International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

Essential Features Extraction from Aaroh and Avroh of Indian Classical Raag - YAMAN

Bharavi U. Desai[1], Kunjal I. Tandel[2], Rahul M. Patel[3]

[1]Master's Student, Dept. of Electronics and Communication Engineering, Dr. S & S S Gandhy Government Engineering College, Gujarat, India.
[2]Professor, Dept. of Electronics and Communication Engineering, Government Engineering College, Gujarat, India.
[3]Professor, Dept. of Electronics and Communication Engineering, Dr. S & S S Gandhy Government Engineering College, Gujarat, India.

Abstract: *Algorithms of Audio signal processing generally involves analysis of signal, its properties extraction, behavior prediction, recognition of any pattern within the signal, and also the way in which a specific signal is correlated to a similar audio signal. Audio signals may comprise of speech, music and environmental sounds. Over the past few years, such audio signal processing has developed considerably in terms of signal analysis and classification. Also, it has been proven that many existing issues are often solved by integrating the smart machine learning (ML) algorithms with the audio signal processing techniques. The performance of any ML algorithm depends on the parameters/features on which the training and testing is concluded. Hence, feature extraction is an important part of a machine learning process. The intend of this research work is to summarize the literature of the audio signal processing which specializes in the feature of aaroh and avroh of Indian classical raag. During this survey the frequency domain, temporal domain, wavelet domain, cepstral domain and time frequency domain features are extracted and discussed thoroughly.*

Key words: ICM, Aaroh-Avroh, Machine learning, MFCC, STFT, CQT, RMSE, Delta RMSE, Onset detection.

1. INTRODUCTION:

Feature extraction is one of the important step in audio analysis. Mainly, feature extraction is an essential processing step in pattern recognition and machine learning tasks. The objective is to extract a set of features from the dataset of interest. These extracted features must be informative in terms of desired properties of the original data. One can consider feature extraction as a procedure for data rate reduction since we need our analysis algorithms to be based on a comparatively small number of features. Here, we have extracted features for aaroh and avroh of raga Yaman and analyzed them. LibROSA, which is a python package, is used for music and audio analysis. It helps in providing the building blocks required to create music information retrieval systems. Time domain feature, chroma features and spectral features are covered in this paper. Extracted features are used for audio classification, recognition, generation, Genre classification and Mood detection.

2. Indian Classical Raag YAMAN:

A raga is only a melodic structure with fixed notes in North Indian traditional music with set of rules [16]. Raag Yaman for Peace, Happiness is generally performed distinctly during the early night. It passes on a temperament that is quiet, quiet, and tranquil and simultaneously euphoric and energetic [18, 19]. The notes in a Raag Yaman generally relate to the accompanying notes in the western scale, in the key of D:

Thaat: Kalyan
Aaroha: Ni Re Ga Ma(Kori Ma/tivra Ma i.e. Ma#) Pa Dha Ni Sa,
Avroh: Sa Ni Dha Pa Ma ((Kori Ma/tivra Ma i.e. Ma#)) Ga Re Sa.

3. AUDIO FEATURE EXTRACTION:

Sound features can be defined as mathematical algorithms that can be put into practice, either by software or hardware, to pull out helpful information from the signal which is not observable from the raw data. Features are extracted either from the frequency (spectral) domain or the time (temporal) domain. The signal is analyzed with respect to time in time domain, while in the frequency domain the signal is analyzed with respect to frequency. Spectrograms are also used to extract features that hold spectral as well as temporal information.

4. Time Domain Features :

Time domain refers to the analysis of mathematical functions, time series of economic or environmental data or physical signals, with respect to time. In the case of time domain, the signal or function's values are identified for all real numbers, be it a case of continuous time or discrete time. An oscilloscope is a tool generally used to visualize real-world signals in the time domain.

4.1 Zero crossing rate:

The rate at which a signal changes its sign during the frame is known as "Zero crossing rate (ZCR)". It highlights the number of times the signal changes value, from positive to negative and vice versa, which is to be divided by the total length of the frame [1].

$$Z(i) = \frac{1}{2N}\sum_{n=0}^{N-1}|sgn[x_i(n)] - sgn[x_i(n-1)]| \quad \text{... (1)}$$

ZCR is considered as a measure of the noisiness for any signal. Likewise, it generally gives higher values in the case of noisy signals. It is also known to reflect, up to some extent, the spectral characteristics of a signal.

Characteristics:

a) Noise and unvoiced sound have high ZCR.
b) ZCR is generally used in endpoint detection, particularly in recognition of the start and end of unvoiced sound.
c) To differentiate noise/silence from unvoiced sound, normally we include a shift before calculating ZCR.

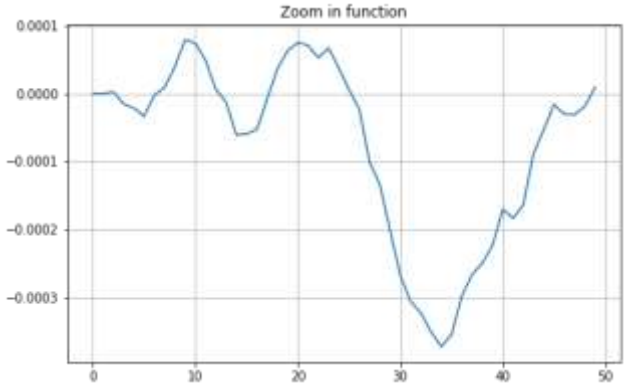

Fig.-4.1(a): Zoom in for 8000< n <9000 (Aaroh)

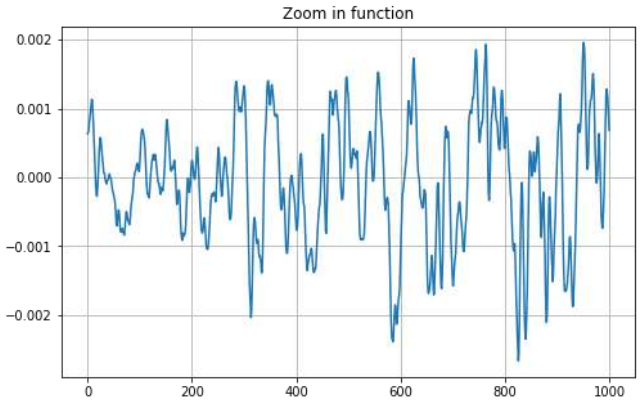

Fig-4.1(b) : Zoom in for 8000< n <9000 (Avroh)

4.2 Autocorrelation :

The autocorrelation method is related to the time domain feature detector. The autocorrelation of the signal can be projected in below mentioned equation :

$$x(n) \otimes x(\tau) = \sum_{i=q}^{q+N-1} x(n) \times x(n+\tau) \quad \text{... (2)}$$

Here, the speech sample sequence – x(n) is multiplied by a rectangular window of length N, and τ is the lag number. The value of τ varies between 0 and N-1. The major peak in the autocorrelation function is on the zero lag location (τ = 0). The location of the subsequent peak provides an approximation of the period, and the height provides an indication of the periodicity of the signal. For analog signals this estimation is specified by equation[15].

$$r(\tau_{max}) = max_{\tau}\, r(\tau) \quad \text{... (3)}$$

Below figure shows the Autocorrelation of a particular audio segment from aaroh and avroh respectively of Raag Yaman.

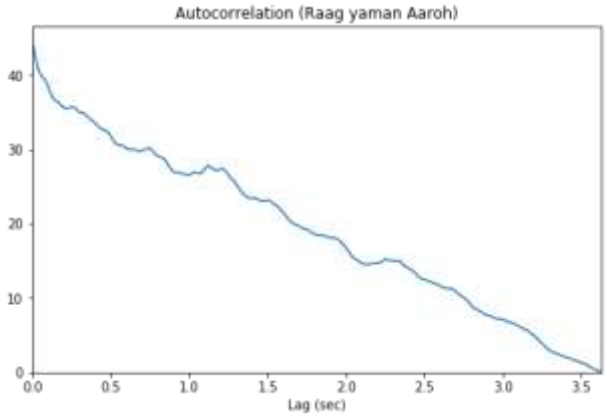

Fig.-4.2(a): Autocorrelation of a segment (Aaroh)

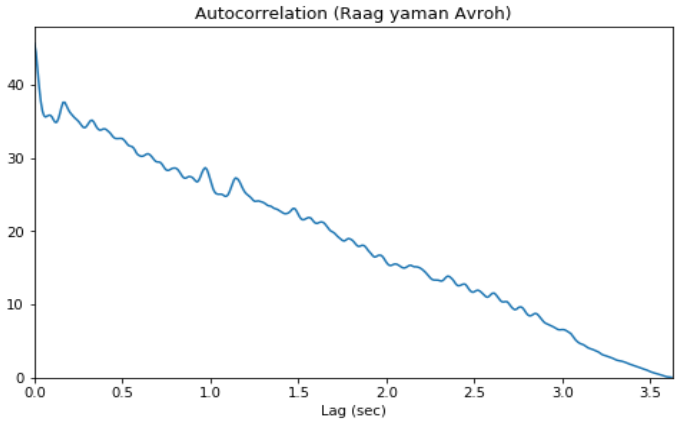

Fig.-4.2(b): Autocorrelation of a segment (Avroh)

5. SPECTRAL FEATURES:

Spectral features are also known as frequency based features and can be obtained using the Fourier Transform, by converting the time based signal into the frequency domain like: spectral density, spectral centroid, spectral flux, fundamental frequency, spectral roll-off, frequency components etc. These features are useful to recognize the notes, pitch, rhythm and melody.

5.1 Spectrogram:

In feature extraction of a speech, frequency (spectral) analysis is very important. Human speech can be considered to be reasonably stationary over the study interval of 20- 25 msec. Hence, the signal is analyzed in consecutive narrow time frames of 20-25 ms width. For the speech signals, the spectral analysis is carried out by identifying Discrete Fourier Transform (DFT) of the samples in the frame. Fast Fourier Transform algorithm is used to calculate DFT [14].

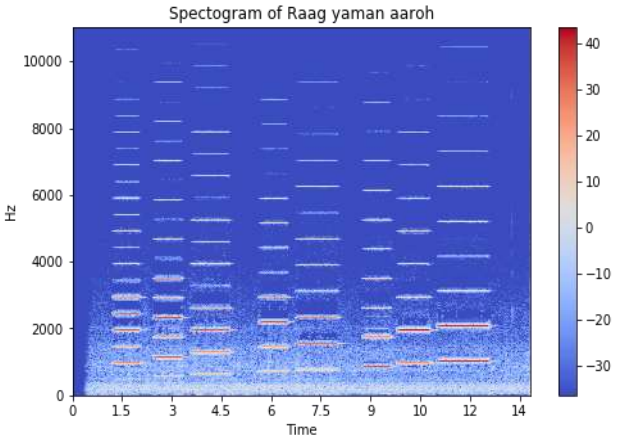

Fig.-5.1(a): Spectogram (Aaroh)

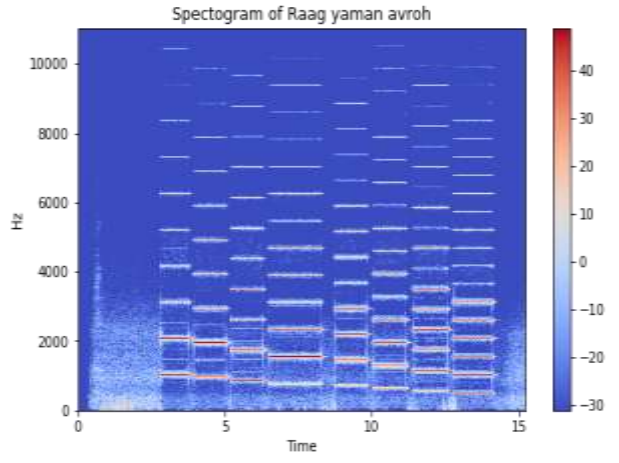

Fig.-5.1(b): Spectogram (Avroh)

5.2 Mel-Spectrogram :

The Mel Scale is the outcome of non-linear transformation of the frequency scale. The Mel Scale is created in such a manner that sounds of equal distance from each other on the Mel Scale, also "sound" to humans as they are equal in distance from one another. For Hz scale, the difference between 500 and 1000 Hz is obvious, while the difference between 7500 and 8000 Hz is hardly noticeable. After having an idea about Spectrogram, and also about Mel Scale, so the Mel Spectrogram, is, a Spectrogram with the Mel Scale as its *y* axis.

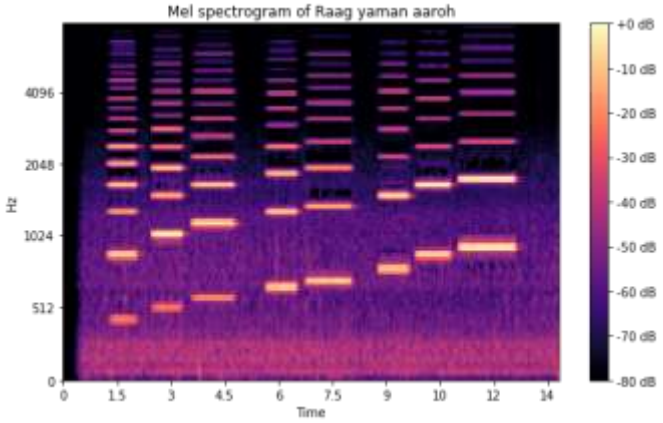

Fig.-5.2(a): Mel Spectrogram (Aaroh)

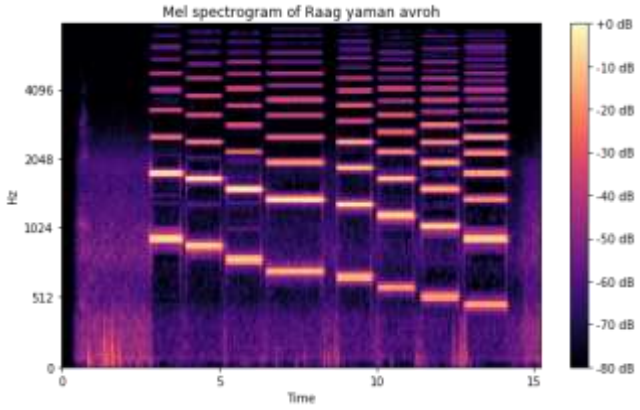

Fig.-5.2(b): Mel Spectogram (Avroh)

5.3 Spectral centroid:

The spectral centroid is a measure used in digital signal processing to characterize a spectrum. It indicates where the center of mass of the spectrum is located [1]. Perceptually, it has a robust connection with the impression of brightness of a sound. The spectral centroid are two simple measures of spectral position and shape. The center of 'gravity' of the spectrum is called spectral centroid. Here, the value of spectral centroid (C_I), is defined for the i^{th} audio frame as:

$$C_i = \frac{\sum_{k=1}^{W_{f_L}} kX_i(k)}{\sum_{k=1}^{W_{f_L}} X_i(k)} \quad \text{... (4)}$$

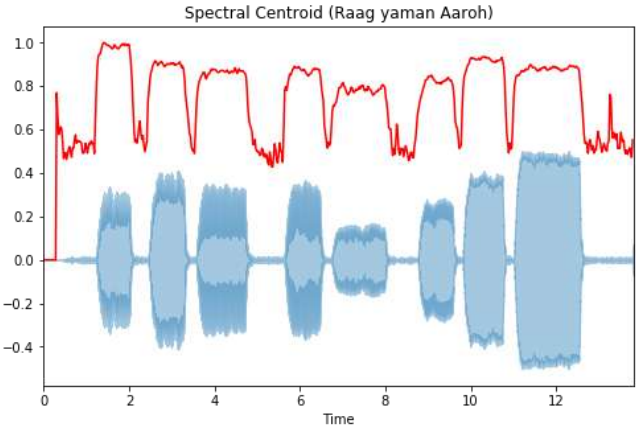

Fig.-5.3(a): Spectral Centroid (Aaroh)

To extract the mean (centroid) from each frame, every frame of a magnitude spectrogram is normalized and treated as a distribution over frequency bins. Higher values of spectral centroid represent the brighter sounds.

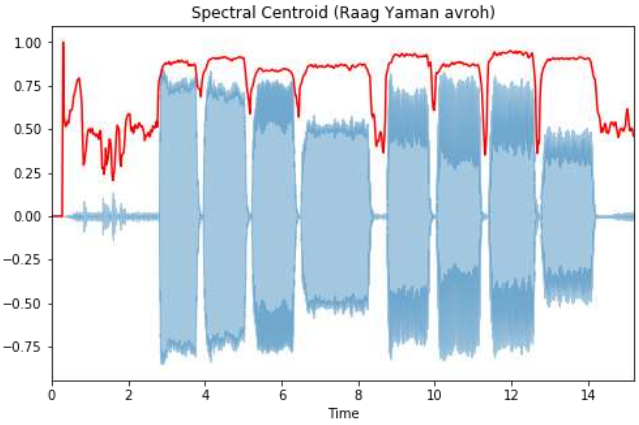

Fig.-5.3(b): Spectral Centroid (Avroh)

5.4 Spectral roll off:

It is the frequency under which around 90% of the magnitude distribution of the spectrum is concentrated. Hence, it satisfies the following equation if the m^{th} DFT coefficient corresponds to the spectral roll off of the i^{th} frame [1]:

$$\sum_{k=1}^{m} X_i(k) = C \sum_{k=1}^{W_{f_L}} X_i(k) \quad \text{... (5)}$$

Where C is the adopted percentage. The spectral roll off frequency is generally normalized by dividing it with W_{f_L}, so that it gets values between 0 and 1. This type of normalization means that a value of 1 corresponds to the maximum frequency of the signal, i.e. to half the sampling frequency.

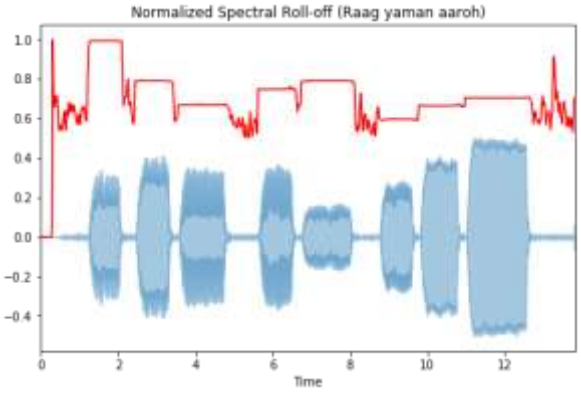

Fig.-5.4(a): Spectral Roll-off (Aaroh)

Spectral roll off can be also treated as a spectral shape descriptor of an audio signal and can be used for differentiating between voiced and unvoiced sounds. It can also be used to differentiate between various types of music tracks.

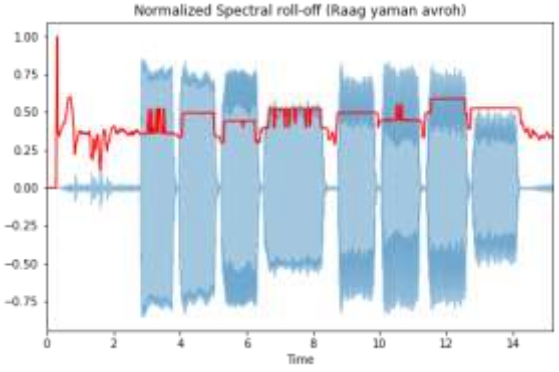

Fig.-5.4(b): Spectral Roll-off (Avroh)

5.5 Spectral skewness:

Spectral skewness is the 3rd order statistical value. Symmetry of the spectrum around its arithmetic mean value is measured by this. It would be equal to zero for silent segments and high for voiced parts.

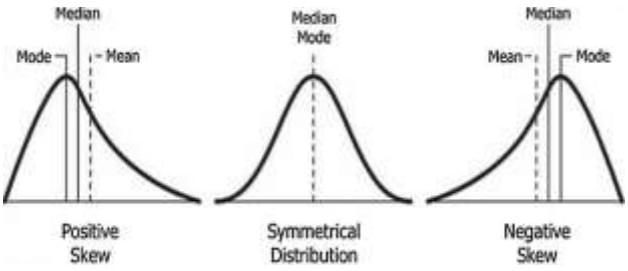

Fig.-5.5: Effect of skewness value

Skewness equal to zero represents symmetric distribution, skewness less than zero is a sign of more energy to the right side of spectral distribution and skewness greater than zero describes more energy components are present on the left side of the spectrum. This feature is used in mood detection, music genre classification, fault finding in motor bearings and Parkinson's disease detection from speech.

5.6 Spectral Contrast:

Every frame of a spectrogram *S* is divided into sub-bands. Comparison of the mean energy in the top quantile (peak energy) to that of the bottom quantile (valley energy) is done to estimate the energy contrast of every sub band.

Especially, narrow-band signals are represented by high contrast values, while broad-band noise are represented by low contrast values.

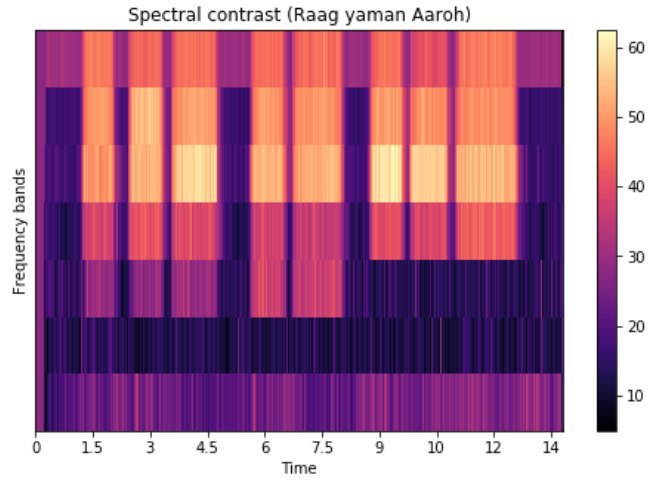

Fig.-5.6(a): Spectral contrast (Aaroh)

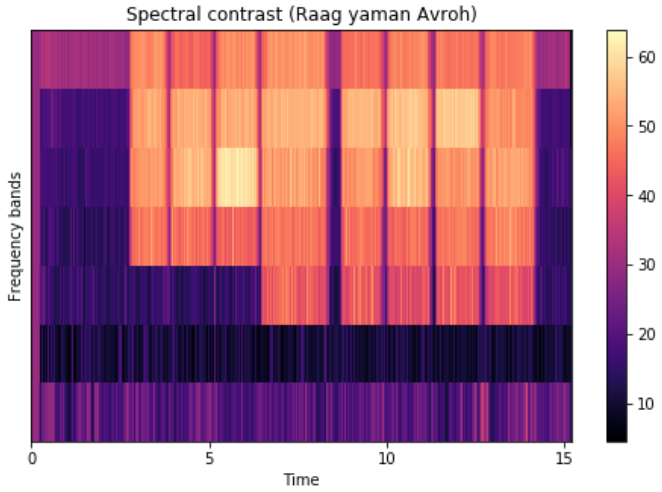

Fig.-5.6(b): Spectral contrast (Avroh)

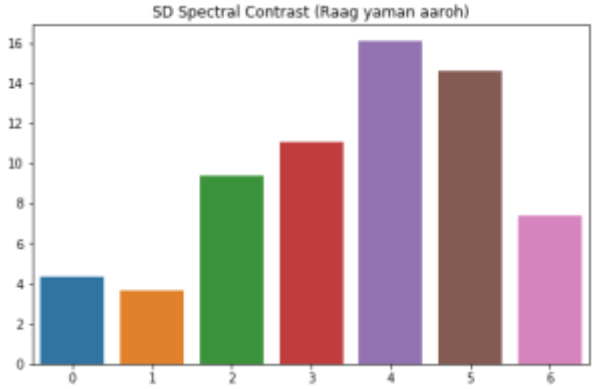

Fig.-5.6(c): Mean Spectral contrast (Aaroh)

Spectral Contrast extracts the spectral valleys, peaks and differences between their each sub-band while the FFT amplitudes are summed up by MFCC.

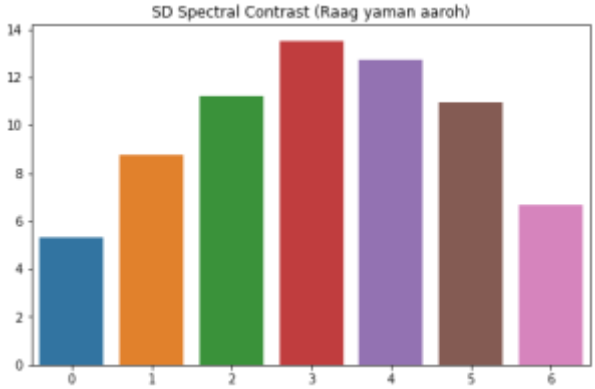

Fig.-5.6(d): Spectral contrast (Aaroh)

Thus, Spectral Contrast feature correspond to the relative spectral characteristics whereas MFCC only involves the average spectral information. Spectral Contrast includes more spectral information in comparison to MFCC.

5.7 Onset Detection :

Onset detection function is mainly an under-sampled version of the original music signal. As per the preprocessing phase, the signal is separated into partially overlapping frames and the ODF consists of one value for each frame [9].

By the definition of an onset, the onset detection is the method of identifying which parts of a signal are comparatively unpredictable. Therefore, each value in an ODF ought to give a fair indication as to the measure of the unpredictability of that particular frame. For onset detection, the vector of these values is passed to the peak-detection algorithm.

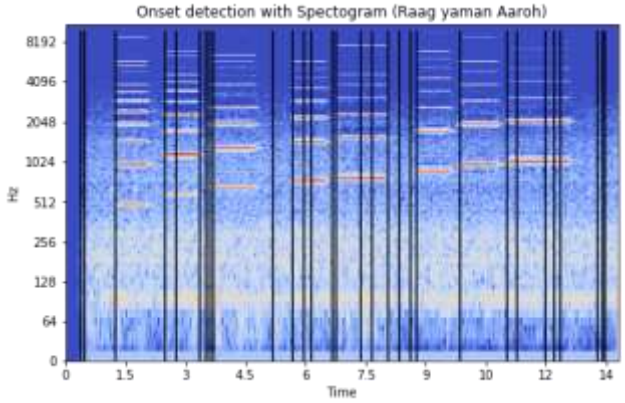

Fig.-5.7(a): Onset detection (Aaroh)

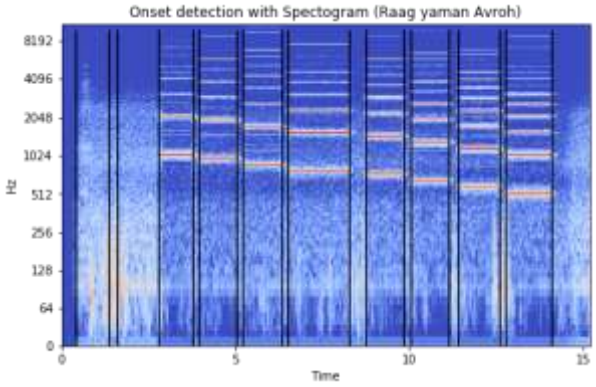

Fig.-5.7(b): Onset detection (Avroh)

5.8 Spectral Bandwidth :

Spectral bandwidth is considered as the 2nd order statistical value which determines the low bandwidth sounds from the high frequency sounds. It is generally used in music classification and environmental sound recognition [1].

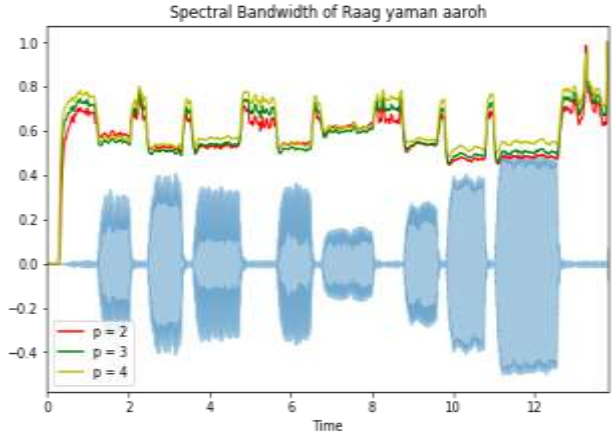

Fig.-5.8(a): Spectral Bandwidth for p=1,2,3 (Raag yaman aaroh)

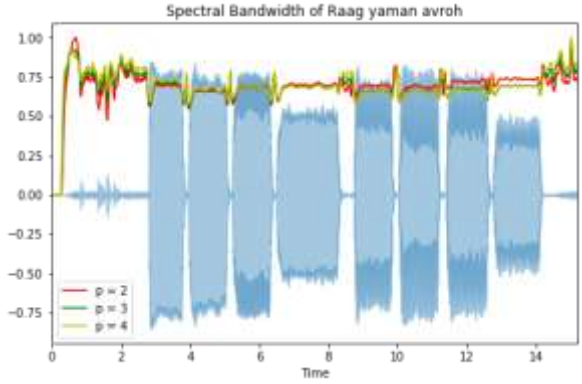

Fig.-5.8(a): Spectral Bandwidth for p=1,2,3 (Raag yaman avroh)

6. CEPSTRAL DOMAIN FEATURES:

A cepstrum is obtained by considering the inverse Fourier transform of the logarithm of the signal spectrum. There are various kinds of cepstrum like complex, power, phase and real cepstrum. Out of all these, power cepstrum is the most significant to the speech signal processing. The analysis of the cepstrum is called as cepstrum analysis, quefrency analysis (corresponding to frequency analysis in spectrum domain) or liftering (corresponding to the filtering in spectrum domain)[3]. For pitch detection, speech recognition and speech enhancement, these cepstrum features are mainly used. The cepstrum/cepstral features are discussed below.

6.1 MFCC features:

MFCC feature progress around perceptual or computational considerations. As this feature captures some of the fundamental properties used in human hearing, it is considered ideal for general audio discrimination. The MFCC feature has been effectively applied to speech recognition, music modeling and audio information retrieval and in recent times, it has been used in audio surveillance [1].

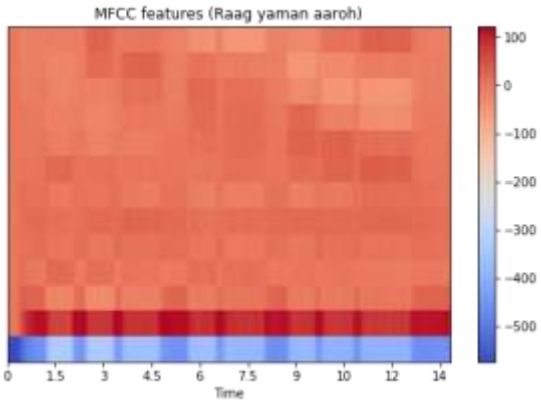

Fig.-6.1(a): MFCC Features (Aaroh)

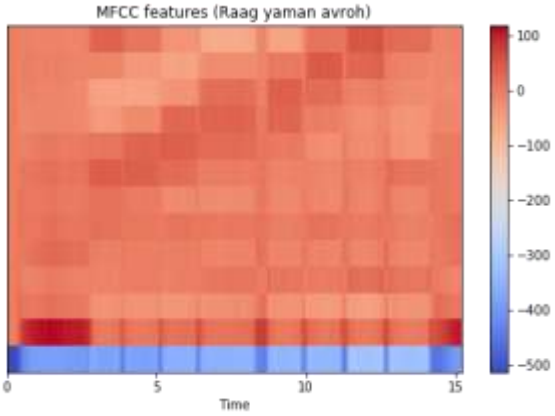

Fig.-6.1(b): MFCC Features (Avroh)

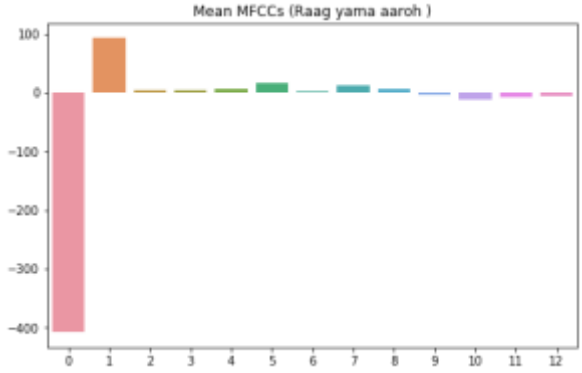

Fig.-6.1(c): Mean MFCCs (Aaroh)

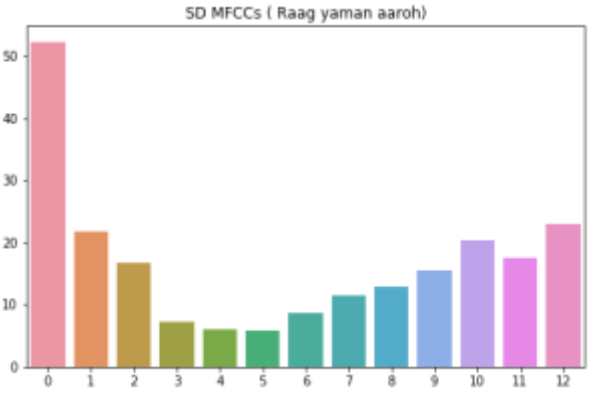

Fig.-6.1(d): SD MFCCs (Aaroh)

The Mel Frequency Cepstral Coefficients (MFCCs) of a signal briefly describe the overall shape of a spectral envelope. By printing the shape of MFCCs you get number of MFCCs against the number of frames where the first value represents the number of MFCCs calculated and another value represents a number of frames available [6].

7. CHROMA FEATURES:

Chroma features are engaging and powerful illustration for music audio. Under this, the whole spectrum is projected on to 12 bins representing the 12 separate semitones (or chroma) of the musical octave.

The chroma feature is a descriptor, which represents the tonal content of a musical audio signal in a condensed form. Therefore chroma features can be considered as important prerequisite for high-level semantic analysis, like chord recognition or harmonic similarity estimation.

Since, in music, notes exactly one octave apart are perceived as particularly similar, understanding the distribution of chroma even without the absolute frequency (i.e. the original octave) can give helpful musical information about the audio and may even reveal perceived musical similarity that is not noticeable in the original spectra[11].

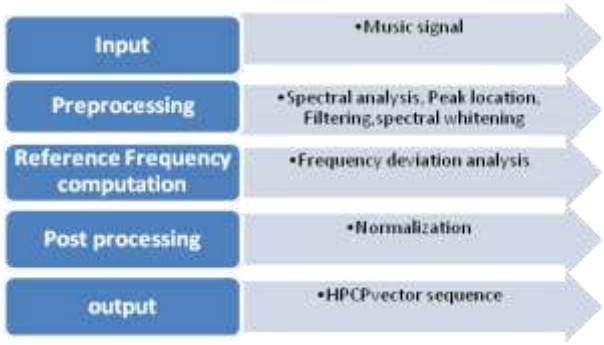

Chart-7: Steps for Chroma feature extraction

7.1 Constant Q- transform:

A constant Q transform is a bank of filters, but in contrast to the fourier transform, it has geometrically spaced center frequencies [11,12]:

$$f_k = f_0 \cdot 2^{\frac{k}{b}} \quad (k = 0, \ldots) \qquad \ldots (6)$$

Here, b indicates the number of filters per octave.

It is important to note that an appropriate choice for f_0 (minimal center frequency) and b the center frequencies directly correspond to musical notes, which makes the constant Q transform so useful.

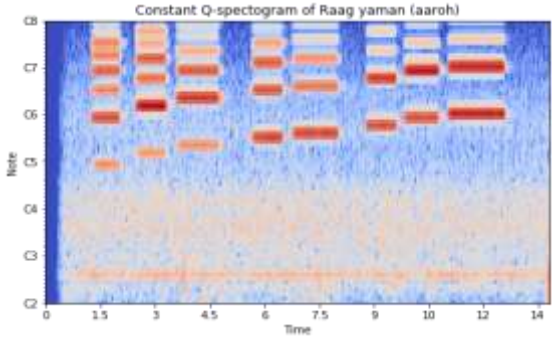

Fig.-7.1(a): Constant Q-Spectogram (Aaroh)

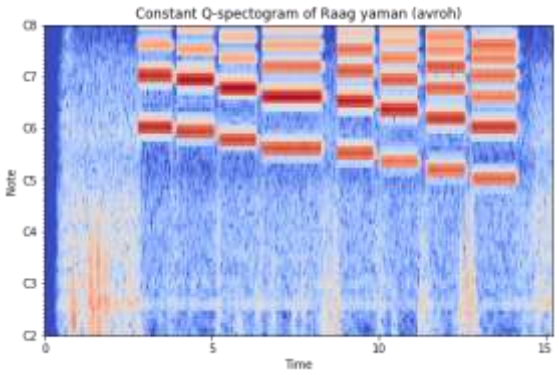

Fig.-7.1(b): Constant Q-spectrogram (Avroh)

7.2 CENS features

CENS (Chroma Energy Normalized Statistics) features can be obtained by some extra degree of abstraction after considering short-time statistics over energy distributions within the chroma bands, CENS features cover a family of scalable and robust audio features. These features are very useful in audio matching and retrieval applications [12]. A quantization is applied based on logarithmically chosen thresholds for computation of CENS features.

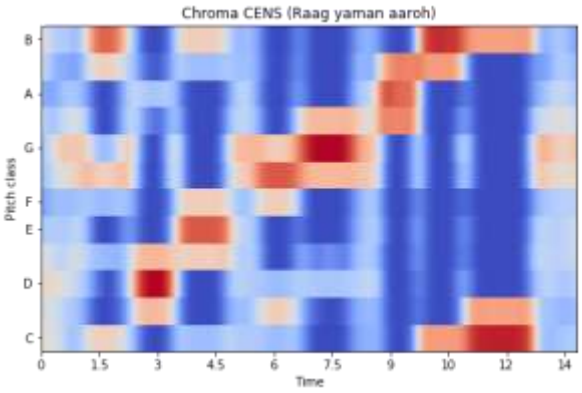

Fig.-7.2(a): Chroma CENS (Aaroh)

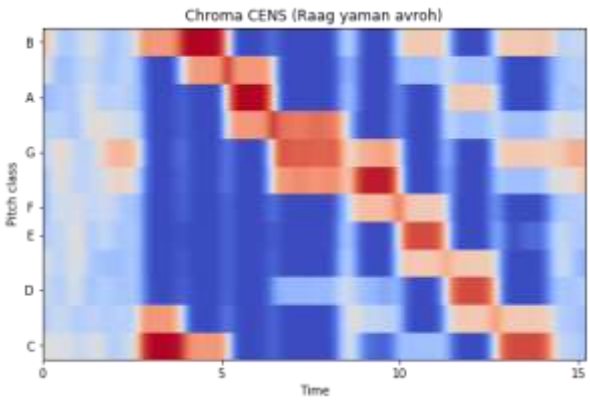

Fig.-7.2(b): Chroma CENS (Avroh)

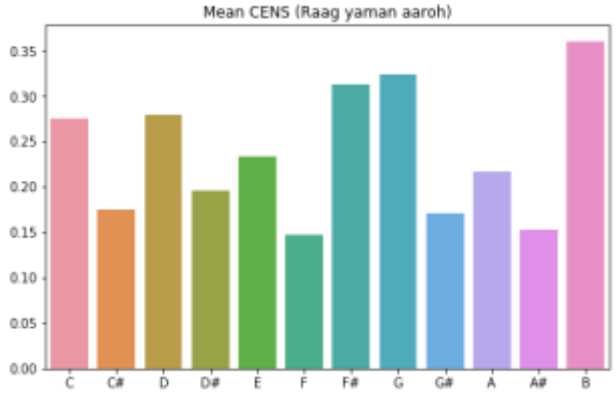

Fig.-7.2(c): Mean CENS (Aaroh)

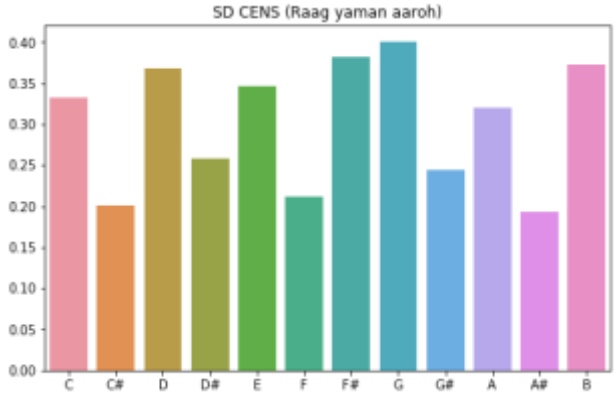

Fig.-7.2(d): SD CENS (Aaroh)

This brings in some kind of logarithmic compression like the CLP[η] features. Additionally, these features allow for introducing a temporal smoothing. Feature vectors are averaged using a sliding window technique depending on a window size denoted by w (given in frames) and a down sampling factor denoted by d [12]. In the following, we do not change the feature rate and consider only the case d = 1 (no down sampling). Therefore, the resulting feature only depends on the parameter w and is denoted by CENS[w].

7.3 CRP features

To increase the degree of timbre invariance, a new family of chroma-based audio features has been established. The common thought is to get rid of timbre-related information in a similar fashion as pitch-related information is discarded in the calculation of mel-frequency cepstral coefficients (MFCCs). After applying a logarithmic compression for pitch features, the logarithmized pitch representation is transformed, using a DCT [12].

After that we need to keep only the upper coefficients of the resulting pitch-frequency cepstral coefficients (PFCCs), apply an inverse DCT and lastly projects the

resultant pitch vectors onto 12-dimensional chroma vectors. These vectors are known as CRP (Chroma DCT Reduced log Pitch) features. The upper coefficients to be kept are specified by a parameter p ∈ [1: 120].

7.4 CP features:

We can obtain a chroma representation from the Pitch representation by mainly adding up the matching values that is pertaining to the same chroma. To archive invariance in dynamics, we normalize each chroma vector relating to the Euclidean norm. The resultant features are referred to as Chroma-Pitch represented by CP.

7.5 CLP Features:

These features utilize to justify the logarithmic sensation of sound intensity, logarithmic compression is applied while computing audio features. Before deriving the chroma representation, the local energy values e of the pitch representation are logarithmized [11]. Whereas each entry e is replaced by the value log (η · e+1), where η is a appropriate positive constant. The resultant features, are referred to as Chroma-Log-Pitch indicated by CLP[η], which depend on the compression parameter η.

8. Energy based Features:

To differentiate between noises, low, medium and high energy regions within the audio signal, Active speech level algorithm (ASL) is used. ASL finds out speech activity factor (Spl) which correspond to the fraction of time where the signal is considered to be active speech and the matching active level for the signal's speech part. The speech activity algorithm calculates the speech energy value at every sample time frame [1].

8.1 Energy:

The short-term energy is computed according to the below mentioned equation:

$$E(i) = \frac{1}{W_L} \sum_{n=1}^{W_L} | x_i(n)|^2 \quad \text{... (7)}$$

Here, x_i (n), n=1, W_L be the sequence of audio samples of the i^{th} frame, where W_L is the length of the frame.

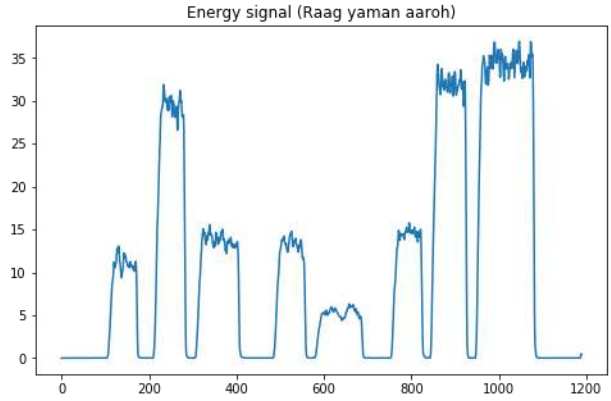

Fig.-8.1(a): Energy signal (Aaroh)

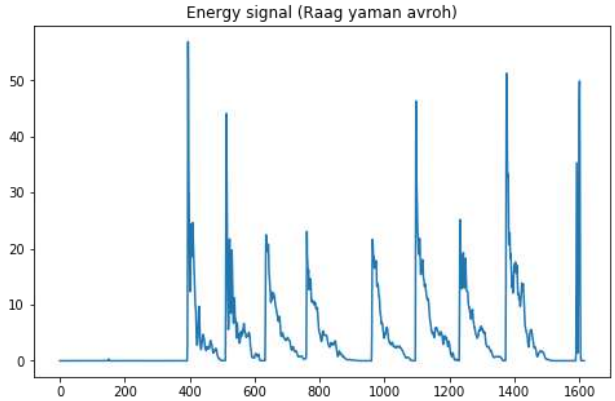

Fig.-8.1(b): Energy signal (Avroh)

8.2 Root Mean Square Error:

The standard deviation of the residuals (prediction errors) is considered as Root Mean Square Error (RMSE). Residuals quantify about distance between the regression line and data points; RMSE is a measure of how spread out these residuals are

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \quad \text{... (8)}$$

In other words, it tells about how concentrated the data is around the line of best fit. Root mean square error is usually used in climatology, forecasting, and regression analysis to verify investigational results [5]. Below figures are having the plot of RMSE, Delta RMSE and Energy Novelty functions in a single Frame.

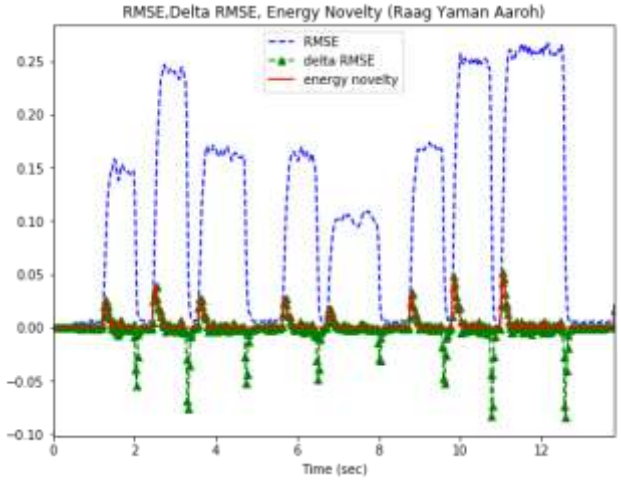

Fig.-8.2(a): Energy based features (Aaroh)

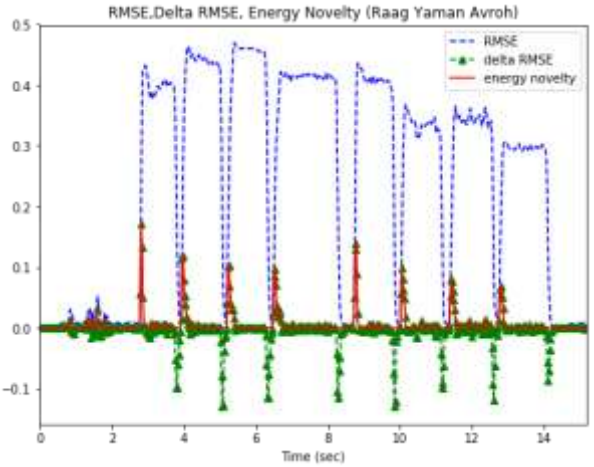

Fig.-8.2(b): Energy based features (Avroh)

8.3 Novelty Function:

To identify note onsets in audio, we need to locate sudden variations in the audio signal which mark the start of transient regions. Many times, an increase in the signal's amplitude envelope will indicate an onset candidate.

Sometimes notes can vary from one pitch to another without changing amplitude, e.g. slurred notes played on violin. Novelty functions denote local variation in signal properties such as energy or spectral content [17]. We will analyze about two novelty functions named below:

i. Energy-based novelty functions
ii. Spectral-based novelty functions

The graphs shown in the figure illustrates the energy Distribution of entire signal.

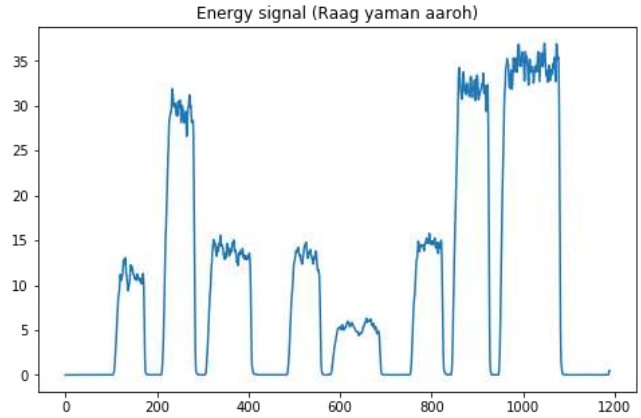

Fig.-8.3(a): Energy based Novelty function (Aaroh)

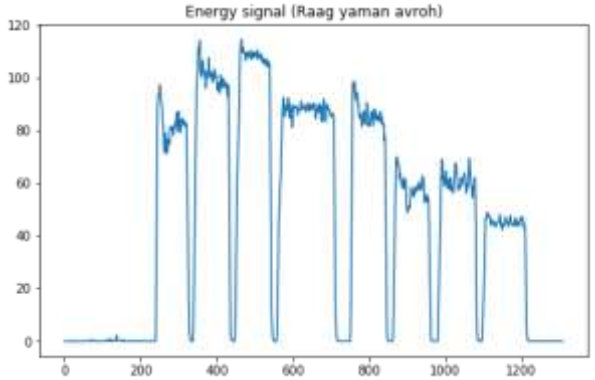

Fig.-8.3(b): Energy based Novelty function (Avroh)

9. Other Audio Features:

More to this there are other Audio features that I have tried to extract from aaroh and avroh of Raag yaman, which are given below. Those are also very informative and could help in to classifying genre and detection of mood from the given audio data.

Fig.9 (a) is showing the Bit locations on a given waveform where Fig.9(b) shows the Harmonic and its percussive plotting in mono frame.

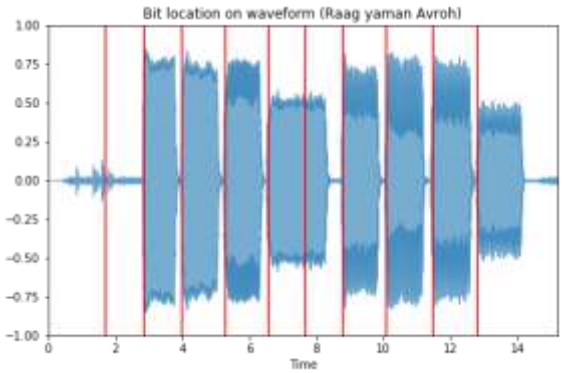

Fig.-9(a): Bit locations (Raag yaman avroh)

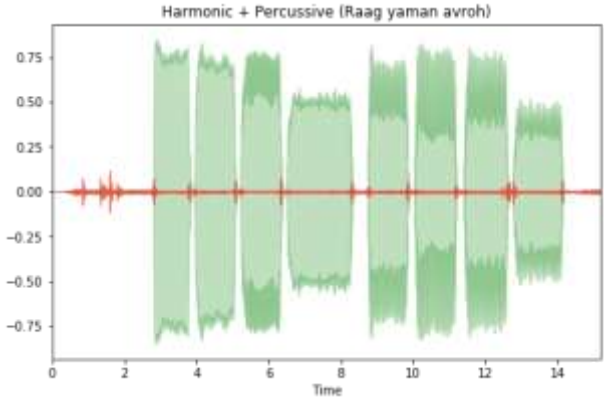

Fig.-9(b): Harmonic with percussives (Raag yaman avroh)

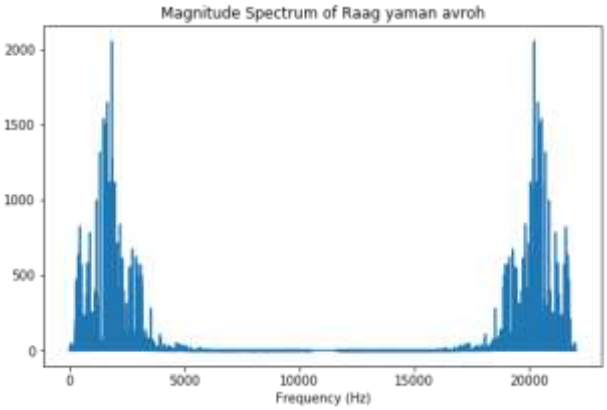

Fig.-9(c): Magnitude Spectrum (Raag yaman avroh)

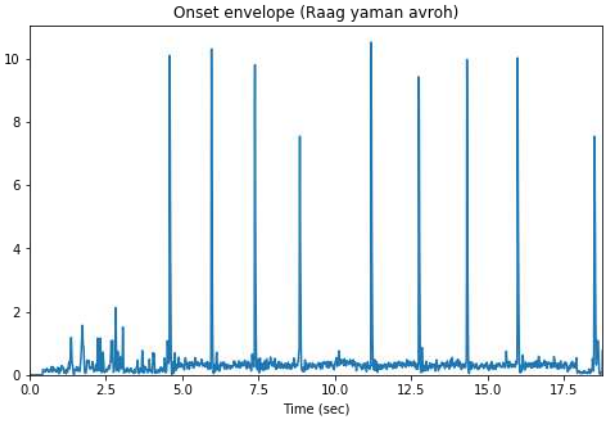

Fig.-9(d): Onset envelope (Raag yaman avroh)

Fig.9(c) have the plot of magnitude spectrum of Raag yaman avroh, Fig.9(d) figure shows the envelope of Onsets,Fig.9(e) shows the fourier tempogram.Fourier tempogram is basically a magnitude spectogram of the novelty function and the last figure shows the histogram of interval between adjacent beats. These are unique features by their characteristics and helps to Retrive the information of Raag yaman.

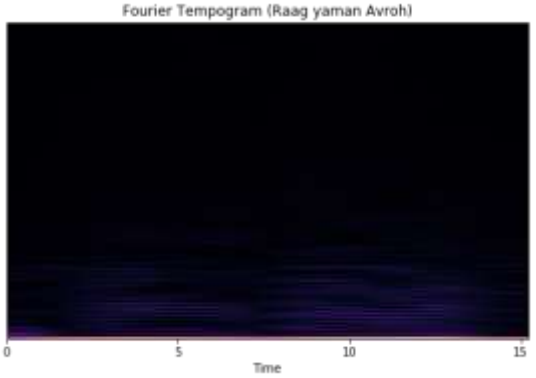

Fig.-9(e): Fourier Tempogram (Raag yaman avroh)

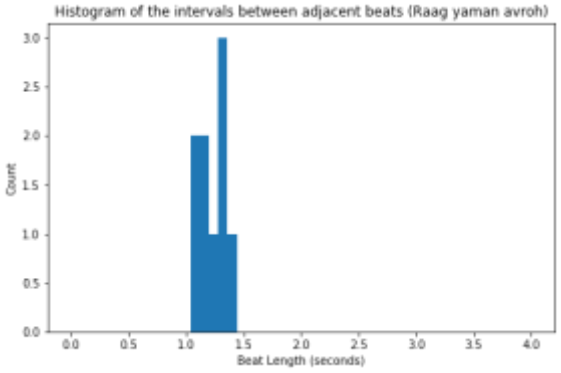

Fig.-9(d): Interval Histogram (Raag yaman avroh)

CONCLUSION:

In this paper, we have extracted the most essential features with the help of LibROSA and have got the understanding about the features and their uses. The software platform we have used for coding is jupyter notebook. Indian classical raag YAMAN have very sharp characteristics and from this extracted information, further raag identification, raga recognition and raga classification can be done with excellent accuracy.

ACKNOWLEDGEMENT:

This paper and the research behind it would not have been possible without the exceptional support of my external guide as well as co-author, Prof. Kunjal I. Tandel. His enthusiasm, knowledge and exacting attention to detail have been an inspiration and kept my work on track from my first encounter with the Technical aspects of Music to the final draft of this paper.

Moreover, I also like to appreciate the efforts of my internal guide and co-author, Prof. Rahul M. Patel. Without their guidance I never have reached to this level. I also would like to thank my college, Dr. S & S S Gandhy government Engineering College and its staff and their technical support.

REFERENCES:

[1] Sharma, G., Umapathy, K., & Krishnan, S. (2020). Trends in audio signal feature extraction methods. Applied Acoustics, 158, 107020. doi:10.1016/j.apacoust.2019.107020

[2] Bhat AS, Amith VS, Prasad NS, Mohan DM. An efficient classification algorithm for music mood detection in western and hindi music using audio feature extraction. In: 2014 fifth international conference on signal and image processing. p. 359–64.

[3] Saunders J. Real-time discrimination of broadcast speech/music. 1996 IEEE international conference on acoustics, speech, and signal processing conference proceedings, vol. 2. IEEE; 1996. p. 993–6.

[4] Stevens KN. Autocorrelation analysis of speech sounds. J Acoust Soc Am 1950;22(6):769–71. https://doi.org/10.1121/1.1906687.

[5] Baniya BK, Lee J, Li ZN. Audio feature reduction and analysis for automatic music genre classification. In: 2014 IEEE international conference on systems, man, and cybernetics (SMC). IEEE; 2014. p. 457–62.

[6] Zhu Y, Kankanhalli MS. Precise pitch profile feature extraction from musical audio for key detection. IEEE Trans Multimedia 2006;8(3):575–84.

[7] Ghoraani B, Krishnan S. Time-frequency matrix feature extraction and classification of environmental audio signals. IEEE Trans Audio Speech Lang Process 2011;19(7):2197–209. https://doi.org/10.1109/TASL.2011.2118753.

[8] Umapathy K, Krishnan S, Rao RK. Audio signal feature extraction and classification using local discriminant bases. IEEE Trans Audio Speech Lang Process 2007;15(4):1236–46. https://doi.org/10.1109/TASL.2006.885921.

[9] Gong, Rong & Serra, Xavier. (2018). Towards an efficient deep learning model for musical onset detection.

[10] Adel, Salwa. (2015). Analyze Features Extraction for Audio Signal with Six Emotions Expressions.

[11] Shah, Ayush & Kattel, Manasi & Nepal, Araju & Shrestha, D. (2019). Chroma Feature Extraction.

[12] Müller, Meinard & Ewert, Sebastian. (2011). Chroma Toolbox: Matlab Implementations for Extracting Variants of Chroma-Based Audio Features.. 215-220.

[13] McFee, Brian & Raffel, Colin & Liang, Dawen & Ellis, Daniel & Mcvicar, Matt & Battenberg, Eric & Nieto, Oriol. (2015). librosa: Audio and Music Signal Analysis in Python. 18-24. 10.25080/Majora-7b98e3ed-003.

[14] Doerfler, Monika & Grill, Thomas. (2017). Inside the Spectrogram: Convolutional Neural Networks in Audio Processing.. 10.1109/SAMPTA.2017.8024472.

[15] Ando, Yoichi. (2013). Autocorrelation-Based Features for Speech Representation. The Journal of the Acoustical Society of America. 133. 3292. 10.1121/1.4805418.

[16] N. P. Patel and M. S. Patwardhan, Identification of Most contributing features for Audio Classification, International Conference on Cloud Ubiquitous Computing Emerging Technologies, 2013, pp. 219-223.

[17] J. S. Downie, "Music information retrieval," Annual review of information science and technology, vol. 37, no. 1, pp. 295–340, 2003.

[18] A.A.Bardekar and Ajay.A.Gurjar, "Study of Indian Classical Ragas Yaman and Todi Structure and its Emotional Influence on Human Body for Music Therapy".

[19] Mathur A, Vijayakumar SH, Chakrabarti B, Singh NC. Emotional responses to Hindustani raga music: the role of musical structure. Front Psychol. 2015;6:513. Published 2015 Apr 30. doi:10.3389/fpsyg.2015.00513