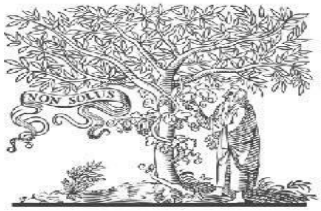
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 23rd Feb 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 03

10.48047/IJIEMR/V12/ISSUE 03/35

Title **ANALYSING BABY VOICE FOR FEELING AND HEALTH DETECTION USING MACHINE LEARNING**

Volume 12, ISSUE 03, Pages: 245-249

Paper Authors
Dr G. Rajesh Chandra, Myneni Sai Revanth, Mendem Poojith, Nuttaki Jaya Manikanta, Kaki Moses Daniel

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Analysing Baby Voice for Feeling and Health Detection Using Machine Learning

Dr G. Rajesh Chandra[1], Myneni Sai Revanth[2], Mendem Poojith[3], Nuttaki Jaya Manikanta[4], Kaki Moses Daniel [5]
Professor[1], Final Year BTech students[2,3,4,5], Department of Computer Science and Engineering, KKR & KSR INSTITUTE OF TECHNOLOGY AND SCIENCES (JNTUK) Guntur, India.
grajeshchandra@gmail.com[1],revanthmyneni@gmail.com[2], mpoojith015@gmail.com[3],manikantastar87@gmail.com[4],kmosesdaniel@gmail.com[5]

Abstract
Communication in different type in which crying is one of the forms. Infants can only express their feelings through crying only. Baby's cry can be characterized according to its natural periodic tone and the change of voice. Detection of a baby cry in speech signals is a crucial step in applications like remote baby monitoring. This study of sound recognition involves feature extraction and classification by determining the sound pattern Signal processing of crying signal and make the pattern classification using machine learning algorithms. Mel frequency cepstral coefficient (MFCC) method of feature extraction uses cry signal and classify the tone patterns of the audio signal. In this we are trying to increase the accuracy of the classification using K-NN algorithm. The KNN classifier consistently yields better results compared to other classifiers.

Keywords: Speech Recognition, Audio Processing, Baby Crying, Signal Patterns, Speech signal Processing, Feature Extraction, MFCC, K- Nearest Neighbor

Introduction

Crying is one of the major ways babies communicate with their surroundings, and is intended to point out any distress and attachment needs to their caregivers. Automatic detection of a baby's cry in audio signals can be used for various purposes – from everyday applications to academic research. Using this technique, we are able to take care of infant needs based on infant cry sounds. The baby cry may be of various reasons such as hunger, pain, discomfort, fatigue etc. These reasons can be identified using baby cry and gestures etc. In real time it becomes a complex problem to interpret the infant cry, but with the advancement of technology we can make this problem easier. We can also monitor the sleep phases, vital signals to detect systems symptoms of infections and the external environments factors such as temperature, humidity and air quality index that affects the behaviour of the child in this paper we are going to use machine learning algorithms that include low-level audio features selections from cry audio recordings. In this we use 50 to 100 audio data sets which help us to increase the accuracy of the machine learning model. 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.

Related Work

An infant's cry audio signal consists of inspiration and expiration parts, which are accompanied by vocalization and audible inspirations (INSV) and expirations (EXP). One of the fundamental challenges faced by such a system is implementing a method that can search for INSV and EXP precisely within a cry signal. The problem of cry detection is different from unvoiced and voiced segmentation because a typical hear-able infant cry audio signal contains each of the unvoiced and voiced parts.

The main issue with cry detection in a lot of domestic environments with lots of domestic household noise is not easy to resolve using VAD (Voice Activity Detection) modules since VAD deals with the search or finding of speech patterns from different parts of the considered audio signal. Any other auditory active

pattern may be of any type, such as 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. The most popular and widely used VAD methods involve two basic and significant methods: Feature Extraction and Decision Making. Features of a signal that allow computation of energy, cepstral coefficients decision rule computation based on frame-by-frame and very simple rules for thresholding.

Background

It is challenging to select 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.

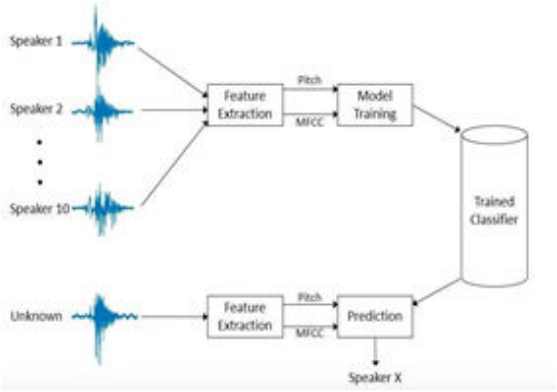
This is because traditional VAD modules are based on spectral-based features, which are unable to capture the temporal characteristics of the signal needed to distinguish between inhalation and exhalation.

In order to avoid limiting the problem of threshold adjustment, a statistical approach is a feasible solution. That is why due consideration is given to statistical model-based approaches.

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

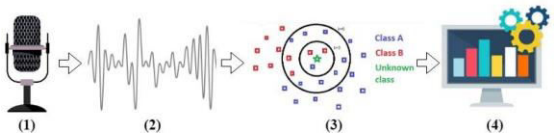
Methodology

We used MFCC (Mel Frequency Cepstral Coefficients) for feature extraction in this project and used a KNN machine learning model to classify the reasons for baby crying based on the literature survey above. Not only is this model effective, but it provides better results when it comes to sound-based classification and audio files.



System Architecture

We have to upload a file that contains an unknown baby cry and its reason for classification. The input audio file is pre-processed to remove empty audio frames and unwanted noise. Then the audio is converted to cepstral coefficients to extract the features (here MFCC technique is used to obtain cepstral coefficients). Once the cepstral coefficients are obtained, the mean of the coefficients is used for further analysis. Next, the KNN classifier is applied to classify the reason from the already trained model. By considering the 'n' nearest neighbour, which has the highest degree of accuracy, one can determine the reason for a baby's cry. As an output, the reason for the baby's cry is displayed.



System overview

The flow of the system proceeds like this:

Various audio files containing baby cries are taken. Then feature extraction is done for the sound files. After feature extraction, the output will be pitch and MFCC (Mel Frequency Cepstral coefficients). In feature extraction, noise is removed and unwanted empty segments of the audio file are removed. The outputs of feature extraction are used for fitting the model. For training the model, the

MFCCs for each audio file are observed and the mean of the coefficients is 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 pre-processed and features are extracted. By combining the extracted features with the predictive model, the reason can be classified/predicted from the information already gathered. The reason for a baby's cry is classified, then the cry is classified as an output.

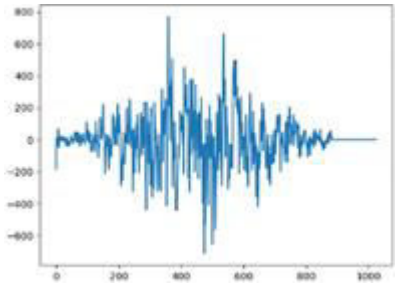
Silence Removal

The speech signals contain many areas of silence or noise. The silence parts of the signal are useless for recognition, because they provide no information. And if we keep the silence part of the signal, it will make the processing signal too large. This will take more time and memory when extracting information or features from the speech signal. Therefore, only the signal parts that contain actual speech are useful for recognition.

Normalization

Normalization is a technique to adjust the volume of audio files to a standard level. This is because different recording levels can cause the volume to differ greatly from word to word. Recording sound samples with different volumes and possibly some DC offset should have no impact on the detection system. Getting rid of this is as simple as normalizing the signal, for example by scaling, and then offsetting the signal by decibels so that it can be put between levels -1 and 1.

Windowing



Windowing

Speech signals are generally unstable, meaning their statistical properties do not remain the same over time. But in a short interval of time, generally 10-30 ms, speech signals can be regarded as stationary. This is carried out by multiplying the speech samples with a windowing function to cut out a short segment of the speech signal. A window is the period of time during which the signal is considered or processed. The data belonging to the window is called a frame.

Mel Frequency Cepstral Coefficients (MFCCs)

MFCC is one of the most popular feature extraction techniques used in automatic speech or speaker recognition systems. It is based on the Mel scale, which is based on the human ear scale. It is based on the nonlinear human perception of the frequency of sounds. These coefficients represent audio dependent on perception. They are derived from the Mel frequency curve. The spectral information can then be converted to MFCC by passing the signals through band pass filters. In these filters, higher frequencies are artificially boosted, and the result is the inverse Fast Fourier Transform. It combines the advantages of the cepstrum analysis with a perceptual frequency scale based on critical bands. MFCCs are a compact representation of the spectrum (When a waveform is represented by a summation of a possibly infinite number of sinusoids) of an audio signal.

The spectral information can then be converted to MFCC by passing the signals through band pass filters where higher frequencies are artificially boosted, and then applying an inverse Fast Fourier Transform (FFT) to it. It combines the advantages of particle analysis with a perceptual frequency scale based on critical bands. As a result, higher frequencies are becoming more prominent. Since the Mel frequency spectrum can represent a listener's response system clearly, therefore MFCC is always considered to be the best available approximation of the human ear.

K-Nearest Neighbor Algorithm

K-nearest-neighbour is a simple nonparametric classification method, which means it does not make any assumptions about the underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset. This is because at the time of

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

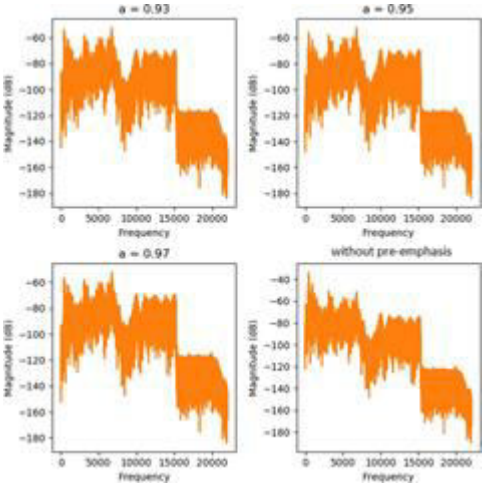
classification, it performs an action on the dataset. The KNN algorithm in the training phase just stores the dataset. When it gets newly acquired information, it classifies that data into a category that is much similar to the new data. It classifies a sample according to its kth nearest neighbours in the feature space, which correspond to the majority class. The time complexity of training a K-nearest neighbour's model is where n, d, and k are the number of instances, data dimension, and the number of neighbour's, respectively.

K-Nearest Neighbours is one of the most basic yet essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining and intrusion detection. It is widely applicable in real-life scenarios since it is non- parametric, meaning, it does not make any underlying assumptions about the distribution of data (as opposed to other algorithms such as GMM, which assume a Gaussian distribution of the given data).

KNN process

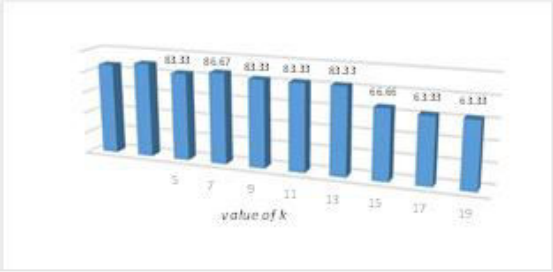
Results

The previous study, the use of MFCC as feature extraction and Euclidean distances for classification, gets high accuracy. The model can produce a truth value for infant cry classification higher than 93% which in that study removes silent tones at the beginning and end of the speech signal. In this system, several tests conducted on a child's crying detection system designed using the MFCC method and KNN classification. MFCC method has been tested by comparing the pre-emphasis values, the filter bank values, the cepstral values, and K values in the KNN classification.



Signal magnitude using different pre-emphasis value

The different pre-emphasis values result different magnitude signals after the emphasized process. The higher the coefficient value of the pre-emphasis the sharper the sound at high frequency. The difference of magnitude signal between using pre-emphasis process and without pre-emphasis process. This different magnitude does not give impact to sound classification. They get the same accuracy although there are different pre-emphasis values in MFCC



Graph of testing the K value

The application of k values more than 14 gets lower accuracy to 66.66% and getting smaller so that it can be concluded that the greater k values can lead to the smaller classification accuracy. It shows in figure 11. The

training data quality can affect the detection accuracy. In this system, training data and testing data are relative clear from any kind of noise. The use of pre-emphasis on the mFCC method does not give significant impact to the classification process but we recommend to use pre-emphasis to avoid sound data with low quality. The selection of the filter bank value applied must be greater than the cepstral value applied and the cepstral values are adjusted to get the best performance in baby's cry detection. The highest accuracy is 90% using the cepstral value of 8 with the nearest neighbor value of 3, where all parameters are set at the best condition based on the test results.

Conclusion

The use of MFCC as feature extraction method and K-Nearest Neighbour (K-NN) as classification method can detect the baby is crying or not. Therefore, it can be used as a way for parents to monitor their children remotely only when certain conditions are met by their children. Tests on the pre-emphasis value, filter bank value, cepstral value, and K value on K-NN have different calculation scenarios. The use of coefficients in the pre-emphasis does not give significant impact to improve the accuracy in the classification process but it affects the quality of the feature extraction results. The choice of filter bank value and cepstral value can affect the accuracy of the classification process even though it is not significant. The use of K value will affect accuracy during the classification process as well as the quality of the testing data. According to our study, the most accurate results in the testing scenario are: the filter bank number must be greater than the cepstral value, and cepstral values are adjusted in order to achieve the most accurate performance in detecting baby's cry. The highest accuracy is 90% using the cepstral value of 8 with the nearest neighbour value of 3. In this case, all parameters are set at the optimal condition based on the test results. We can conclude that the use of the MFCC method can be implemented in the baby crying detection system. If training data and test data are free of noise, it will produce high characteristic values. Analysis of the crying sound of a baby is not sufficient. There must be a normalization process and a pre-processing process to reduce noise prior to the core process.

Future Enhancements

The classification of baby's cry can be conducted by using various data from different ethnic baby, wider range of age, various sounds similar to baby sounds, etc. Further testing must be carried out in a noisy and not noisy place. Many other feature extraction method and classification method can be used to compare their performance.

References

Osmani A., Hamidi M., Chibani A. Machine learning approach for infant cry interpretation. Proceedings of the IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI); 6 November 2017; Boston, MA, USA. IEEE; pp. 182–186.

[2] Barajas-Montiel S. E., Reyes-Garcia C. A. Identifying pain and hunger in infant cry with classifiers ensembles. Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCAIAWTIC"06); 28 November 2005; Vienna, Austria. IEEE; pp. 770–775.

[3] Orlandi S., Reyes Garcia C. A., Bandini A., Donzelli G., Manfredi C. Application of pattern recognition techniques to the classification of full term and preterm infant cry. Journal of Voice. 2016;30(6):656–663. doi:10.1016/j.jvoice.2015.08.007.

[4] Ashwini K., Pm D. R. V., Srinivasan K., Chang C. Y. Deep convolutional neural network-based feature extraction with optimized machine learning classifier in infant cry classification. Proceedings of the International Conference on Decision Aid Sciences and Application (DASA); 8 November 2020; Sakheer, Bahrain. IEEE; pp. 27–32.

[5] Dezecache G., Zuberbühler K., Davila-Ross M., Dahl C. D. A machine learning approach to infant distress calls and maternal behaviour of wild chimpanzees. Animal Cognition . 2020;24:1–13