Internship Report

ON

Glass Breaking

&

Baby Crying

Audio Classification

Shivansh Mittal

Arindam Jain

1st-June-1st August 2019

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.

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

Results:-

Features

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

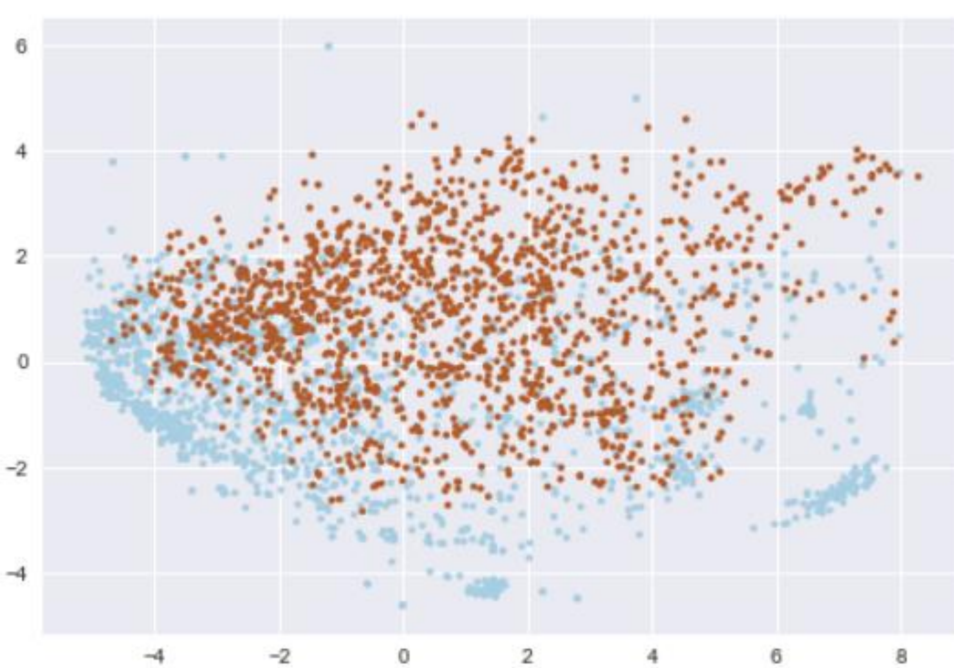
FINAL RESULTS

S.No	Dataset	Features	Pre-emphasis/ Frequency	Parameters	Accuracy		
					Cry	Glass	Noise
1.	Cry+Noise(Sir)	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(Sir)	MFCC	1/16kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf'	81.4%	80%	82.6%
4.	Cry+Glass+Noise(Sir)	MFCC	5/16kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf'	82.4%	80.4%	82.7%
5.	Cry+Glass+Noise(Sir)	MFCC	10/16kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf'	81.9%	84.17%	80.4%
6.	Cry+Glass+Noise(Sir)	MFCC	3/16kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf'	82.4%	81.2%	80.9%
7.	Cry+Glass+Noise(New)	MFCC	5/16kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf'	80.9%	89.2%	81.6%
8.	Cry+Glass+Noise(New)	MFCC	1/16kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf'	77.7%	84.7%	88.3%
9.	Cry+Glass+Noise(New)	MFCC	1/22.5kHz	'C': 10, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf'	83.2%	82.7%	82.01%
10.	Cry+Glass+Noise(New)	MFCC	5/22.5kHz	'C': 15, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf'	86.7%	86.7%	81.7%
11.	Cry+Glass+Noise(New)	MFCC	5/22.5kHz	'C': 15, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf'	83.5%	81.6%	80.8%
12.	Cry+Glass+Noise(New)	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(New)	MFCC Normalized-Audio	5/22.5kHz	'C': 30, 'degree': 2, 'gamma': 0.0001, 'kernel': 'rbf'	91.05%	84.7%	75.7%

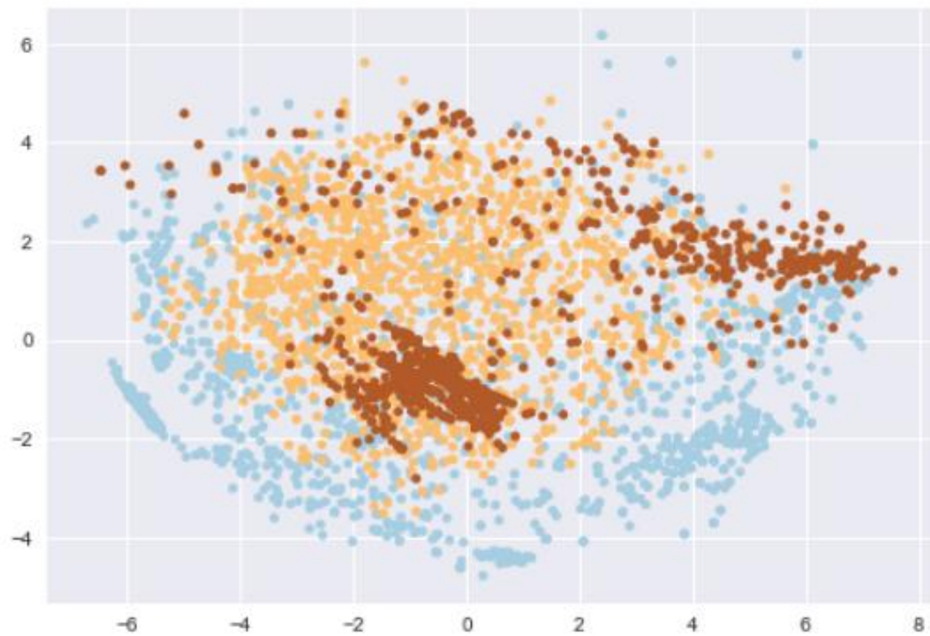
S.No	Dataset	Features	Pre-emphasis/ Frequency	Parameters	Accuracy		
					Cry	Glass	Noise
14.	Cry+Glass+Noise(New)	MFCC Normalized-Audio+Feature	5/22.5kHz	'C': 35, 'degree': 2, 'gamma': 0.1, 'kernel': 'rbf'	93.6%	80.6%	80.1%
15.	Cry+Glass+Noise(New)	MFCC Normalized-Feature	5/22.5kHz	'C': 10, 'degree': 2, 'gamma': 0.1, 'kernel': 'rbf'	93.12%	81.6%	73.5%
16.	Cry+Noise(New)	MFCC Normalized-Feature+Audio	5/22.5kHz	'C': 10, 'degree': 2, 'gamma': 0.1, 'kernel': 'rbf'	90.4%	-	78.6%
17.	Cry+Noise(New)	MFCC Normalized-Feature	5/22.5kHz	'C': 5, 'degree': 2, 'gamma': 0.1, 'kernel': 'rbf'	87.3%	-	85.3%
18.	Cry+Noise(New)	MFCC Normalized-Audio	5/22.5kHz	'C': 10, 'degree': 3, 'gamma': 0.01, 'kernel': 'poly'	84.1%	-	77.9%

BEST RESULTS- Hyperplane

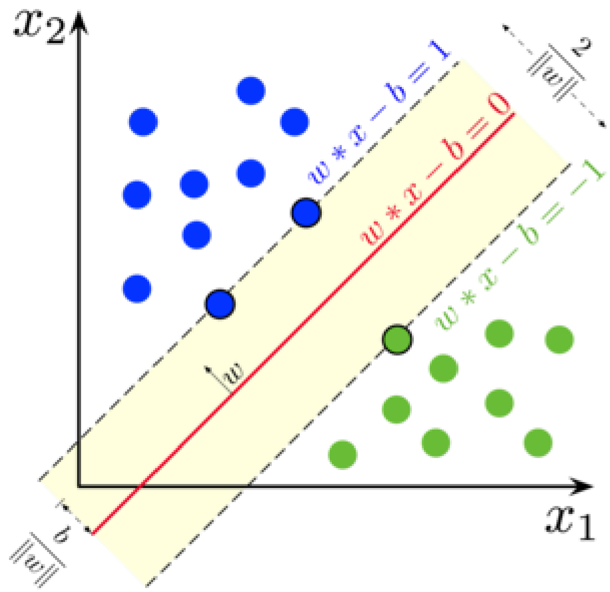
17. Cry+Noise(New)



14.Cry+Glass+Noise(New)



Support Vector Machine



More formally, a support-vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin, the lower the generalization error of the classifier.

MFCC

Mel-frequency cepstral coefficients (**MFCCs**) are coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip (a nonlinear "spectrum-of-a-spectrum"). The difference between the cepstrum and the mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum. This frequency warping can allow for better representation of sound, for example, in audio compression.

MFCCs are commonly derived as follows:

1. Take the Fourier transform of (a windowed excerpt of) a signal.
2. Map the powers of the spectrum obtained above onto the mel scale, using triangular overlapping windows.
3. Take the logs of the powers at each of the mel frequencies.
4. Take the discrete cosine transform of the list of mel log powers, as if it were a signal.
5. The MFCCs are the amplitudes of the resulting spectrum.

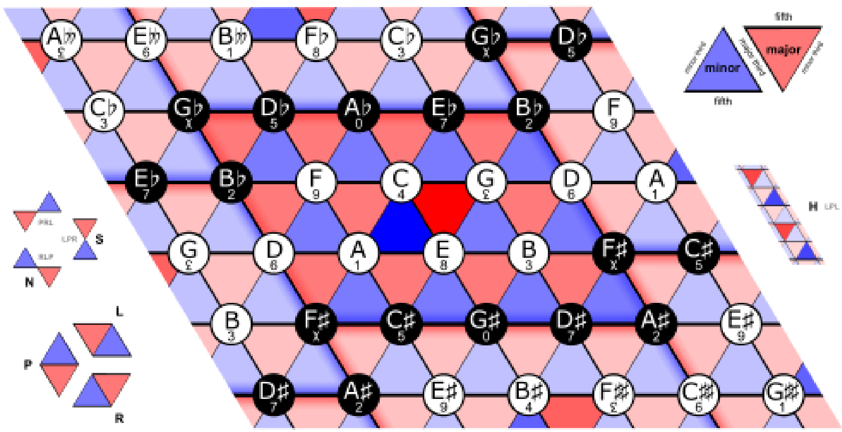
SPECTRAL CONTRAST

Octave-based Spectral Contrast considers the spectral peak, spectral valley and their difference in each sub-band. For most music, the strong spectral peaks roughly correspond with harmonic components; while non-harmonic components, or noises, often appear at spectral valleys. Thus, Spectral Contrast feature could roughly reflect the relative distribution of the harmonic and non-harmonic components in the spectrum.

(a)

TONNETZ

In musical tuning and harmony, the Tonnetz is a conceptual lattice diagram representing tonal space first described by Leonhard Euler in 1739. Various visual representations of the Tonnetz can be used to show traditional harmonic relationships in European classical music.



MEL SPECTOGRAM

A spectrogram is a visual representation of the spectrum of frequencies of a signal as it varies with time. When applied to an audio signal, spectrograms are sometimes called sonographs, voiceprints, or voicegrams. When the data is represented in a 3D plot they may be called waterfalls. When these spectrograms are drawn on the mel scale as y axis they are called mel spectrogram.