iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.58973

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Human Scream Detection and Analysis for Controlling Crime Rate using Machine Learning and Deep Learning

Achyut Ojha[1], Adarsh[2]
Bachelor Of Technology Computer Science & Engineering (Artificial Intelligence and Machine Learning) Inderprastha Engineering College, Ghaziabad

Abstract: ***This paper addresses the problem of automatic detection and recognition of impulsive sounds, such as glass breaks, human screams, gunshots, explosions or door slams. A complete detection and recognition system is described and evaluated on a sound database containing more than 800 signals distributed among six different classes. Emphasis is set on robust techniques, allowing the use of this system in a noisy environment. The detection algorithm, based on a median filter, features a highly ro- bust performance even under important background noise conditions. In the recognition stage, two statistical classifiers are compared, using Gaussian Mixture Models (GMM) and Hidden Markov Models (HMM), respec- tively. It can be shown that a rather good recognition rate (98% at 70dB and above 80% for 0dB signal-to-noise ratios) can be reached, even under severe gaussian white noise degradations.***

Keywords: Impulsive sound detection – Sound recognition – Receiver Operating Characteristic (ROC-AUC) curve – Hidden Markov Models – Multimodels – Convolutional Neural Networks (CNN) – Telesurveillance – Mel-Frequency Cepstral Coefficients (MFCCs).

I. INTRODUCTION

The Human Scream Detection System is a major innovation designed to recognize and respond to human screams in real-time audio data. The main objective of the system is to respond to the immediate needs of people in distress, providing rapid assistance and protection in critical situations Human cries act as a powerful auditory signal of danger, fear or anxiety, crossing language barriers and cultural gaps. The ability to accurately detect and classify these danger signals is not only important for improving public safety, but also applies to a variety of scenarios, including security monitoring, smart home automation, healthcare and emergency response systems.

The Human Scream Detection System represents a fusion of state-of-the-art technologies that combine audio signal processing and machine learning principles to achieve its critical mission.

II. SYSTEM OVERVIEW

A human scream detection system starts by collecting audio recordings from various sources. These recordings undergo pre-processing to improve their quality and remove noise. Riff features such as pitch and intensity are extracted from the pre-processed sound using feature extraction techniques. Next, the classification model is trained to distinguish screams from non-screams using the extracted features. The system and its performance are evaluated through testing, which ensures its accuracy and reliability. After successful validation, the system will be integrated into applications such as security systems for real world use. Continuous improvement is emphasized through constant updates and improvements to improve the accuracy and efficiency of the system over time..

III. SOUND DATABASE

The database used in the experiments described in this article contains 872 sounds from 16 different categories related to situations of intrusion or aggression: 204 door slams, 78 glass breaks, 233 human screams, 62 explosions, 126 gunshots, and 40 other static sounds. The sounds are taken from different sound libraries (freesoundeffects.com, Noise Ex) and there are many variations within the same category (quality differences, signal length differences, signal energy levels, etc.). Some signals were also recorded manually. All signals were digitized and sampled.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

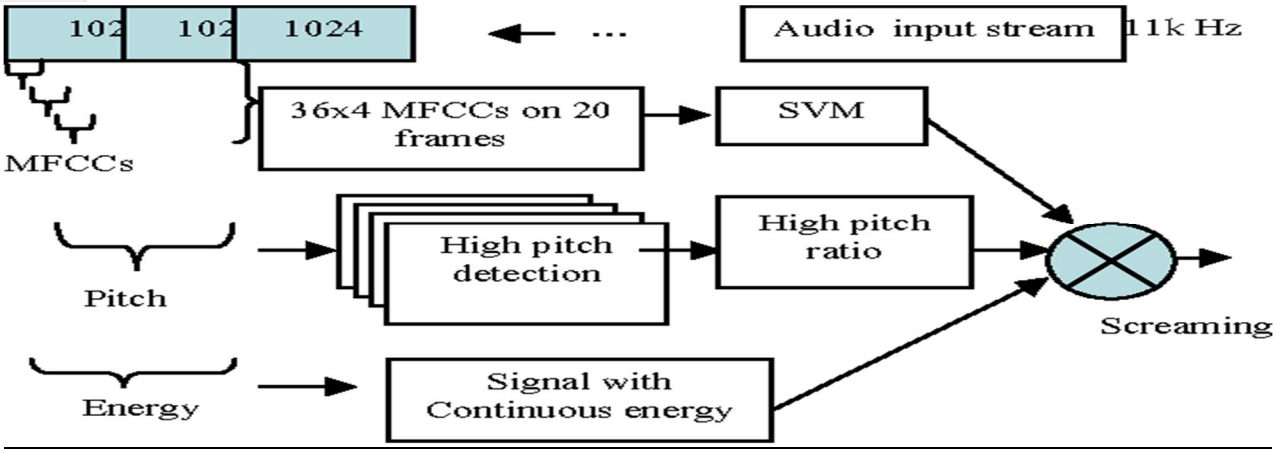


Figure 1: Overview of the Human Scream detection system

Impulses present in the temporal energy sequence More specifically; the expression process proceeds as follows. In the first step, the signal energy is estimated for each 100 ms consecutive block. Then, the obtained energy series is filtered by median filtering and the filter power is subtracted from the energy. This results in a new sequence that is normalized to highlight the significant energy pulses. An adaptive threshold is then applied - depending on the standard deviation of the previous energy sequence with a long window.

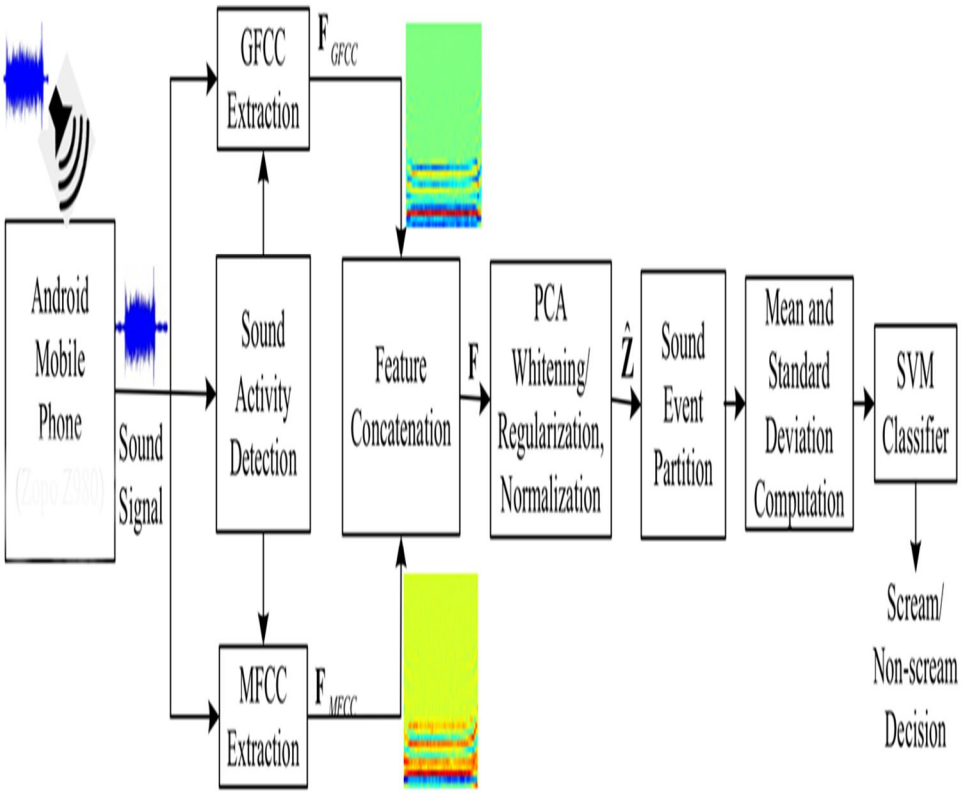


Figure 2: Detection of an scream, random white noise background environment

This method provides a configurable and highly sensitive impulsive signal detection system where impulses can be detected under rather unfavorable background noise conditions with a signal-to-noise ratio (SNR).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Figure 3 shows the achieved performance evaluated by the impulse audio database signals at different levels of Gaussian white noise. 100% correct detection rate above 0 dB SNR and very low false detection risk (0% above 5 dB, 0.7% maximum below) are guaranteed.

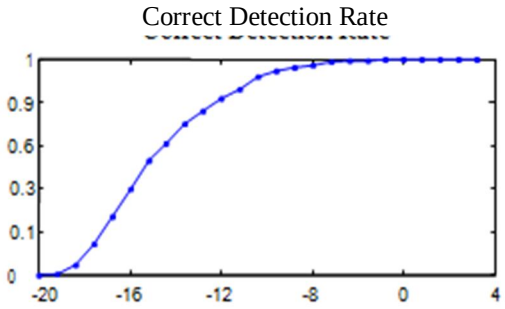


Figure 3: Sound detection Performance evaluation on the sound database.

IV. SCREAM DETECTION SYSTEM

A. Features Extraction

The first step of the recognition algorithm consists in an analysis of the signal to be classified, with a view to extractingsome typical features.

In this work, the spectrum of the signal is calculated for every successive time frame of 512samples. For each frame, the energy of N spectral bands isthen derived, covering the frequency range from 0 to 20 kHz in a uniform manner. In this way, every feature vector is composed of N parameters, representing the spectral energy distribution of one-time frame.

1) Gaussian Mixtures Models (GMM)

The statistical behavior (probability density functions, pdf) of each sound class can be modeled by a combination of Gaussians. This model is characterized by the number of Gaussians, their relative weights, and mean/covariance parameters.

During the training process, the system learns the GMM parameters by analyzing a subset of the audio database. To find the best model for each sound class, the likelihood is maximized using 20 iterations of the EM (expectation maximization) algorithm. In the detection process, the signal to be classified is compared with the models of each class to find the most likely signal**.**

2) Hidden Markov Models (HMM)

On the other hand, the use of left-right HMMs in the pattern recognition step offers the advantage of taking into account the temporal evolution of the signal features. This paper considers M = 3 consecutive states for signal characteristics, which roughly correspond to the attack, steady, and decay phases of a pulse.

During the pattern recognition process, the most probable class of signal is determined by a long-likelihood estimation. Instead of the Forward-Backward Algorithm, the likelihood is evaluated using the Viterbi approximation .

B. Implementation of Model

To implement a model for human scream detection, the process typically involves several key steps. Initially, data collection is paramount, requiring the assembly of a diverse dataset containing audio recordings featuring both human screams and non-scream sounds.

This dataset serves as the foundation for subsequent stages. Preprocessing the audio data follows, involving techniques such as spectrogram conversion, Fourier transforms, or MFCC extraction to derive pertinent features for the model. Once the data is preprocessed, model selection becomes crucial. Options range from traditional machine learning algorithms like SVM or k-NN to deep learning architectures such as CNNs, RNNs, or Transformers. With the model chosen, training ensues, where the dataset is split into training and validation sets to enable the model to learn to differentiate between screams and non-scream sounds by adjusting its parameters to minimize a specified loss function.

Finally, if the model meets performance criteria, it can be deployed in real-world applications, albeit with careful consideration of ethical implications and safeguards to ensure responsible use. Throughout the process, iteration and experimentation are essential for refining the model and optimizing its performance.

Table 1: Performance of the sound recognition algorithms

Signal-to-Noise Ratio (dB)	Rec. Rate (%) GMM Classifier	Rec. Rate (%) HMM Classifier
70	97.96	99.74
60	36.83	97.07
50	94.39	95.85
40	51.74	94.67
30	89.76	94.63
20	96.24	70.63
10	80.73	88.54
0	65.85	81.95
-10	52.44	59.01

Following training, evaluation measures the model's performance, employing metrics like accuracy, precision, recall, and F1-score, along with techniques like cross-validation and ROC curves. Subsequently, testing on unseen data gauges the model's generalization ability, ensuring robustness.

V. DETECTION AND RECOGNITION SYSTEM

When the detection stage is cascaded with the subsequent recognition stage, the classification performance listed in *Table 1* is observed to decrease, due to possible mis-matches between the true pulse time-location of the input signal, and the detected one. For that reason, the time location of the pulse start must be refined (time resolution of 200 samples) with an adaptive amplitude- based thresholding process. Global system recognition results are shown on *Table 2*, using a fixed signal durationof 0.75 seconds. In *Table 2*, bad-detection situations are ruled out, when a detection position error higher than 1 second appears on the attack of the signal.

A Detection and Recognition System for Human Scream Detection encompasses several interconnected components to accurately identify screams within a variety of sound environments. It begins with the collection of audio data, including screams, from diverse sources. This data undergoes preprocessing to clean out noise and enhance signal.

In real-time, incoming audio streams are processed, segmenting them into short frames for feature extraction and classification. Post-processing techniques refine detection results, improving accuracy by smoothing predictions and adjusting thresholds. Evaluation metrics validate system performance, ensuring precision and recall meet requirements

Table 2: Performance of the whole sound detection andrecognition system after removal of bad-detected signals:

SNR (dB)	Bad-detected Signals (%)	GMM Rect. Rate (%)	HMM Rec. Rate (%)
70	0	98.32	98.54
60	0	96.88	96.10
50	0	91.71	95.37
40	0	90.93	92.32
30	0	90.89	90.53
20	0	86.54	93.54
10	0	79.02	83.09
0	0.26	61.57	66.30
-10	18.24	42.06	59.15

VI. CRIME DETECTION AND ANALYSIS

Detecting and analyzing human screams for crime detection involves a multi-step process combining audio signal processing and machine learning techniques. Initially, audio data is captured from various sources like microphones or surveillance cameras.

This raw data undergoes pre-processing to remove background noise and distortions. Features such as pitch, intensity, and spectral characteristics are then extracted from the pre-processed audio. Machine learning models, like Support Vector Machines or Convolutional Neural Networks, are trained on labeled datasets to classify audio segments into scream and non-scream categories. Real-time detection involves segmenting audio streams, classifying each frame for screams, and integrating the results into existing crime detection systems.

Challenges include handling variations in acoustic environments, speaker characteristics, and ethical considerations regarding privacy and misuse of surveillance technologies. Despite challenges, robust feature extraction and classification methods are crucial for building accurate and reliable scream detection systems, facilitating rapid responses to critical situations.

VII. CONCLUSION AND FUTURE WORK

In the realm of human scream detection, researchers have identified key acoustic features that distinguish screams from other sounds, including high intensity, high pitch, and irregular waveform patterns. Leveraging machine learning techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), significant progress has been made in automatically detecting screams from audio.

However, this progress relies heavily on the availability of annotated datasets containing labeled scream instances to train and evaluate detection systems.

Despite advancements, challenges persist, including background noise interference, variations in scream characteristics, and ethical considerations regarding privacy and consent. Nonetheless, the practical applications of scream detection in security systems, emergency response, and healthcare underscore its importance.

VIII. ACKNOWLEDGEMENT

We take this opportunity to thank our teachers and friends who helped us throughout the project.

First and foremost, I would like to thank my guide for the project (Dr. Kumud Kundu, Head of Department, CSE(AI&ML)) for her/his valuable advice and time during development of the project.

We would also like to thank Dr. Kumud Kundu (HoD, CSE(AI&ML)) for her constant support during the development of the project.

REFERENCES

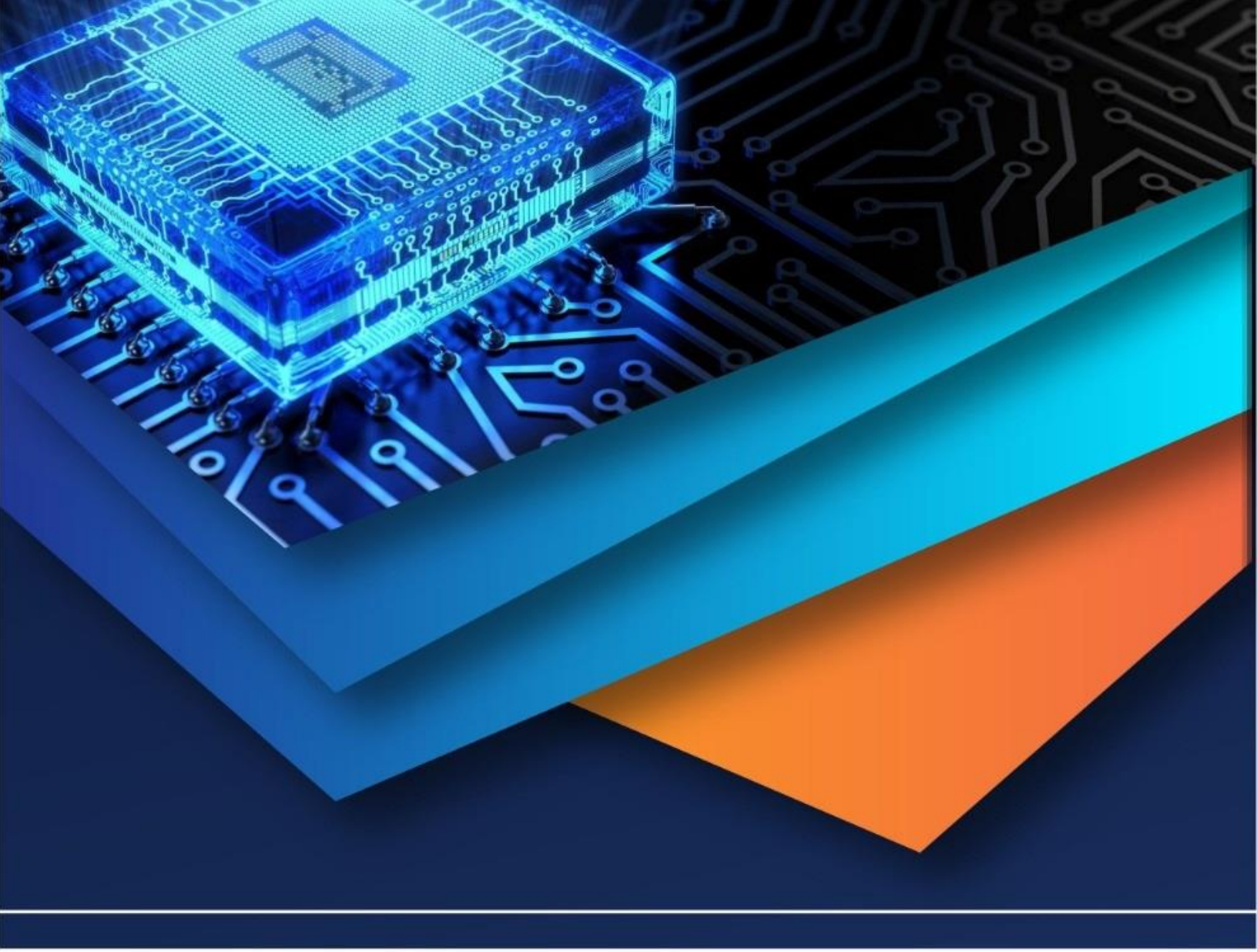
[1] M. Akbacak and J.H.L. Hansen, “Environmental sniffing: noise knowledge estimation for robust speech systems,” in Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP’03). 2003 IEEE International Conference on. IEEE, 2003, vol. 2, pp. II–113Leonardo Software, Santa Monica, CA 90401,

[2] S. Matos, S.S. Birring, I.D. Pavord, and D.H. Evans, “Detection of cough signals in continuous audio recordings using hidden Markov models,” Biomedical Engineering, IEEE Transactions on, vol. 53, no. 6, pp. 1078–1083, 2006.

[3] W.H. Liao and Y.K. Lin, “Classification of non-speech human sounds: Feature selection and snoring sound analysis,” in Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on, Oct 2009, pp. 2695–2700.

[4] L.S. Kennedy and D.P.W. Ellis, “Laughter detection in meetings,” in NIST ICASSP 2004 Meeting Recognition Workshop, Montreal. National Institute of Standards and Technology, 2004, pp. 118–121.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)