Home Security and Surveillance using glass breaking and baby crying detection and Audio Classification

Shivansh Mittal[1] | Sachin Garg[2]

[1]UG student, IT & Maharaja Agrasen Institute Of Technology,Delhi,India
[2]Asst. Prof. Department of IT & ,Maharaja Agrasen Institute Of Technology,Delhi,India

To Cite this Article
Shivansh Mittal and Sachin Garg, "Home Security and Surveillance using glass breaking and baby crying detection and Audio Classification", *International Journal for Modern Trends in Science and Technology*, 6(12): 459-462, 2020.

ABSTRACT

The main objective of this project is to detect audio which may be vulnerable to people living in a family environment and ensure security at their homes.

All the process from data collection to data cleaning and then training and testing the finally made Machine Learning model will be done for audio's like Glass breaking and Baby Crying Detection.

This model can be used in smart home devices to ensure safety and security at the homes of people.

KEYWORDS: *Glass Breaking Audio Detection, Baby Crying Audio Detection, Audio Classification, Librosa*

I. INTRODUCTION

This project is based on classification of Audio Samples and detection of audio. We will be first collecting the audio dataset for the project. Then will be cleaning the whole data and splitting it into equi-length samples. This will be followed by feature extraction from the audio samples fitting it into machine learning models and collecting the results providing best accuracies against test samples. The audio detection will be of audio data like Glass breaking, Baby crying, Gun Shot etc. to ensure safety and security in homes.

This project provides an overview to how Audio can be used as a surveillance tool to ensure safety and security at homes. Librosa plays a very vital role in Audio feature extraction in the project and helps in extraction of features like MFCC, Chroma, Spectral Contrast etc.

The project will be able to give a very a through idea about Audio Classification using Machine Learning and improving security in Future Homes.

Overview

Problem: Problem is to use Audio detection for ensuring security in homes.

Data: The data used here is taken from various sources like YouTube, FreeSound and has been cropped to a particular sample size.

The data taken is for audio classification and detection of audio like Glass breaking, Baby crying, Gun Shot detection etc.

Evaluation: The evaluation would be on the basis of weather the final model is able to detect the audio correctly or not.

Features: Would be using a lot of features to classify audio samples like MFCC (Mel- Frequency

Spectral Coefficient), Chroma, Spectral Contrast, Mel, Tonnetz etc.

Dataset-
• 1045 glass breaking mono audio samples of 6 seconds at 22.5kHz.
• 1350 baby crying mono audio samples of 6 seconds at 22.5kHz.
• 1700 noise mono audio samples of 6 seconds at 22.5kHz.

II. METHODOLOGY

The process description of this work include the following:

1. Setting up the data: In this the data is arranged and formatted such that it is compatible for the training usage and there are no missing values or misguiding arrangements.

2. Preparing the Audio data: In this process the Audio samples are now cropped to a particular length such that all the samples are of the same length and frequency.

3. Feature Selection and Fitting the model: Features are selected and fit to the machine learning model and predictions are made from these data sets and accuracies are recorded.

4. Improve the accuracy through experiments: The model accuracy has to be improved by changing the hyper parameters and making random and associated valued to these variables.

5. Save and reload the model for future predictions: The model that is trained is saved or logged in some manner such that the prediction labels can be determined by the trained model.

III. RESULTS

Features Extracted- 64 features using MFCC
Parameters Used for MFCC-
• n_fft= 2048
• hop_length= 512
As n_fft should be greater than sample length and should be a power of 2.

Creating Dataset
• Extracted MFCC features from dataset and created '.npy' files for training.
Training Model Used- Support Vector Machine
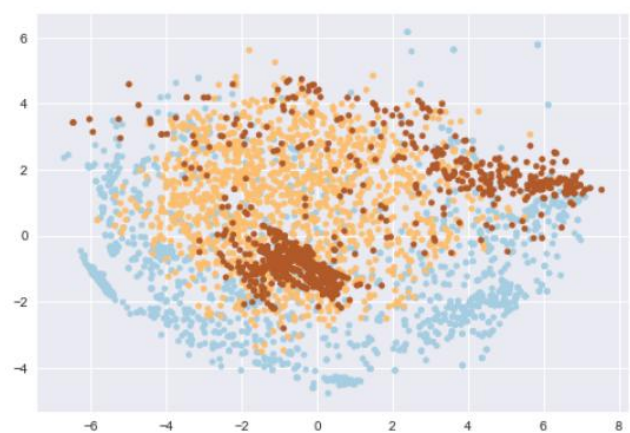
Features Table

S.No	Features	Explanation	Working
1.	MFCC	Graph plotting shows great contrast and less overlapping	YES
2.	Mel	Averaging lead to bad results	NO
3.	Tonnetz	Overlapping of Points	NO
4.	Spectral Contrast	Overlapping of Points	NO
5.	Chroma	Overlapping of Points	NO

Hyperplane Plots of Best Results

17. Cry+Noise(New)
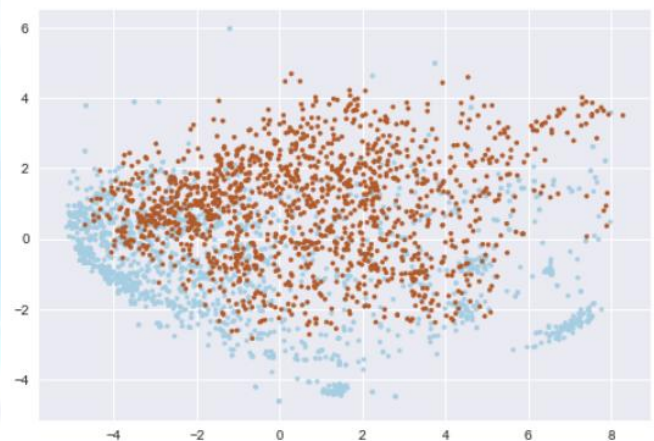
14. Cry+Glass+Noise(New)

RESULTS TABLE

S.No	Dataset (Sounds)	Features	Pre-emphasis/ Frequency	Parameters	Accuracy		
					Cry	Glass	Noise
1.	Cry+Noise	MFCC	1/16kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf	75.7%	-	90.6%
2.	Cry+Glass	MFCC	1/16kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf		91.05%	100%
3.	Cry+Glass+ Noise	MFCC	1/16kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf	81.4%	80%	82.6%
4.	Cry+Glass+ Noise	MFCC	5/16kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf	82.4%	80.4%	82.7%
5.	Cry+Glass+ Noise	MFCC	10/16kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf	81.9%	84.17%	80.4%
6.	Cry+Glass+ Noise	MFCC	3/16kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf	82.4%	81.2%	80.9%
7.	Cry+Glass+ Noise	MFCC	5/16kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf	80.9%	89.2%	81.6%
8.	Cry+Glass+ Noise	MFCC	1/16kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf	77.7%	84.7%	88.3%
9.	Cry+Glass+ Noise	MFCC	1/22.5kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf	83.2%	82.7%	82.01%
10.	Cry+Glass+Noise	MFCC	5/22.5kHz	'C': 15, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf	86.7%	86.7%	81.7%
11.	Cry+Glass+ Noise	MFCC	5/22.5kHz	'C': 15, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf	83.5%	81.6%	80.8%
12.	Cry+Glass+Noise	MFCC Normalized-Audio	5/22.5kHz	'C': 15, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf	89.9%	75.7%	82.7%
13.	Cry+Glass+Noise	MFCC Normalized-Audio	5/22.5kHz	'C': 30, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf	91.05%	84.7%	75.7%
14.	Cry+Glass+ Noise	MFCC Normalized-Audio+Feature	5/22.5kHz	'C': 35, 'degree': 2, 'gamma': 0.1, 'kernel': 'rbf	93.6%	80.6%	80.1%
15.	Cry+Glass+ Noise	MFCC Normalized-Feature	5/22.5kHz	'C': 10, 'degree': 2, 'gamma': 0.1, 'kernel': 'rbf	93.12%	81.6%	73.5%
16.	Cry+Noise	MFCC Normalized-Feature+Audio	5/22.5kHz	'C': 10, 'degree': 2, 'gamma': 0.1, 'kernel': 'rbf	90.4%	-	78.6%
17.	Cry+Noise	MFCC Normalized-Feature	5/22.5kHz	'C': 5, 'degree': 2, 'gamma': 0.1, 'kernel': 'rbf	87.3%	-	85.3%
18.	Cry+Noise	MFCC Normalized-Audio	5/22.5kHz	'C': 10, 'degree': 3, 'gamma': 0.01, 'kernel': 'poly'	84.1%	-	77.9%

IV. CONCLUSION

This project provides an overview to how Audio can be used as a surveillance tool to ensure safety and security at homes. Librosa plays a very vital role in Audio feature extraction in the project and helps in extraction of features like MFCC, Chroma, Spectral Contrast etc.

The project will be able to give a very a through idea about Audio Classification using Machine Learning and improving security in Future Homes.

Glass Breaking and Baby crying detection has been done and good accuracies have been found using SVM model with 64 normalized MFCC features with normalized audio.

V. REFERENCES

[1] Music Type Classification By Spectral Contrast Feature1 The work was performed at Microsoft Research Asia Dan-Ning Jiang, Lie Lu, Hong-Jiang Zhang, Jian-Hua Tao, Lian-Hong Cai

[2] B. Matityaho and M. Furst. "Neural Network Based Model for Classification of Music Type", in Proc. of 18th Conv. Electrical and Electronic Engineers in Israel, pp. 4.3.4/1-5, 1995.

[3] K. P. Han, Y. S. Pank, S. G. Jeon, G.C. Lee, and Y. H. Ha, "Genre Classification System of TV Sound Signals Based on a Spectrogram Analysis", IEEE Transactions on Consumer Electronics, VOL. 55, No. 1, pp. 33-42, 1998.

[4] H. Soltau, T. Schultz, Martin Westphal, and Alex Waibel. "Recognition of Music Types". ICASSP 1998, Vol. II, pp. 1137-1140, 1998.

[5] D. Pye. "Content-Based Methods for the Management ofDigital Music". ICASSP 2000, Vol. IV, pp.2437-2440, 2000.

[6] D. Li, I. K. Sethi, N. Dimitrova, T. McGee, "Classificationof General Audio Data for Content-BasedRetrieval",PatternRecognition Letters, VOL. 22, No. 5, pp. 533-544, 2001.

[7] S. Sadie as Editor, "The Cambridge Music Guide",Cambridge University Press, 1985.