ABSTRACT

Cry is a form of communication for children to express their feelings. Baby's cry can be characterized according to its natural periodic tone and the change of voice. Through their baby's cry detection, parents can monitor their baby remotely only in important conditions. Detection of a baby cry in speech signals is a crucial step in applications like remote baby monitoring and it is also important for scholars, who study the relation between baby cry signal patterns and other developmental parameters. This study of sound recognition involves feature extraction and classification by determining the sound pattern. We use MFCC as a feature extraction method and K-Nearest Neighbor (K-NN) for classification. K-Nearest Neighbor (KNN) is a classification method that is often used for audio data. The KNN classifier is shown to yield considerably better results compared to other classifiers.

Key Words*:* Signal patterns, developmental parameters, sound recognition, Feature extraction, MFCC, K-Nearest Neighbour (KNN), speech signal processing.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

Cry signals or cry patterns have been under research analysis for many years. Scholars and analysts found that the cry signals can give detailed pictures about the physical and mental states of newborns.

From the research of WHO, “every year, nearly 40% of infant deaths are happening in their initial 30-50 days of life, 72% of infant deaths happen in the first week of their birth, and up to 2/3 of infant lives can be saved if the cause is known much earlier.

The techniques that can allow us to identify the former signs of infant health and hygiene can help us reduce infant mortality. To be precise this is the superior goal of our thesis is to develop or implement a reliable system that allows us to know diseases based only on cry sound examination. Development of such a type of system initially mentions the problem in finding the reliable cry components or patterns in an input waveform. The NCDS system probably confuses if the input speech signal contains unnecessary noises other than the cry signal alone. Hence, the greatest challenge in designing and implementing a diagnostic system is to implement an automatic detecting machine to precisely search the inspiratory and expiratory parts of a cry pattern. After a lot of research on diseases and cry signals and their relation some useful results happened to develop automatic audio segmentation of expiratory and inspiratory parts of infant cries. If we are able to segment audio cry signals and examine vital parts of a pre-recorded sound signal, it would be very helpful and simpler to develop a completely automatic system which helps in understanding diseases. This system can definitely be used to support our decisions on understanding infant cries. Through this we can determine the symptoms earlier and take necessary steps at an efficient and cheap cost.

The recent studies on infant cries have shown that infants are crying for several reasons based on their requirements like hunger, fatigue, uncomfortable feelings, pain, and many other reasons. Researchers and scholars such as pediatricians health workers can relate between various types of infant cries and thus pre estimate the

infant's requirement by using the cry sounds, gestures and other behaviour produced by the infant however, this is a real time considerable problem (issue) for the parents not so skilled to those who can't take proper care of the baby.This project provides an Automatic method for infant cry classification which is trained using a data set of five different infant cries.

Hence the main objective is to extract useful features from the cry audio signal i.e The infant cry and test the unknown cry signal with the classified trainer and know the meaning of the infant cry, thereby taking care of the infant accordingly.

1.1 Need for baby cry classification

Crying is one of the major means of infants to communicate with their surroundings, and is intended to point out any distress and attachment needs to their caregivers. Automatic detection of a baby cry in audio signals can be used for various purposes – from every-day applications to academic research. Through this Detection on baby/infant cry sounds we can provide/take care of infant needs.

1.2 Problem Statement

If a system is developed to support our decisions on understanding the infant cries, we can determine the feelings and take necessary steps accordingly so that it would be a contributing factor in understanding their feelings and many of the infant lives can be saved if the cause is much known earlier.

2. LITERATURE SURVEY

2.1 Components of Infant cry audio signal:

The important and key components of an infant cry audio signal are inspiration and expiration parts with vocalization and hear-able inspiration(INSV) and expiration (EXP). The vital challenges faced in this type of system is implementing a method that can effectively search INSV, EXP exactly within a respective cry signal. The problem of cry detection is different from unvoiced, voiced segmentation because a typical hear-able infant cry audio signal contains each of the unvoiced and voiced parts.

2.2 Voice Activity Detection:

The literature proposed by Kuo, 2010, has stated that the main problem is detection of cry sound using a system recorded in a lot of noisy domestic local environment is not too easy to be solved by VAD (Voice Activity Detection) modules, VAD deals with the problem of searching or finding speech patterns from other auditory active regions of a considered audio signal. The other auditory active patterns may be any type like silence, noise or a doorbell warning. The Signal to Noise Ratio “SNR " is a key parameter and it might result in a lot of unwanted errors. VAD is vital in several audio communication systems like automatic speech recognition, telephones, other digital resources, and transmission of speech in real time. Some of the Common and very widely used VAD methods contain two basic and important methods: Feature Extraction and Decision making. Features of a signal that allows computation of energy, cepstral coefficients, ZCR, Marzinzik and Kollmeier, proposed spectrum analysis in 2002, Wang and Tasi, proposed wavelet ad entropy transforms in 2008 ; Juang et al., proposed decision rule computation based on frame-by-frame and very simple rules for thresholding. In 2009, After applying familiar VAD algorithms from Rabiner-Sambur and G.729b method in detection of cry signal parts or segments.

2.3 Findings:

The findings were:

- It is hard to select the threshold settings in a noisy domestic environment
- While data acquisition, the Traditional VAD module is unable to differentiate between EXP and INSV (cry signal segments) and recorded speech signal segments.
- Traditional VAD modules are unable to distinguish expiration (EXP) from inspiration (INSV) parts of a cry audio signal

Statistical approach is a good solution to avoid restricting the problem of adjusting thresholds. That is why due consideration is given to statistical model-based approaches proposed by AbouAbbas in 2015b, 2015c modules.

2.4 Existing Systems:

There are systems to detect whether a sound file provided is a Baby cry or not. The techniques used LFCC (Linear Frequency Cepstral Coefficients) for feature extraction. There is also a system to classify the reasons for baby cry and in this system various classifiers are used to classify the reasons from the pre-classified data set.

3. METHODOLOGY

Based on the above literature survey, we finally decided to implement this project by using MFCC (Mel Frequency Cepstral Coefficients) for feature extraction and make efficient use of KNN machine learning model to classify the reasons for baby cry. Not only that this model is efficient, but it provides better results in case of sound based classification and audio files.

This approach has the following steps:

1. An audio file containing the unknown baby cry and whose reason is to be classified is uploaded.

2. The input audio file is preprocessed to remove empty audio frames and unwanted noise.

3. Then the audio is converted to cepstral coefficients to extract the features (here MFCC technique is used to obtain cepstral coefficients)

4. Once the cepstral coefficients are obtained, the mean of the coefficients is taken for further process.

5. Next, the KNN classifier is applied to classify the reason from the already trained model.

6. The reason for baby cry is obtained by considering the ‘n’ nearest neighbour which has the highest accuracy.

7. The reason for the baby cry is displayed as output.

3.1 Proposed System:

Our proposed system will be having the following main modules:

- Get the audio file to be classified.
- Preprocess the audio file.
- Obtain the Cepstral Coefficients.
- Apply to the KNN model for classifying the reason behind the cry.
- Output the final classified result.

3.2 System Architecture:

The flow of the system architecture goes like this,

- Various audio files containing baby cries are taken.
- Then the feature extraction is done for the audio files.
- After feature extraction, the output will be pitch and MFCC (Mel Frequency cepstral coefficients)
- In feature extraction, the noise is removed and unwanted empty segments of the audio file are removed.
- The outputs of feature extraction are used for training the model.
- For training the model, the MFCCs for each audio file is observed and the mean of the coefficients are taken for further classification.
- Then the classifier is trained with all the available data.
- Now, an unknown audio file is given for classification.
- The audio file is preprocessed and the features are extracted.