International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 9 Issue 3, May-Jun 2025 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Automated Speech Recognition System to Detect Babies' Feelings through Feature Analysis using SVM Algorithm

Mr. K N V Krishna Sai Ram[1], Mr. Parasurama N[2], Prashanthi Maddala[3]

[1]PG Studies, Department of Computer Science and Engineering,
[2]Associative Professor, Department of CSE,
[1,2]JNTUK University College, PPDCET Vijayawad, Andhra Pradesh, India
[3]Assistant Professor, Department of CSE, Vikas Group of Institutions, Vijayawada Rural, Andra Pradesh, India

ABSTRACT

This "paper investigated the signal processing and feature extraction of the aforementioned speeches in order to investigate the intelligent machine technology for comprehending infants' needs and emotions from speech signals and, as a result, assisting parents in child rearing. It would appear that the infant's speech signals alone were not sufficient to achieve a high level of precision and dependability when it came to dealing with a variety of needs and emotions. Taking into account the combined characteristics of acoustic characteristics and rearing behaviors, an efficient recognition strategy was therefore proposed using the SVM classification algorithm. The findings of the experiment demonstrated that the majority of infants' typical physiological and psychological states, such as happiness, hunger, and slumber, can be correctly identified with a relatively high level of accuracy.

KEYWORDS: *SVM classification algorithm, digital processing, Matlab, MFCS, Speech signal*

How to cite this paper: Mr. K N V Krishna Sai Ram | Mr. Parasurama N | Prashanthi Maddala "Automated Speech Recognition System to Detect Babies' Feelings through Feature Analysis using SVM Algorithm" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-9 | Issue-3, June 2025, pp.852-857, URL: www.ijtsrd.com/papers/ijtsrd81124.pdf

IJTSRD81124

I. INTRODUCTION

The emotions and needs of an infant can be seen in a variety of information sources, including their attitudes, movements, activities, locations, and electrophysiological signals (such as heart rate, blood pressure, electrocardiogram, electroencephalogram, and so on)." of "these, the talk sign is accurately and completely gotten, and it also possesses numerous common features of both phonetic information and paralinguistic information.

In our paper, we propose a recognition approach considering MFCC parameters combined with infant's age and rearing behaviors. The features associated with infant's common emotions and MFCC parameters. Before the calculation of those parameters, a standard data pre-processing which usually includes signal sampling and quantizing, pre-emphasis, framing and windowing, and endpoint detection should be employed to regulate the recorded original signals. The Feature extraction not only acquires MFCC parameters, but also includes the infant's age, last feeding time, and slept time. Before the pattern recognition, a template is necessary by the training of annotated speeches which have been classified into different kinds of emotions and needs. Using SVM classification, we can classify Happy, Hungry and sleepy emotions.

The examination of talk signals and the management of sign procedures are part of talk preparation. Talk preparation is a remarkable example of mechanized sign planning associated with talk signals because the signs are typically depicted in a propelled manner." The verification, "control, accumulation, movement, and output of talk signals are all functions of parts of talk taking care of fuses. The results are referred to as talk association and the data as talk affirmation.

Crying is a child's first and most important communication skill. The cry of a newborn child is a fixation that urges parents or guardians to move forward with the plight. Dunstan Baby Language (DBL) is a current framework for learning about the size of a newborn child's cry from 0 to more than a month old. Priscilla Dunstan, an Australian skilled worker who has acquired the ability to review a wide range of sounds, also known as sound photo, is the pioneer of DBL.

The DBL illustration demonstrates that there are four infant language expressions: "neh" denotes hunger, "owh" denotes sluggishness, "eairh" denotes stomach pain, and "heh" denotes ungainly. The sporadic nature of newborn child tyke's crying, such as squinting cry enunciations and consolation, is illustrative. Because "there is visit opening and closing of the vocal folds, which conveys cyclic excitation, sound is transmitted by using a skillful surge of air through the larynx break. In order to produce the cry sound, which typically has a key/pitch rehash of 250-600 Hz, this excitation travels through the vocal tract.

A child's cry's sound-related flag contains irrelevant information about their physical and physiological state, such as their constitution, weight, personality, sex, and estimations. Zero-intermingling rate and central rehash [1], Fast Fourier changes coefficients, were identified and isolated here for cry disclosure in order to perceive crying signs." The "examination is based on establishing a legal separation between cries that could be used for basic reasoning purposes, and one can barely care about racket-free conditions in clinical settings. The finding issue does not criticize racket-free conditions in clinical settings. In the end, quality of seeing crying signs in amusing and ear-splitting situations is always the main focus.

II. RELTED WORK

I am working on 15 investigate papers. Get data The Testing signal from my speech dataset only because of training parameters.

A literature survey provides a broad overview of existing research on a topic, focusing on gathering and summarizing information rather than in-depth analysis or critical evaluation.

In this research, we aim to help other researchers by making a systematic literature review of automatic speech recognition that will provide them with the most significant topics published in the last six years. Also, this research will help to specify the recent major challenges and the research gaps in automatic speech recognition. Moreover, it will provide them with future research directions in this area.

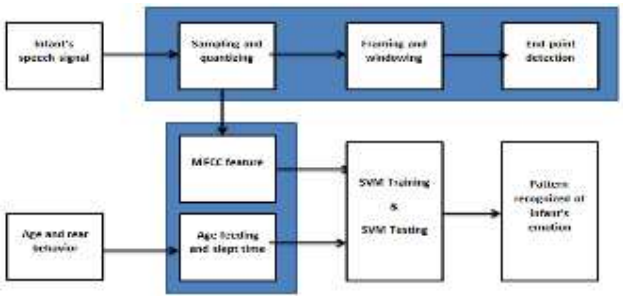
The specific search in each database was by using titles with keywords as in the following: "artificial intelligence " AND ("speech recognition" OR "automatic speech recognition ")

III. Existed system

The recognition of infant's speech signals has been paid close attention for many years . According to the physiological findings, the generating mechanism of infant's laugh and cries has a lot of similarities to the speeches of adults. It has been justifiable that the signal processing and pattern recognition techniques, such as Cepstrum Analysis, Dynamic Time Warping (DTW) and Hidden Markov Modeling (HMM), can also be applied to the analysis of infant's speech data. Existed So the acoustic parameters of infant's speech signals, such as duration, pitch, and energy etc., have been considered as the main features in existing recognition systems, among which, duration corresponds to the lasting time that infant begins uttering to the end. It has low accuracy.

IV. Proposed system

In our paper, we propose a recognition approach considering MFCC parameters combined with infant's age and rearing behaviors. The features associated with infant's common emotions and MFCC parameters. Before the calculation of those parameters, a standard data pre-processing which usually includes signal sampling and quantizing, pre-emphasis, framing and windowing, and endpoint detection should be employed to regulate the recorded original signals. The Feature extraction not only acquires MFCC parameters, but also includes the infant's age, last feeding time, and slept time. Before the pattern recognition, a template is necessary by the training of annotated speeches which have been classified into different kinds of emotions and needs. Using SVM classification, we can classify Happy, Hungry and sleepy emotions.



Proposed block diagram.

Parameters of MFCC. Preceding the estimation of those parameters, a standard data pre-taking care of which more frequently than prohibits sign investigating and quantizing, pre-emphasis, enclosing and windowing, and endpoint area should be used to control the recorded novel sign. The Feature

extraction gains MFCC parameters, yet what's more joins the child's age, last continuing time, and snoozed time. A catch on the organized electronic device makes it possible to successfully record those data.

V. Support Vector Machine (SVM) Algorithm

SVMs aim to find the best possible line, or decision boundary, that separates the data points of different data classes. This boundary is called a hyperplane when working in high-dimensional feature spaces

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. It tries to find the best boundary known as hyperplane that separates different classes in the data. It is useful when you want to do binary classification like spam vs. not spam or cat vs. dog.

The main goal of SVM is to maximize the margin between the two classes. The larger the margin the better the model performs on new and unseen data**.**

Support Vector Machines (SVMs) are a versatile machine learning algorithm with applications in various fields. They are commonly used for classification and regression tasks, particularly in areas like image and text classification, bioinformatics, and time series prediction

Key Concepts of Support Vector Machine

- **Hyperplane**: A decision boundary separating different classes in feature space and is represented by the equation wx + b = 0 in linear classification.
- **Support Vectors**: The closest data points to the hyperplane, crucial for determining the hyperplane and margin in SVM.
- **Margin**: The distance between the hyperplane and the support vectors. SVM aims to maximize this margin for better classification performance.
- **Kernel**: A function that maps data to a higher-dimensional space enabling SVM to handle non-linearly separable data.
- **Hard Margin**: A maximum-margin hyperplane that perfectly separates the data without misclassifications.
- **Soft Margin**: Allows some misclassifications by introducing slack variables, balancing margin maximization and misclassification penalties when data is not perfectly separable.
- **C**: A regularization term balancing margin maximization and misclassification penalties. A higher C value forces stricter penalty for misclassifications.
- **Hinge Loss**: A loss function penalizing misclassified points or margin violations and is combined with regularization in SVM.
- **Dual Problem**: Involves solving for Lagrange multipliers associated with support vectors, facilitating the kernel trick and efficient computation.

VI. Digital Image Processing Algorithms using MATLAB

MATLAB stands for "Matrix Laboratory". It's a programming language and software environment developed by MathWorks primarily used for numerical computation, data analysis, and technical computing

An image could also be defined as a two-dimensional function, f(x, y), where x and y are spatial (plane) coordinates and therefore the amplitude of f at any pair of coordinates (x, y) is named the intensity or grey level of the image at that time. When x, y, and therefore the amplitude values of f are all finite, discrete quantities, we call the image a digital image**.**

In MATLAB, the IPT is a collection of functions that extends the capability of the MATLAB numeric computing environment. It provides a comprehensive set of reference-standard algorithms and workflow applications for image processing, analysis, visualisation and algorithm development.

Carrying your research on **Speech processing projects using MATLAB** is one of the best ways to implement your innovative ideas. **Speech processing** refers to the techniques used in *recognizing and processing audio signals*. Speech processing technologies have developed exponentially in recent times.

- Speech technologies nowadays have improved quite significantly for analyzing complex data input.
- This has created tremendous use of artificial intelligence and machine learning algorithms into speech processing techniques.

The following article is a brief analysis of speech processing technologies and their various aspects. The analysis includes the following.

- Application of cognitive sciences in speech recognition
- Analysis and processing of speech signals
- Machine learning algorithms and tools
- Production and perception of speech
- Understanding multiple languages

VII. PATTERN RECOGNITION

Data Talk signal data are assembled from the infants inside 1year's old, who are generally under the conventional condition. We consider three common sentiments and needs in their consistently life: perky, hungry, and drowsy states. Those statuses are generally easy to be surveyed and checked by discernment or experience. The sound sign were tried at 16 kHz using a 16-bit easy to-electronic converter. Each talk signal ordinarily contained a movement of disengaged squares. We detached the squares into brief time ranges with 256 reviewing data for each packaging. The waveforms of three sorts of talk sign and their packaging data are displayed as from Figure1 to Figure3:

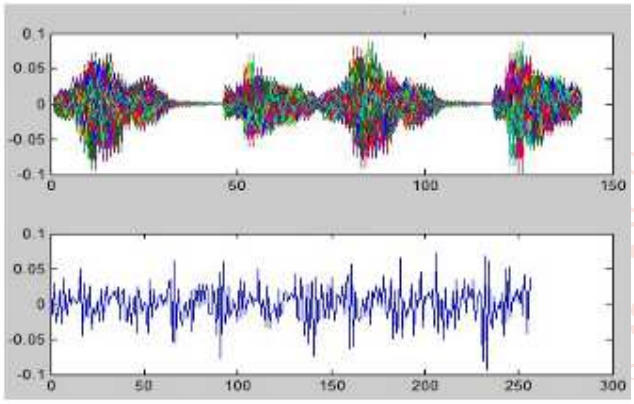


Fig.1. Waveforms of happy signal (upper) and one of its frame data (below)

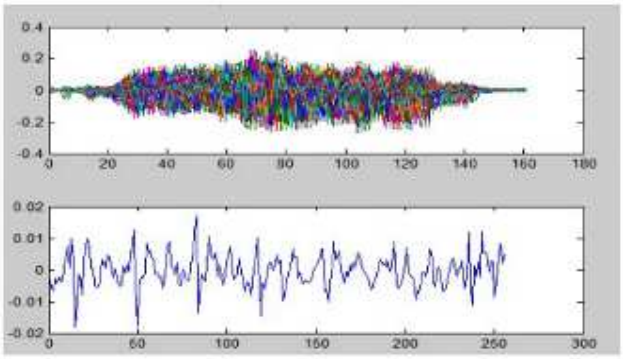


Fig.2. Waveforms of hungry signal (upper) and one of its frame data (below)

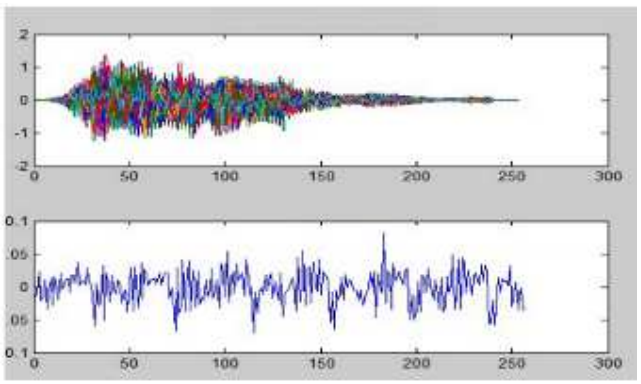
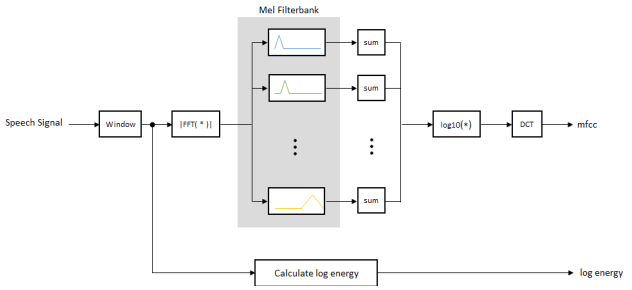


Fig.3. Waveforms of sleepy signal (upper) and one of its frame data (below).

VIII. SPEAKER FINDING USING MFCC

This model demonstrates an AI strategy for managing and recognizing individuals based on features removed from recorded speech. The features used to set up the classifier are: pitch of the voiced segments of the talk, and the Mel-Frequency Cepstrum Coefficients (MFCC). This is a closed set speaker recognizing verification - the sound of the speaker under test is contemplated against all the open speaker models (a restricted set) and the closest match is returned."



MFCC"

Mel-Frequency Cepstrum Coefficients (MFCC) are common features isolated from talk signals for use in affirmation tasks. In the source-channel model of talk, MFCCs are fathomed to address the channel (vocal tract). The repeat response of the vocal tract is commonly smooth, while the wellspring of voiced talk can be shown as an inspiration train. The result is that the vocal tract can be assessed by the powerful envelope of a talk section."

IX. RESULTS"

In this examination, we immediately tried the displays of MFCC parameters in its various solicitations and selected the parameters with the best performance to be arranged in our strategy. We recorded 300 genuine talk signals from infant children between the ages of two and one year to determine the best MFCC parameters.

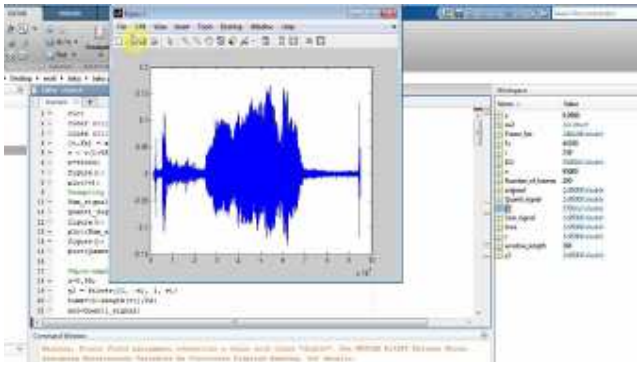


Fig 9.1 .First step execution

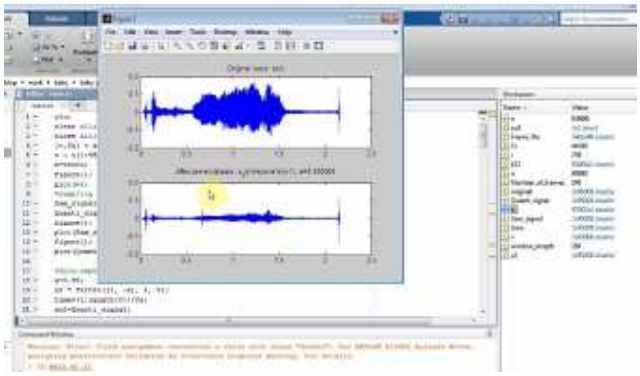


Fig 9.2.Pre emphasis

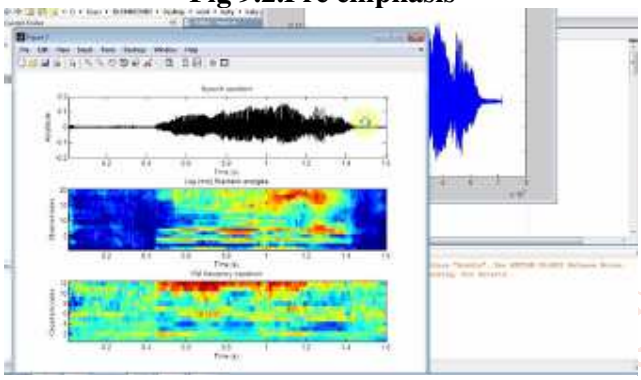


Fig9.3 Anlyzing waveform

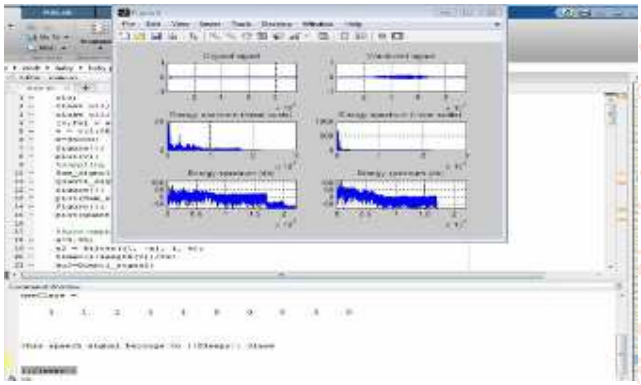


Fig9.4 Baby speech sleepy

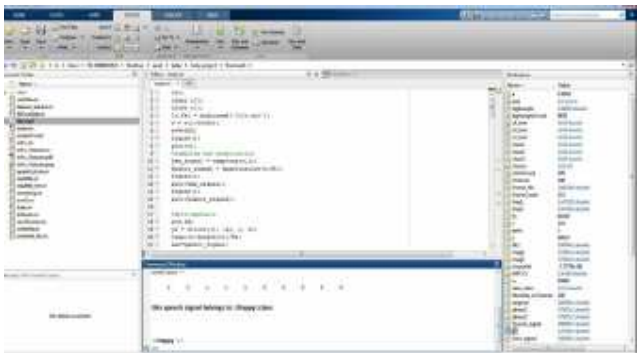


Fig 9.5 Baby speech happy

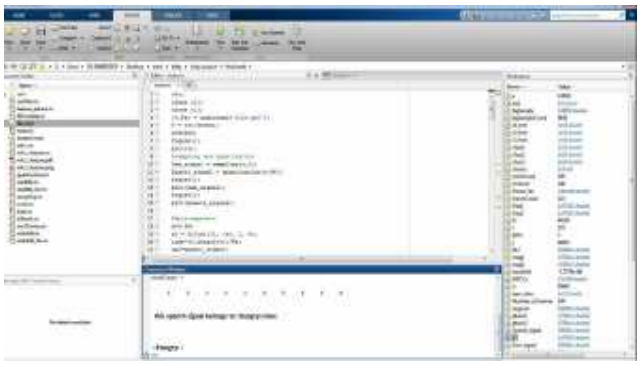


Fig9.6 Baby speech hungry

"We put the shaped structure into the use of 13 infant children for tenacious every month and a half in order to check the presentation and congruence of the above approach throughout the infant's life. This structure was worked by those infants' people, and arranged by their assessments of the infant kids' sentiments and needs through oneself altering count. The shifts in typical affirmation exactness as the weeks pass are depicted in Figure.

X. CONCLUSION

In this Paper, we examined the sign dealing with, incorporate extraction, and model affirmation of infant's talks. It suggests that the high precision and strong affirmation results of various infant needs and feelings could not be obtained from the talk banner, in a sense. Thusly, we explored the joined features of acoustic traits with the additional information, for instance, infant tyke's age and raising practices, and proposed a convincing technique subject to self-adapting figuring. Application test exhibited that the three ordinary emotions" and prerequisites (happy, hungry, and tired states) "in infant kids' each day life can be seen viably at a for the most part high exactness of over 80% by this technique.

REFERENCES

[1] G. Várallyay Jr., Z. Benyó, A. Illényi, Z. Farkas, L. Kovács, "Acoustic analysis of the infant cry: classical and new methods", in Proceedings of the 26th Annual International Conference of the IEEE EMBS, 2004, pp. 313-316

[2] J. O. Garcia, C. A. Reyes García, "Acoustic features analysis for recognition of normal and hypoacoustic infant cry based on neural networks", Lecture Notes in Computer Science: Artificial Neural Nets Problem Solving Methods, vol. 2687, pp. 615-622, 2003.

[3] N. R. Carlson, Foundations of Physiological Psychology, USA: Pearson Education Inc., 2005

[4] Wasz-Hockert, et al., The Infant Cry: A Spectrographic and Auditory Analysis, London: William Heinemann Medical Books Ltd., 1968

[5] F. L. Yuan, "The infant cries for acoustic feature analysis and recognition", Xihua University, Chengdu, 2001.

[6] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, "Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques", Journal of computing, vol. 2, no. 3, pp.138-143, 2010.

[7] S. C. Sajjan, C. Vijaya, "Comparison of DTW and HMM for isolated word recognition", in Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering of the IEEE, 2012, pp.466-470.

[8] M. Haro, J. Serrà, Á. Corral, P. Herrera, "Power-law distribution in encoded MFCC frames of speech, music, and environmental sound signals", in Proceedings of the 21st International Conference Companion on World Wide Web of the ACM, 2012, pp. 895-902.

[9] B. Vlasenko, A. Wendemuth, "Heading toward to the natural way of human-machine interaction: the NIMITEK project", in Proceedings of IEEE International Conference on Multimedia and Expo, 2009, pp.950– 953.

[10] W. Guo, R. H. Wang and L. R. Dai, "Mel-Cepstrum integrated with pitch and information of voiced/unvoiced", Journal of Data Acquisition and Processing, vol. 22, no.2, pp. 229–233, 2007.

[11] Y. S. Lei, "Analysis and pattern classification of infants' cry", Shandong University of Science and Technology, Qingdao, 2006.

[12] W. H. Dai, D. M. Han, Y. H. Dai, and D. R. Xu, "Emotion recognition and affective computing on vocal social media", Information & Management, vol. 52, no. 7, pp.777- 788, 2015.

[13] A. Zabidi, W. Mansor, L. Y. Khuan, "Discrete mutative particle swarm optimisation of MFCC computation for classifying hypothyroidal infant cry", in Proceedings of International Conference on Computer Applications and Industrial Electronics (ICCAIE) of the IEEE, 2010, pp.588 – 592.

[14] M. Sheikhan, D. Gharavian, F. Ashoftedel, "Using DTW neural-based MFCC warping to improve emotional speech recognition". Neural Computing & Applications, 2011, Doi:10.1007/s00521-011-0620-8.

[15] Y. J. Yuan, P. H. Zhao, Q. Zhou, "Research of speaker recognition based on combination of LPCC and MFCC", in Proceedings of IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS), pp.765-767, 2010.