Emotion Detection from Infant Cries

Samiksha Dharmadhikari
Department of Computer Engineering
Pimpri Chinchwad College of Engineering, Pune University, India

Abstract— Infant cry is a way in which baby tries to communicate specific messages. As crying is the only way to express discomfort, so this signal carries a lot of information about infants physical and psychological condition. Therefore, the aim to extract and classify the state of an infant based on the patterns exhibited by the crying sound signal. More specifically we propose a methodology able to distinguish among the following five states: (a) hungry, (b) uncomfortable, (c) need to burp, (d) in pain, and (e) need to sleep. There are many feature extraction techniques evolved such as Linear Predictive Codes, Perceptual Linear Prediction, Mel Frequency Cepstral Coefficients (MFCC), RASTA Filtering etc evolved over past few years. This paper is to study MFCC technique to detect the emotion of crying infant. Among all MFCC is the most prevalent and dominant method used to extract spectral feature. The recognition accuracy is high. That means the performance rate of MFCC is high. So here we compare the performance of various feature classification techniques that are used with MFCC and other extraction techniques.

Key words: Infant Cry, Feature Extraction, Mel Frequency Cepstral Coefficients, Feature Classification

I. INTRODUCTION

Crying is the only way of communication for an infant to communicate specific messages. Crying is a positive intimation of new healthy life. Also crying is all a baby can do to express any discomfort. Therefore, the aim is to classify the state of an infant based on the patterns exhibited by the crying sound events. There are a variety of reasons behind a baby cry sound event associated with a specific situation that is unpleasant for the infant (pain, uncomfortable, stress, need to burp, need to sleep etc.) and each one of these situations needs to be treated in a different way in order secure the infant. Mel Frequency Cepstral Coefficients, a favourite extraction technique used to derive significant acoustic features from voice-based signals, such as infant cry. This intelligent machine recognition technology can help parents to understand their babies in an even better fashion and assist them in the practices of child rearing.

In past decade, different studies analysed the acoustic characteristics of the infant or baby cry, for various purposes and applications. These include, for example, recognition of potential neurological insults or the medical status of newborns, discrimination between normal and hearing-impaired babies, enhancing social robot's behaviour in childhood education settings, or identification of babies solely by the sound of their cry.

In this paper, we describe an analysis of the cry sound of infants and present an algorithm for cry detection, which is aimed at notifying parent's infants physical and psychological condition. The scope of this research is: feature extraction, in which MFC (Mel-frequency cepstral) coefficients parameters are extracted from the infant cries. This contributes in further classification divided into groups of hungry babies, tired/sleepy baby, like burping baby, infants experience pain in the stomach, and uncomfortable baby. It is used to identify the meaning of 0-3-month-old infant cries.

Widely used speech features for auditory modelling are cepstral coefficients obtained through Linear Predictive Coding (LPC). Another well-known speech extraction is based on Mel-frequency Cepstral Coefficients (MFCC). Methods based on Perceptual Prediction which is good under noisy conditions are PLP and RASTA-PLP (Relative Spectra Filtering of log domain coefficients). There are some other methods like RFCC, LSP etc. to extract features from speech. MFCC, PLP and LPC are the most widely used parameters in area of speech processing.

II. LITERATURE SURVEY

A. *Existing System*

1) Rahmat Hidayati, Ketut Eddy Purnama, Mauridhi Hery Purnomo:

Pitch and formants feature extraction technique is used. K-means classifier technique for feature classification is used. Database used id Voice Pen IC Recorder (SONY ICD-TP620). Performance Accuracy – 90%.

2) Advantages:

The results of sound processing module show high accuracy for the 'pain' type of cries.

3) Disadvantages:

Other types of cries resulting in more errors in the overall detection process.

B. *Bhagatpatil Varsharani V, V. M. Sardar:*

Feature extraction technique used is Linear Frequency Cepstrum Coefficients (LFCC). Feature classification technique K-means algorithm. Database used cry signals were recorded from the Neonatal Intensive Care Unit (NICU) of Department of Neonatology, Nobal Hospital, and Pune. Performance Accuracy -higher about 94%.

1) Advantages:

LFCC resulting higher formant frequencies in speech. LFCC is as robust as MFCC.

2) Disadvantages:

Not used widely due to its linear frequency.

C. *Swarnalexmi Nagarajan, Rajeswari Rengarajan, Nivethitha Manoharan, Kanniga Devi Baskaran:*

Feature Extraction technique used is Mel-frequency and Linear predictive coding method.Feature classification is done by GUI and Microsoft Excel for statistical analysis. Database used is infant cry signals are stored in test folder which are collected from the site (http://www.soundJay.com). Performance accuracy mentioned is LP technique can give us and most accurate results in infant cry analysis.

D. Rami Cohen, Yizhar Lavner:

Feature extraction technique used is MFCC (mel-frequency cepstrum coefficients) and short-time energy parameters. Feature classification is done by using k-NN algorithm. Database created by G. Varallyay. Performance detection rate of almost 100% in high SNR (signal to noise ratio)

1) Advantages:

Demonstrate both high detection rate and robustness in the presence of noise, even at low SNR values

2) Disadvantages:

Scope is limited to be classified either as 'cry' or 'no cry'

E. Medhanita Dewi Renanti, Agus Buono , Wisnu Ananta:

Feature extraction technique used is MFCC. Feature classification technique used is codebook of clusters using k-means clustering. Database Dunstan Baby Language. Performance accuracy recognition of infant cries with the higher about 94%.

1) Advantages:

Maximum accuracy in varying frame lengths.

2) Disadvantages:

The silent is only be cut at the beginning and at the end of speech signal.

III. SYSTEM DESIGN

A. Feature Extraction

MFCCs are the most popularly used for speaker/speech recognition. MFCC integrates a Mel frequency scale which is based on human ear scale. These were presented by Davis and Mermelstein in the 1980's. Earlier to the introduction of MFCCs, LPC, LPCC, RASTA filtering were the main feature extraction techniques in speech recognition.

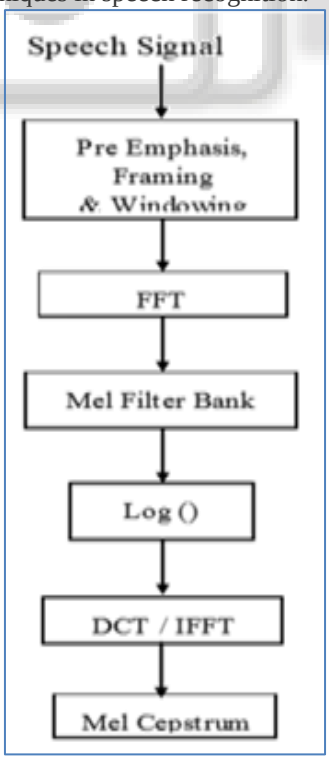

Fig. 1: MFCC Derivation

The Fig.1 is illustrates calculation of MFCC which is explained in following steps

1) Pre-Emphasis: Boosting of signal to high frequencies to provide more information.
2) Framing: Blocking the speech signal into frames of n samples.
3) Windowing: Minimize the signal discontinuities. Each frame now connects with the beginning of next smoothly.
4) Fourier Transform: actual frequency measured in hertz is converted to pitch measured by scale i.e. mel scale. Main purpose: mimic the behaviour of human ear
5) Map the powers onto mel scale: It converts each frame from time domain to frequency domain.
6) Log: Taking the log of powers at each of mel frequencies.
7) DCT: Taking Discrete cosine transform of list of mel frequencies Convert the log mel spectrum into the time domain

1) Advantages:

1) Accuracy is high, so performance is also high.
2) It approximates the human system response more closely than any other technique does.

2) Why so popular:

1) Efficient to compute.
2) Incorporates a Mel Frequency Scale.
3) Separates source from filter.

3) Disadvantages:

1) In background noise MFCC does not give accurate results.
2) Performance might be affected by number of filters.

4) Application:

1) MFCCs are used to automatically recognize numbers spoken into a telephone.
2) MFCCs are used in music information retrieval applications.

B. Feature Classification

Different neural network paradigms are available for classifying the patterns. KNN is the most widely used algorithm with MFCC as per the survey. Objects are divided into k groups, and then k is chosen as priori. Cluster is then found out by finding the centroid of each of k groups and assign each group the closest centroid. This approach minimizes the overall within-cluster dispersion by iterative reallocation of cluster members.

IV. CONCLUSIONS

In this work we surveyed many research papers which have proposed various techniques for analysing the reason behind infant cries. There are many feature extraction techniques which exists in this modern era. Each technique has its own merits and demerits. Feature extraction used with correct feature classification technique produces accurate and efficient results. MFCCs are one of the most popular feature extraction techniques used in speech recognition based on frequency domain using the Mel scale which is based on the human ear scale. In sound processing, the mel-frequency cepstrum is a representation of the short-term power spectrum of a sound. In MFCC, the frequency bands are equally spaced on the Mel scale, which approximates the human auditory system response more closely than the linearly- spaced frequency bands used in the normal cepstrum. Other feature extraction techniques like LPC, PLP-RASTA, PLP have less

accuracy when compared with MFCC. In MFCC the recognition accuracy is high. That means the performance rate of MFCC is high. MFCC is observed as most widely used techniques for the application of analysis of the cry sound of infants.

In a future research we plan to extend the evaluation of the proposed algorithm, using a broader set of cry and noise signals. The infant cry cannot only be treated as a means of communication, or a sign of mood as it is very important acoustic signal, so it can be used to diagnose several diseases and accordingly the neonates can be treated in an effective manner. Further studies will develop a device capable to analyse crying and giving a help in diagnostics.

REFERENCES

[1] Rahmat Hidayati, Ketut Eddy Purnama , Mauridhi Hery Purnom -"The extraction of acoustic features of infant cry for emotion detection based on pitch and formants", International Conference on Instrumentation, Communication, Information Technology, and Biomedical Engineering, 2009

[2] Bhagatpatil Varsharani V, V. M. Sardar –"An Automatic Infants Cry Detection Using Linear Frequency Cepstrum Coefficients (LFCC)", International Journal of Technology Enhancements and Emerging Engineering Research

[3] Swarnalexmi Nagarajan, Rajeswari Rengarajan, Nivethitha Manoharan, Kanniga Devi Baskaran –" Infant Cry Analysis for Emotion Detection by using Feature Extraction Methods", International Journal of Industrial Electronics and Electrical Engineering, May,2017

[4] Rami Cohen, Yizhar Lavner -"Infant Cry Analysis and Detection", 2012 IEEE 27-th Convention of Electrical and Electronics Engineers in Israel

[5] Medhanita Dewi Renanti, Agus Buono , Wisnu Ananta Kusuma –" Infant Cries Identification by using Codebook as Feature Matching, and MFCC as Feature Extraction", Journal of Theoretical and Applied Information Technology, 2013

[6] Namrata Dave – "Feature Extraction Methods LPC, PLP and MFCC in Speech Recognition", July 2013, International Journal for Advance Research in Engineering and Technology, 2013

