used to represent cry signal. The performance of nine different segmentation systems has been evaluated. The experiments for each system have been repeated several times with different training and testing datasets, randomly chosen using a 10-fold cross-validation procedure. The lowest global classification error rates of around 8.9% and 11.06% have been achieved using a Gaussian mixture models classifier and a hidden Markov models classifier, respectively. Among all IMF combinations, the winner combination is IMF3+IMF4+IMF5.

Key Words: automatic segmentation–empirical mode decomposition–wavelet packet transform–Gaussian mixture models–hidden Markov models.

INTRODUCTION

Crying is the only possible way for newborns to express their needs and their physical conditions because they are not able to communicate with words. Cry signals have been studied for many years, and it has become evident that cry signals can provide valuable information concerning physiological and physical states of infants. Most research on infant cry focused on extracting information from infant cry signals with known medical problems such as prematurity asphyxia, hypoglycemia, Down syndrome, and meningitis. For example, the cries of infants with neonatal asphyxia and meningitis are high-pitched, and the cry duration is very short or unusually long with melody type rising or falling-rising in comparison with healthy infants. Preterm babies have higher minimum fundamental frequency than normal babies. Cries of infants with hyperbilirubinemia have significant changes in fundamental frequency over a 100-ms period. For the reason of cries, features such as pitch and loudness are able to distinguish hunger cry from pain cry.[1–4]

Given these pieces of evidence, many researchers have suggested an automatic system to classify infant cries, which is more like a pattern recognition problem, similar to automatic speech recognition (ASR) systems. The aim of the automatic classification system is to give clinicians an early diagnostic result if a baby may have high probability to get specific types of medical diseases. As in any ASR system, a cry classification system needs a segmentation module that can detect useful parts of recorded signal and reject other acoustic activities to be thereafter classified.

Infant cry signals consist of a sequence of audible expiratory and inspiratory phases separated by a period of silence or by unvoiced phases of cry (inaudible expiratory and inspiratory phases during a cry). A cry signal recorded in a real environment usually contains different acoustic activities other than the cry, such as background noise, speech, sound of medical equipment, and silence. This work aims to retrieve most relevant and audible sections from cry signals recorded in a realistic clinical environment, as well as distinction between expiratory and inspiratory phases of the cries. One way to address this problem is to manually segment recorded audio signals and pick out important cry parts. However this manual task is tiresome and prone to errors when the volume of data is large. It is therefore essential to design a segmentation system able to automate this tedious task and be implemented in a real-time clinical decision support tool. Typical waveforms of a cry signal, expiratory phase, and inspiratory phase of cry signals are shown in Figures 1, 2, and 3, respectively.

Some attempts to segment cry signals have been reported in the literature. Many studies used the spectrogram to segment cry signals manually through visual and audio monitoring.[5] On one hand, automatic segmentation is often desired to manipulate all automated diagnostic systems, and on the other hand, because the manual segmentation is an extremely long, tedious task and is prone to errors especially when the amount of data is large. A number of recent works have been done on infant cry segmentation based on the time domain characteristics of the signal. The problem of cry segmentation was being considered as the problem of voice activity detection. Refs. 6, 7 used high-pass filter to reduce most of the background noise, and to distinguish between important and less important parts of the cry signals, they applied short-term energy or/and zero crossing rate by using a satisfactory threshold. However, these methods perform

Accepted for publication May 24, 2016.

From the *Electrical Engineering Department, École de Technologie Supérieure, Montreal, Canada; and the †Electrical Engineering Department, Polytechnique Montreal, Canada.

Address correspondence and reprint requests to Lina Abou-Abbas, Electrical Engineering Department, École de Technologie Supérieure, Montreal, Canada. E-mail: Lina.abou-abbas.1@etsmtl.net

Journal of Voice, Vol. 31, No. 2, pp. 259.e13–259.e28

0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.05.015

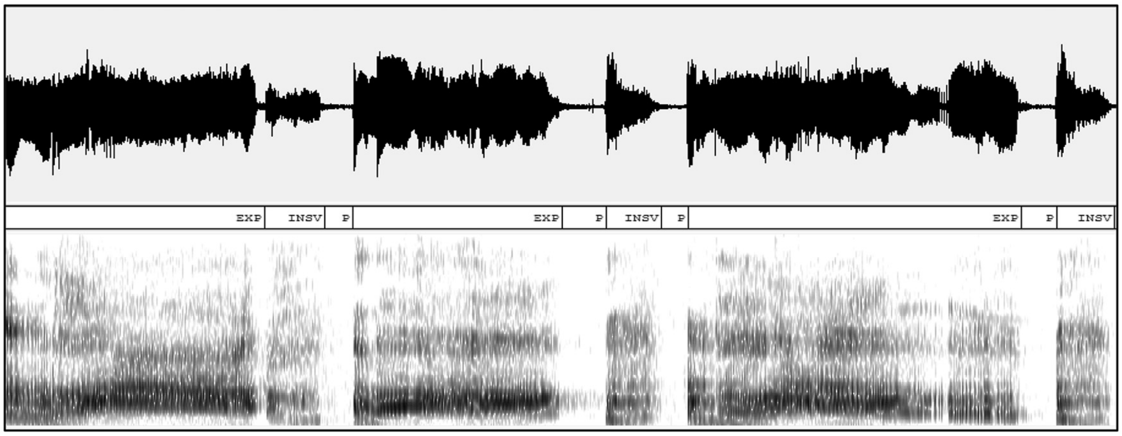

FIGURE 1. An example of a portion of cry signal with its corresponding components expiration (EXP), audible inspiration (INSV), and pauses (P).

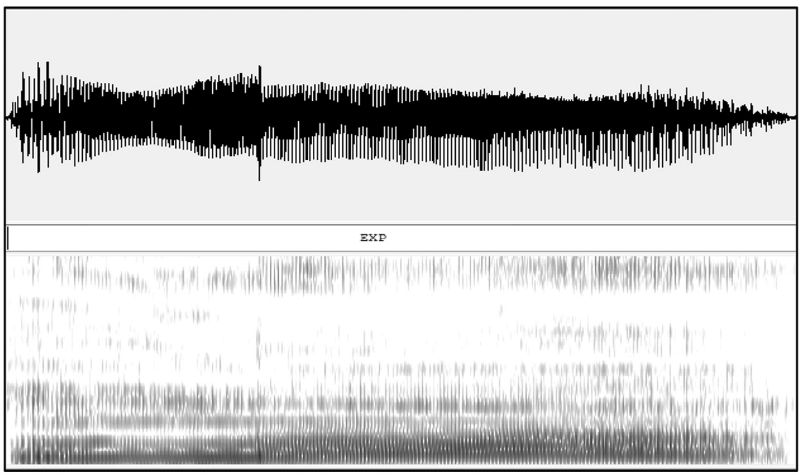

FIGURE 2. An example of a waveform and spectrogram of an expiration phase.

well when cries have been recorded within a laboratory environment and fail under noisy or clinical environment.

In other research efforts, the cry detection problem was considered as the problem of start and end points detection of a cry

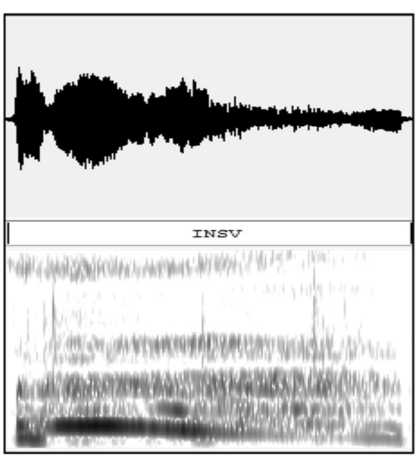

FIGURE 3. An example of a waveform and spectrogram of an inspiration phase.

unit. Based on the hypothesis that cry segments have four times more energy than unvoiced segments, authors in Refs. 8, 9 defined some guidelines to detect cry units based on a dynamic threshold for each record. In these works, authors eliminate not only useless sounds from the signals but also inspiratory sounds of the cry. Another technique used in Ref. 10 considers the problem of cry segmentation as the problem of Voiced/Unvoiced decision. In Ref. 10, authors modified a well-known fundamental frequency estimation method, the Harmonic product spectrum, to check the regularity of the segment analyzed to classify it as an important or not important cry part. In Ref. 11, authors used the simple inverse filter tracking algorithm to detect the voiced frames of cry based on a threshold of the autocorrelation function. The use of threshold limits the attractiveness of the mentioned approaches and decreases their performance in low signal-to-noise ratio levels.

Inspiratory cry parts have been proven to be important in identification of newborns at risk for various health conditions.[12] Despite this evidence, it is thus surprising that in most research analyzing cry signals, the inspiratory parts of a cry were ignored and not considered in the analysis and the main focus was only on extraction of acoustical data of expiratory parts.