IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 08, Issue 6 (June. 2018), ||V (VII) || PP 23-26

Emotion Recognition in Vocal Music: A test with Hindi, Bengali and Odia

Sanghamitra Mohanty
(Department of Computer Science and Application,
Utkal University, Bhubaneswar – 751004, Odisha, India)
Corresponding Auth0r: Sanghamitra Mohanty

Abstract: Music signals are processed like speech signals in computers. Music also carries information about the emotions with which it is sung by the singer. To recognize the emotions in the piece of music be it accompanied with instruments or not is tried here. The Cochleogram Model is selected as the features of the music signals, through the robust Gammatone Frequency Cepstral Coefficients(GFCC). These coefficients are calculated using the ERB filters and the emotions in the music are recognized successfully. The optimizations technique followed here is the FFBPNN. The result is satisfactory tough provoking to new challenges.

Date of Submission: 15-06-2018 Date of acceptance: 30-06-2018

I. INTRODUCTION

Music is a form of art and it represents the cultural activity, whose medium is generally sound and silence and it exists in time. Music exists through the common elements like pitch, rhythm, dynamics and the sonic qualities. Vocal music is a type of music generally performed by one or more singers, with or without the use of instrumental accompaniment and in it, singing is the main focus of the speech. Vocal music represents the sung words called lyrics in general.

Emotion, in everyday speech, is any relatively brief conscious experience characterized by intense mental activity and a high degree of pleasure or displeasure. It is often interwined with mood, temperament, personality, disposition and motivation. Music being a form of speech, we can relate it with emotion as it is done for speech signals. Vocal music is still a potential phenomena, where emotion is attached with. Vocal music be it with or without accompaniments carries the speech signal part and this has the emotion merged in it. Emotion in music[1] involves psychological theory of emotions and are represented by the evidence obtained due to the introspections of musician, aestheticians, and listeners and the objective data gathered from the observation of the behavioral study of them. With respect to Music, Emotion can be defined in a different way as it generates a psychological state of mind to sing with a particular emotion, which is different from acting in an emotion and singing. According to Yand and Chen[2] broadly emotion with respect to Music can be stated in 28 different types[2], which is a challenge for the researchers of Emotion in Speech. They have made the classification with respect to Arousal and Valency. Below is the figure depicting 28 emotions in music.

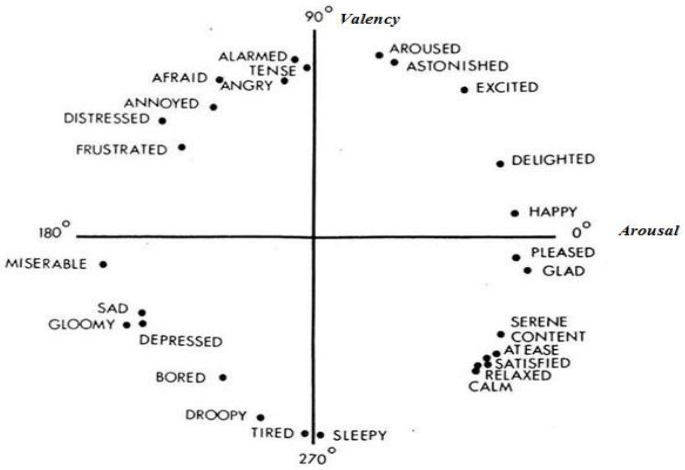

Fig. 3 28 types of emotion for Music according to Yand and Chen.

In this piece of work, an attempt has been made to recognize the emotion lying in a piece of Music be it with or without accompanying instruments. With respect to computer analysis Music is a speech signal when it is represented inside the computer digitally. The music database was made up of Hindi, Bengali and Odia Music. The result is quite satisfactory for the trained and test samples. Further studies are on considering features like arousal and valency with respect to the music signal to classify the emotions further.

II. MATHEMATICAL DERIVATIONS FOR GFCC

Here is an attempt to recognize the emotion in the vocal music applying the psychophysical observations of the auditory periphery and this filter-bank is a standard model of cochlear filtering. This auditory feature parametric analysis is the Gammatone Frequency Cepstral Coefficient (GFCC)[2] as it is modeled on the Cochleogram unlike Spectrogram. It is a Time Frame analysis involving a bank of Gammatone filters[3,4]. The impulse response of a Gammatone filter centered at frequency f is given by

$$g(t) = at^{(n-1)}\ e^{-2\pi bt}\cos(2\pi f_c t + \phi)\quad , t \geq 0 \qquad (1)$$

where f_c is the central frequency of the filter, and ϕ is the phase which is usually set to 0. Constant a controls the gain and n is the order of the filter which is usually set to be equal to or less than 4. And b is the decay factor which is related to f_c and is given by

$$b=1.019 * 24.7 * (4.37 * f_c /1000 + 1) \qquad (2)$$

where the central frequency of a filter or channel is a measure of a central frequency between the upper and lower cutoff frequencies. It is either the arithmetic mean or geometric mean of the lower cutoff frequency and upper cutoff frequency of a band-pass system or a band-stop system.

A set of Gammatone Filters (GF) with different f_c results in forming a Gammatone filter bank. GF being derived from measured impulse response, has complete amplitude and phase information. This GF is applied to obtain the music signal characteristics at different frequency and the resultant is the temporal frequency representation like FFT-based short time spectral analysis. When there is an effort to simulate the human auditory behavior of the signal, the central frequencies of the filterbank are equally distributed on the Bark Scale.

This Gammatone filters, which is otherwise known as Cochleogram is a time frequency representation of the input signal which is mimicking the components of a cochlea, the sensitive part of the human auditory system. Here the frequency is downsampled into frequency bands with Equivalent Rectangular Bandwidth (ERB)[5] scale, given by

$$ERB(f_c) = 6.23\ f_c^{\ 2} + 93.39.\ f_c + 28.52 \qquad (3)$$

This is due to Moore and Glassberg in 1983 who latter in 1990 gave another linear equation

$$ERB(f_c) = 24.7.(4.37.\ f_c\ + 1) \qquad (4)$$

Where f_c is in kHz and ERB(fc) is in Hz.

$ERB(f_c)$ is a measure used in psychoacoustics, which gives an approximation to the bandwidths of the filters in human hearing, using the unrealistic but convenient simplification of modeling the filters as rectangular band-pass filters. After the Gammatone Filter bank is defined, it is applied to the raw music signal to generate the respective cochleogram, which represents transformed raw music signal in the time and frequency domain. The advantage of using a cochleogram over spectrogram is that the features of a cochleogram is based on ERB scale with finer resolution at low frequency than the Mel-scale used in spectrogram. Besides it allows more number of coefficients in comparison to MFCC. The Mel filter-bank for a power spectrum is with 257 coefficients[6] while the GFCC is with 512 coefficients.

Experiment and Result:

In this experiment we have used vocal music of films and have tried to get the emotion through our programs. The experiment is done using 24 channels, frequency lowered upto 50Hz, summation window 0.025 secs., hop between successive windows is 0.01 secs.. The ERB filter is made first and then the Filter Bank. Appling FFT the Gammatone filters are generated. The coefficients thus generated are matched with the trained dataset using Feed Forward Back Propagation Neural Network method. First training of the database is done with epoch 5000. Then the test dataset are experimented with the epoch being dynamically decided by the program. It accommodates mono as well as stereo recordings and recognizes the emotion lying in it. Music without accompanying instruments are also tested and the results found are satisfactory.

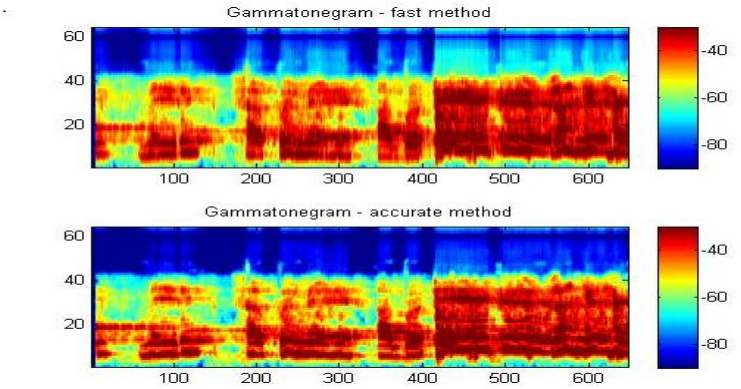

Fig. 1 Cochleogram for the vocal music tere_mono.wav

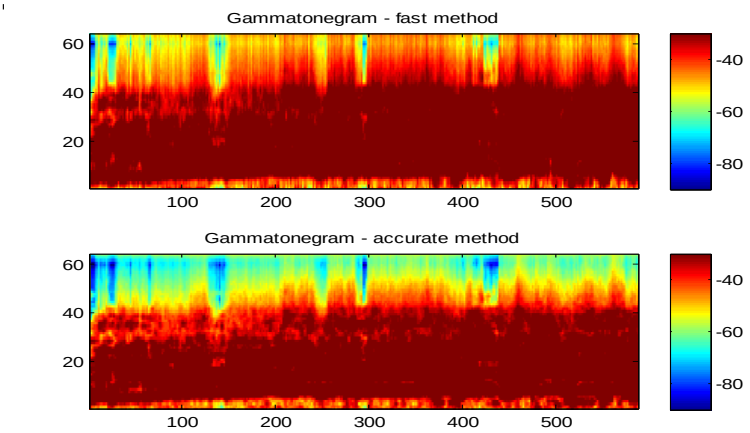

Fig.2 Cochleogram for the vocal music baharo_mono.wav

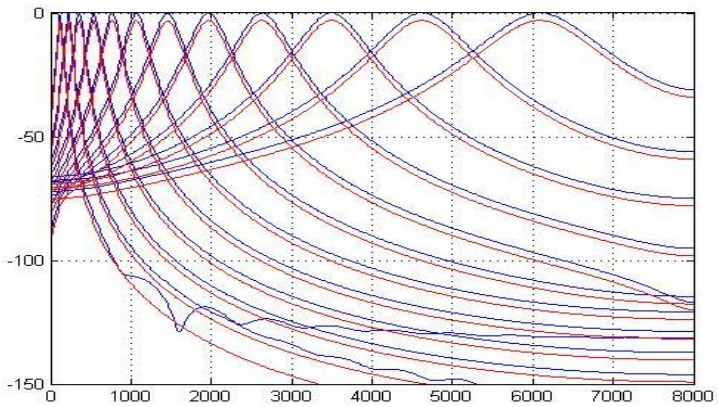

Fig. 3 Gammatone Frequency Cepstral Coefficients for the music for Ellis and Malcolm a comparison.

The GFCC coefficients are plotted using ERB Filters with respect to frequencies are shown below[8]. The same is checked with Malcolm's toolbox[9]. Fig. 3 shows the comparision of Gammatone Frequency Cepstral Coefficients for Ellis[8], and for Malcolm's [9] with a difference of 3 DB.

Fig. 1 and Fig 2. represent the Cochleogram of two music pieces while Fig.3 shows the GFCC. The experiment is setup with 7 layers of training network. The test is for two types of Emotions namely (1) Joy and Sadness during the experiment. In every case it runs till the SSE value saturates with value << 1.

III. CONCLUSION

The experiment is performed with MATLAB. The Feed Forward Back Propagation Neural Network optimization technique is used for the optimization of the results during emotion recognition. The database is with popular film music of male and female singers with different emotions. Ten singers in three different languages i.e. Hindi, Bengali and Odia are considered. The algorithm was trained by 80% of the speech database and 20% are kept for testing. Accuracy for Joy or happiness is found to be correct upto 90% while it has some confusion for the rest. A detailed study is on for other near emotions like joy with serene and sadness, as serene is a state with less arousal and less valence value but it is more towards joy.

REFERENCES

[1]. Meyer L. B. " Emotion and Meaning in Music", The University of Chicago Press, 1956.
[2]. Y.H. Yand and H. H. Chen, "Music Emotion Recognition", CRC Press, 2011.
[3]. Darling A. M. "Properties and Implementation of Gammatone Filter: A Tutorial", 1991.
[4]. Yang Shao, Z. Jin, D. Wang and S. Srinivasan, "An Auitory Based Features For Robust Speech Recognition", ICASSP 2009.
[5]. A. Tjandra, S. Sakhi, G. Neubig, T. Toda, M. Adrian and S. Nakamura, "Combination of Two-Dimensional Cochleogram and Spectrogram Features For Deep Learning-Based ASR", ICASSP 2015.
[6]. B. C. J. Moore and B. R. Glasberg, "Suggested formulae for calculating auditory-filter bandwidths and extraction patterns" Journal of Acoustical Society of America Vol. 74 1983.
[7]. Honig F. , Stemmer G. Hacker C. And Brugnara F. " Revising Perceptual Linear Prediction (PLP)", INTERSPEECH 2005.
[8]. D.P.W.Ellis(2009)."Gammatonelikespectrograms",webresource,http://www.ee.columbia.edu/~dpwe/resources/matlab/gammatonegram/.
[9]. Malcolm Slaney (1998) "Auditory Toolbox Version 2", Technical Report #1998-010, Interval Research Corporation, 1998, http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/.

IOSR Journal of Engineering (IOSRJEN) is UGC approved Journal with Sl. No. 3240, Journal no. 48995.

Sanghamitra Mohanty "Emotion Recognition in Vocal Music: a test with Hindi, Bengali and Odia "IOSR Journal of Engineering (IOSRJEN), vol. 08, no. 6, 2018, pp. 23-26