Speech, music and multifractality

Susmita Bhaduri[1,]* and Dipak Ghosh[1,2]

[1]Deepa Ghosh Research Foundation, Kolkata 700 031, India
[2]C. V. Raman Centre for Physics and Music, Jadavpur University, Kolkata 700 032, India

Audio signal categorization is one of the rudimentary steps in applications like content-based audio information retrieval, audio indexing, speaker identification, and so on. In this work, a rigorous, non-stationary methodology capable of categorization among speech and various music signals is proposed. Multifractal detrended fluctuation analysis method is used to analyse the internal dynamics of the acoustics of digitized audio signal. The test data include speech (non-musical), drone (periodically musical) and music samples of Rāgas (having different musicality) from Indian classical music (INDIC). It is found that the degree of complexity and multifractality (measured by width of the multifractal spectrum) changes from the start towards the end of each audio sample. However, the range of this variation is the smallest for speech and drone. The normalized value of the width of the multifractal spectrum is strikingly different for speech and drone. Experimental results show that this parameter can effectively classify speech and drone signals. Further, we have experimented with a number of clips of INDIC Rāgas with a range of variation in musicality and mood content. The results show that the width of the multifractal spectrum of the signals can categorize different music signals. In contrast with the conventional stationary techniques for audio signal analysis, we have used the method of complexity analysis without converting the non-stationary audio signals in frequency domain. We have used basic waveforms of the audio signals after de-noising them.

Keywords: Classical music, drone, multifractal analysis, speech.

MUSIC is conventionally defined as an ordered arrangement of sounds of different acoustic frequencies (pitches, tones) in succession (melody), of sounds in combination (harmony), and of sounds spaced in temporal succession (rhythm)[1]. However, which one of these is mainly responsible for the musicality of the signal, still remains unresolved. According to Mandelbrot[2], the quality of a certain kind of sound remains unaffected even with a change in playing speed, which he termed 'scaling noise'. White noise is the simplest scaling noise. The power spectral density $S(f)$ of a time series produced in accordance with the temporal variation of white noise varies with f-frequency content, according to the relation $S(f) \propto f^{\beta}$, where β is the scaling exponent. Brownian noise is another type of scaling noise with scaling exponent $\beta = 2$.

According to Mandelbrot[2], fractal is a geometric pattern which is iterated at smaller or larger scales to produce self-similar, irregular shapes or surfaces that cannot be represented using Euclidian geometry. Fractal systems are infinite, i.e. they can extend to insurmountably large values of their coordinates, outwards in all directions from the centre. Another important feature of fractals is their self-similarity, i.e. smaller and bigger fragments of a system look similar, but are not necessarily identical, to the entire fractal system. To express self-similarity of the large and small, power law [a mathematical pattern in which the frequency of occurrence of a given size is inversely proportionate to some power (n) of its size.] is applied. This n is defined as the scaling exponent or the fractal dimension of the system. Fractals can be of two types: monofractals and multifractals. Monofractals are those whose scaling properties are the same in different regions of the system. Multifractals are more complicated self-similar objects which consist of differently weighted fractals with different non-integer dimensions. Hence their scaling properties are different in different regions of the systems[3].

In nature, there exist many geometries which are fractals, like the profile of a mountain or shape of snowflakes. If we investigate them closely, we can deduce self-similarity of the system. Music was originated in the sounds that nature produces, and hence music also has a fractal property like many other naturally occurring fluctuations. Vincenzo Galilei was the first to analyse the numerology of music[4]. He pointed out that the octave can be obtained through different ratios of $2n : 1$. It is 2 : 1 in terms of string length, 4 : 1 in terms of weights attached to the strings, which are inversely related to the cross-section of the string, and 8 : 1 in terms of volume of sound-producing bodies, such as organ pipes. Studies have shown that an octave can be divided by the rule of equal temperament[1,5]. It is one of the most popular methods considered to play a musical composition harmonically in all keys. The first fractal analysis of music was carried out by Voss and Clarke[6]. They showed that it is pink noise or $1/f$ noise. Bak *et al.*[7] also showed that this type of noise occurs often in nature. Tricot[8] applied some fractal theories on self-affine functions, and found a power law relationship between power spectra and fractal dimension. Hsü and Hsü[1,9] analysed the variations in pitch interval between successive notes in a series of music scores composed by Bach and Mozart, and showed that the incidence frequency approximately exhibits a power-law relationship.

Detrended fluctuation analysis (DFA) is a scaling analysis method where the scaling exponent (similar to a single-scale Hurst exponent) is used to quantify the long-range correlation of stationary and non-stationary signals[10]. Shi[11] employed the calculation method of the

*For correspondence. (e-mail: susmita.sbhaduri@gmail.com)

Hurst exponent to examine the pitch sequence fashioned in folk songs and piano pieces. Gunduz and Gunduz[12] studied the mathematical structures of six songs by treating them as complex systems. From the above study, it can be summarized that apart from the frequency-domain stationary methods like Fourier power spectrum, techniques like DFA have also been used to compute the Hurst exponent and the value of fractal dimension of the non-stationary music signal. But fractal dimension refers to the overall properties of the song sequence. Recent research with complex systems showed that naturally evolving geometries and phenomena cannot be characterized by a single scaling ratio (as in monofractal system), as different parts of the system are scaled differently. Such a system is better characterized as a multifractal system[13,14]. Multifractal detrended fluctuation analysis (MFDFA) method has been applied successfully to study multifractal scaling behaviour of various non-stationary, scale-invariant time series[15]. MFDFA method is a robust tool for performing scaling analysis in case of nonlinear, non-stationary time series. Results obtained by this method turn out to be more reliable in comparison to methods like wavelet analysis, discrete wavelet transform, wavelet transform modulus maxima, detrending moving average, band moving average, modified detrended fluctuation analysis, etc.[16–18]. It has been applied for analysing various phenomena such as heart-rate dynamics, DNA sequences, neuron spiking, human gait and economic time series as well as weather-related and earthquake signals.

In recent years works have been reported, where non-stationarity and nonlinearity of the time series of music signals have been used to quantify the complexity/ musicality of the acoustic signal. As musicality in audio signal has naturally evolving geometry and non-uniform pattern in its progression, it is necessary to reinvestigate the musical structure from the viewpoint of the multifractal theory. Multifractal analysis of music has been carried out by some researchers[19,20]. Substituting both rhythm and melody by a geometrical sequence of points, Su and Wu[19] showed that these quantities can be considered as multifractal objects. Although some work (as described above) has been done for Western Music under non-stationary conditions, not much has been done for Indian music, specially Indian classical music (INDIC). Nowadays, automatic classification and retrieval of audio or music information has become a significant area of research. An effective music and audio classification system for INDIC built under non-stationary conditions, would be a step forward towards various applications like music or audio indexing, music information retrieval (MIR) and genre classification. Other applications include music learning in distant mode, performer recognition and music synthesis.

We have used MFDFA technique to measure temporal variation of self-similarity of the audio waveforms of various music and speech signals to reveal the internal dynamics of their so-called musicality. The width of the multifractal spectrum for the waveform of each audio sample is measured. First, the excerpts of speech and drone signals are used for the experiment. Drone signal has a perceived periodicity as its musical feature and thereby one can intuitively conclude that it is less multifractal in nature. On the contrary, speech is believed to have less or no musicality in its nature and one can conclude that its multifractality is much higher than that of a drone signal. Experimental results justify this fact. For them, multifractality remains almost consistent throughout the signal and the value of the spectrum width varies within a small range. Further, we have experimented with different kinds of Rāga in the INDIC domain and have compared the results among themselves and with the limiting value of drone. The result shows that the range of variation of the spectrum width of music signals is much higher and different compared to the drone. Using these experimental inerences, the future road-map for a computational system for categorization of speech, drone and music signals from INDIC can be initiated.

To study the musicality in INDIC, some details of the Rāga framework are analysed. It should be noted that acoustic signals like speech and music can be differentiated by their musicality content. A sequence of notes, the rhythm, the mood, the temporal and spatial variation of a note sequence, the time of compositions – all create a multidimensional piece, called a musical composition. Corresponding to the framework of Western music, Rāga is the soul of INDIC. According to some studies[21,22], Rāga, the nucleus of INDIC, may be defined as a melodic structure with fixed notes and a set of rules characterizing a certain mood conveyed by performance. Each Rāga expresses different moods in certain characteristic progressions. This Rāga framework contributes to the musicality of INDIC system and makes it distinctly different from the Western music system. So, Rāga greatly contributes towards fixing the degree of complexity of an audio signal for INDIC system. We can classify various compositions of INDIC by their Rāga-base.

The rest of the communication is organized as follows. First, the details of data are elaborated. The method of analysis and inferences from the test results are then presented followed by concluding remarks.

The experimental data are as follows:

- Speech signal which has less amount of musical content.
- Drone signal which contains more amount of periodically occurring musical content or sequence of notes according to INDIC system.
- Signals of instrumental recordings of four music compositions of different Rāga-s eliciting different moods or emotions. Samples 1, 2 and samples 3, 4 in the test dataset are almost opposite to each other in terms of

mood content of the Rāgas. Table 1 shows the basic mood details for each sample. We have taken the opening (rhythmless) section of each composition, because this part introduces and develops the melodic modes or Rāgas.

The speech, drone and rest of the music samples of 160 sec duration, are in .wav format. Sampling frequency for the data is 44.1 kHz. Audio samples are encoded by 16 bit-stream and are of single (mono) channel. The amplitude waveform is taken for testing. We have used empirical mode decomposition method of Norden *et al.*[23] for noise removal from the original signal.

Each 160 sec sample is divided among 80 samples of 2 sec from start to end. Then, multifractal spectra of the whole signal and the segments are generated. Widths of the spectrum are denoted as follows:

- W – Width of spectrum for the whole signal (160 sec).
- w-Width of spectrum for the segmented samples (2 sec).

These widths are calculated according to the method of Kantelhardt *et al.*[15]. The step-by-step process is described below. Software implementation is done in Matlab. Then the analysis of w and W is done for all the samples and inferences are drawn from it.

Step 1: Each digitized audio signal represents a time series having time instants in the x-axis and for each time instant a corresponding amplitude value in the y-axis. Suppose for a particular audio sample $i = 1, 2, \ldots, N$, are the time instants and corresponding amplitude value is $x(i)$. The mean of this time series is calculated as

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x(i).$$

Then the integrated series is computed according to eq. (1) of Kantelhardt *et al.*[15] as follows

$$Y(i) = \sum_{k=1}^{i}[x(k) - \bar{x}], \quad i = 1,2,\ldots,N.$$

Table 1. Test music data and their characteristics according to the Rāga framework

Music sample	Rāga	Mood content
Sample-1	Bahar	Sprightly feeling
Sample-2	Miyan-ki-malhar	Create tension and restlessness, anticipating separation and unnamed fears
Sample-3	Chhayanat	Happiness
Sample-4	Darbari kanada	Sadness, seriousness and pathos

Step 2: The integrated time series is divided into N_s non-overlapping bins (where $N_s = \text{int}(N/s)$, N is the length of the time series and s is the length of a single bin with respect to the number of time instants), and the fluctuation function is computed. In our experiment s varies from 16 as minimum to 1024 as maximum value in log-scale. For each s, the local RMS variation is calculated as function $F(s, v)$, according to eq. (2) of Kantelhardt *et al.*[15] as follows

$$F^2(s,v) \equiv \frac{1}{S}\sum_{i=1}^{S}\{Y[v-1)s+i] - y_v(i)\}^2,$$

where $i = 1, 2, \ldots, s$ and $v = 1, 2, \ldots, N_s$. Here $y_v(i)$ is the least square fitted polynomial of the bin v. It is defined as $y_v(i) = \sum_{k=0}^{m} C_k(i)^{m-1}$, where C_k is the kth coefficient of the fit polynomial with degree m. Here we have taken m as 1.

Step 3: The qth order overall RMS variation for each scale s is denoted by $F_q(s)$, which is calculated according to eq. (4) of Kantelhardt *et al.*[15] as shown below

$$F_q(s) \equiv \left\{\frac{1}{N_s}\sum_{v=1}^{N_s}[F^2(s,v)]^{\frac{q}{2}}\right\}^{\frac{1}{q}}.$$

For our experiment we have calculated qth order RMS variation $F_q(s)$ for 100 values of q ranging between (–5) and (+5).

Step 4: Steps 2 and 3 are repeated and $F_q(s)$ is calculated for various values of s. If the time series is long-range correlated, the $F_q(s)$ versus s for each q will show power law behaviour as $F_q(s) \propto s^{h(q)}$. When one quantity varies as the power of another, then the quantities are said to be showing power law behaviour. If such a scaling exists, $\log_2(F_q(s))$ will depend linearly on $\log_2(s)$, where $h(q)$ is the slope. The exponent $h(q)$ depends on q. Here $h(q)$ is the generalized Hurst exponent. This $h(q)$ of MFDFA is related to the scaling exponent $\tau(q)$ according to eq. (13) of Kantelhardt *et al.*[15], i.e.

$$\tau(q) = qh(q) - 1.$$

Step 5: Multifractal signals have multiple Hurst exponents. Hence $\tau(q)$ depends nonlinearly on q. If α is singularity strength, the singularity spectrum is $f(\alpha)$. This is related to $h(q)$ according to eq. (15) of Kantelhardt *et al.*[15], i.e.

$$\alpha = h(q) + qh'(q), \quad f(\alpha) = q[\alpha - h(q)] + 1.$$

The resulting multifractal spectrum $f(\alpha)$ is an arc as shown for a test sample in Figure 1. The difference between the maximum and minimum values of α, is called the multifractal spectrum width. The width of the spectrum gives a measure of the multifractality of the time series.

In our experiment, first the width of the multifractal spectrum for the whole signal is calculated for each input. Then for each input, the whole signal is shuffled and the width of the multifractal spectrum for that shuffled signal is calculated. If there are long-range correlations in the original data, they will be removed by this shuffling and the sequence will become uncorrelated. Hence the width of the multifractal spectrum for the shuffled signal will be much less and different from the width of the multifractal spectrum for the original signal. This was found to be true for all the test samples of speech, drone and music signals used in our experiment. Figure 1 shows the multifractal spectrum for one such original sample and its shuffled version. This test result clearly indicates that the multifractality in the speech, drone and music signals is due to their broad probability distribution and long-range correlation. The widths of the multifractal spectrum for the whole signal (W) and the width of the multifractal spectrum for the segments (w) of the same signal are calculated as per the above steps 1–5. The values of W and w are normalized within a range [0, 1].

Frequency histogram of w: For each of the segmented input signals w_is are calculated, where $i = 1, 2, \ldots, 80$. Then, the histogram for frequency of occurrence for a particular w_i throughout the progression of the signal is formed. In this histogram, the ranges of w_is are divided among a number of bins. Figure 2 shows the histogram for a particular musical sample. The peak of the histogram is denoted by w_p, which is considered here as the width of the multifractal spectrum for the segmented signal.

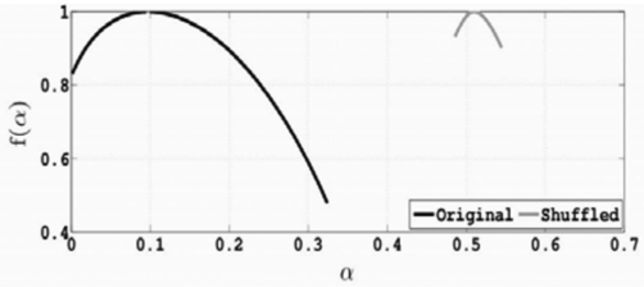

Figure 1. Plot of α versus $f(\alpha)$ for the original and shuffled signal data.

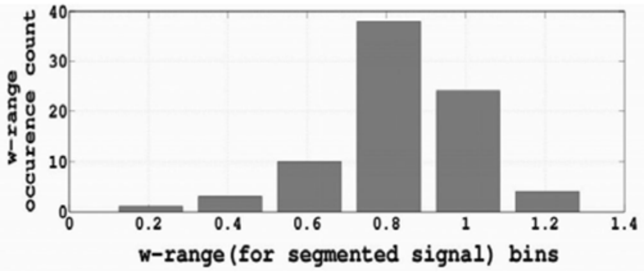

Figure 2. Frequency histogram of w_is for a particular music sample.

Trend of W and w for speech and drone signals: w_i-s, where $i = 1, 2, \ldots, 80$, are plotted from the start towards the end of the signal for speech and drone. What we get from the trend are as follows:

- Figure 3 shows that the respective w_i-s vary within the range $W \pm t$, where $t = 0.15$ for speech and $t = 0.12$ for drone signals. It is also evident from the figure that W for speech is much higher and different from the drone.
- Figure 3 also shows that for both speech and drone signals, the respective W and w_p are almost similar.
- Finally if we compare both W and w_p for speech and the drone (Figure 4), we can see that the parameters have a striking difference in values. Hence we can conclude that speech and drone signals can be distinguished by the width of their multifractal spectrum.

Trend of W and w for music and drone signals: As the difference between speech and drone is now been established, we attempt to compare the progression of w-s for each of the musical samples with that of drone signal, as drone is believed to be most musically periodic in nature. The inferences are as follows:

- Figure 5 shows that although the w_p-s for a particular music sample and drone are quite similar, their W-s are vastly different. This is because the w-s for the music signal change has a wider range of variation as compared to the drone signal. This results in a high value of W for music. So we can conclude that the multifractality of this particular music sample varies from very small to very high.
- We have done similar kind of comparison for the remaining three music samples with the drone and found a range of values for w compared to the drone signal. This range may be considered as a parameter for categorization of different music signals.

We have compared the overall W-s for the music samples with W of the drone sample. The inferences are as follows:

- Figure 6 shows that all W-s of the music samples are consistently higher than that of drone sample. It can be concluded that the drone signal contains the least amount of multifractality among all musical samples.
- Figure 6 also shows that W for sample-1 is much higher compared to sample-2. This is also true for the samples 3 and 4. Table 1 shows that samples 1, 2 and samples 3, 4 have almost opposing mood contents according to the Rāga framework. We may conclude

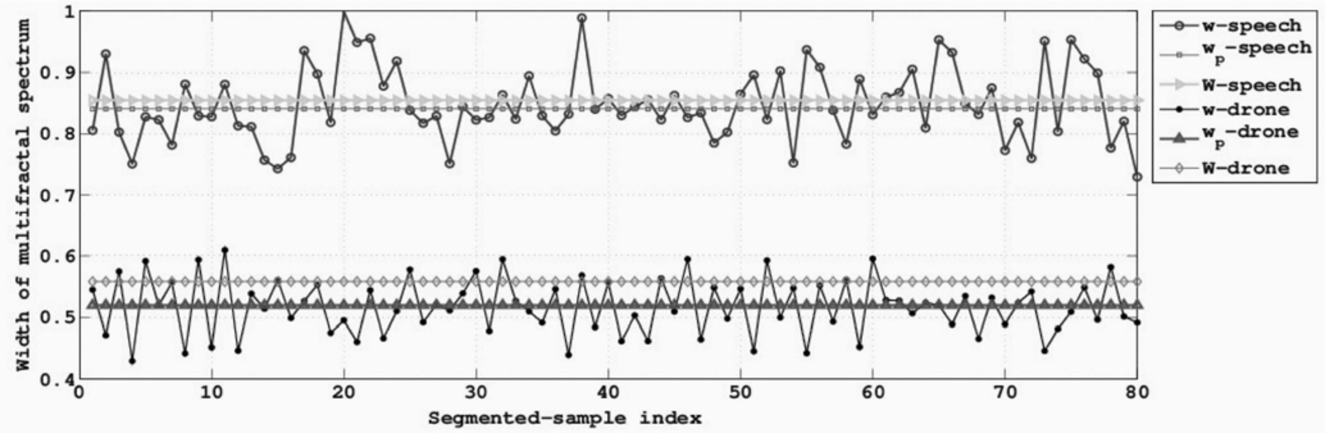

Figure 3. Trend of *W* and *w* for speech and drone signals.

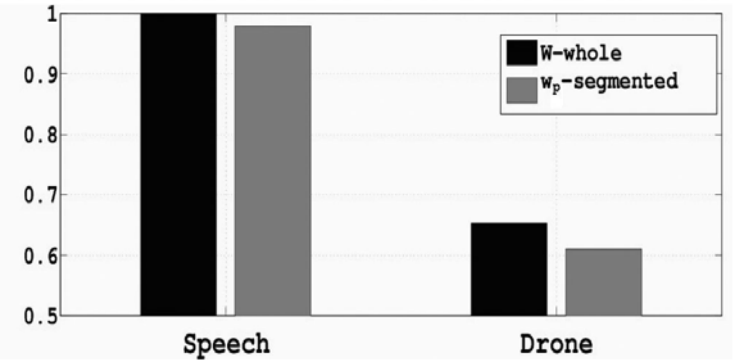

Figure 4. Comparison of *W* and w_p for speech and drone.

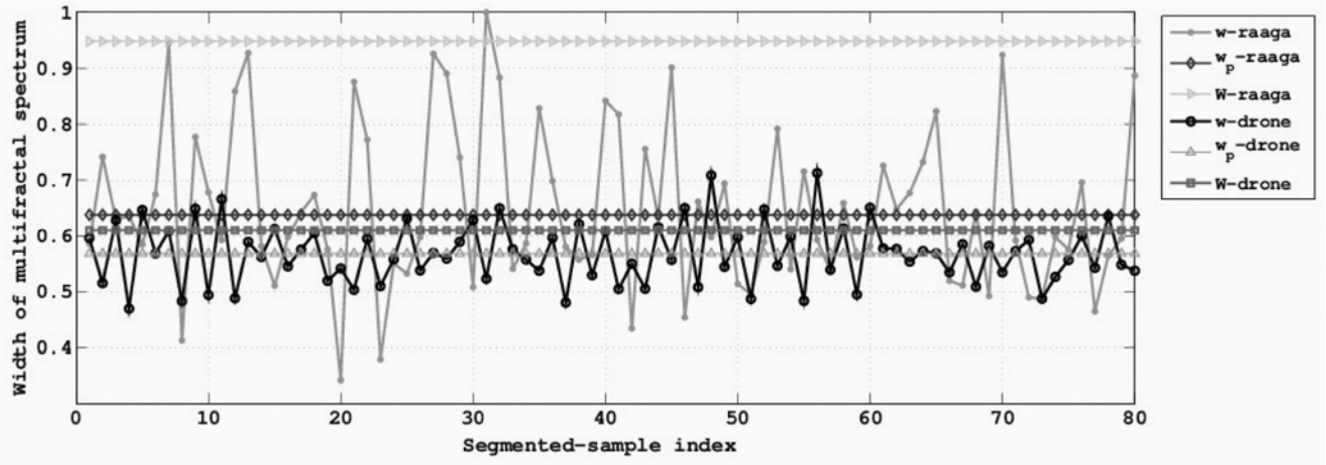

Figure 5. Trend of *W* and *w* for music sample and drone signal.

that this opposing mood content contributes to the difference in *W*-s of the respective samples.

We can infer here that degree of complexity increases with increasing musicality. Hence the width of the multifractal spectrum increases as we move from the drone towards music samples. Also, the degree of complexity varies largely for samples with mutually exclusive musicality and mood content. Hence with the help of this parameter we may try to classify music samples with different musicality or mood content as an extension of this work.

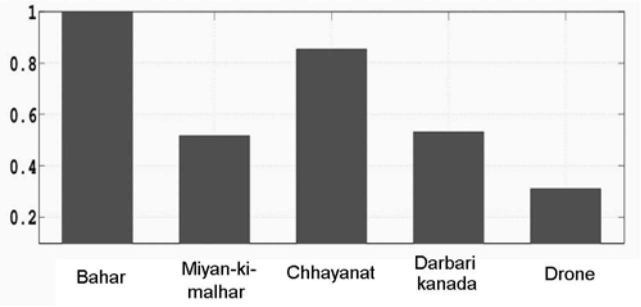

Figure 6. Trend of *W* for music sample and drone signal.

In this work we have taken the initial step towards devising a non-stationary computation model for audio signal classification. Using MFDFA method, the widths of multifractal spectra for the whole signal as well as segmented signal, have been calculated and analysed. These parameters make a clear distinction between speech and drone signals. The proposed parameters can also classify Rāga signals of the INDIC system, containing mutually exclusive musicality. The success in the experiment lends us to believe that we may hereafter be able to differentiate and alienate the musicality and moods in different Rāgas which are mutually inclusive in some aspects and yet different in nature. More rigorous analysis of large samples of INDIC and Western music data with different genres of music conveying a variety of emotions, needs to be done to frame a computational system for MIR. The methodology may be extended to develop a framework to classify different types of music signals according to INDIC and Western music systems.

1. Hsu, K. J. and Hsu, A. J., Fractal geometry of music. *Proc. Natl. Acad. Sci. USA*, 1990, **87**(3), 938–941.
2. Mandelbrot, B.. *The Fractal Geometry of Nature*, Henry Holt and Company, 1983, vol. 51(3), pp. 384–391.
3. Chen, Z., Ivanov, P. Ch., Hu, K. and Stanley, H. E., Effect of nonstationarities on detrended fluctuation analysis. *Phys. Rev. E.*, 2002, **65**(4), 041107–041111.
4. Haar, J. and Palisca, C. V., *Humanism in Italian Renaissance Musical Thought*, Renaiss, Q., 1988, vol. 41(1), pp. 138–156.
5. Madden, C. B., *Fractals in Music: Introductory Mathematics for Musical Analysis*, High Art Press, Salt Lake City, 1999.
6. Voss, R. F., 1/*f* noise in music: Music from 1/*f* noise. *J. Acoust. Soc. Am.*, 1978, **63**(1), 258–263.
7. Bak, P., Tang, C. and Wiesenfeld, K., Self-organized criticality: An explanation of the 1/*f* noise. *Phys. Rev. Lett.*, 1987, **59**(4), 381–384.
8. Tricot, C., Dimension fractale et spectre. *J. Chim. Phys.*, 1988, **85**(1), 379–382.
9. Hsü, K. J. and Hsü, A., Self-similarity of the 1/*f* noise called music. *Proc. Natl. Acad. Sci. USA*, 1991, **88**(8), 3507–3509.
10. Hausdorff, J. M., Purdon, P. L., Peng, C. K., Ladin, Z., Wei, J. W. and Goldberger, A. L., Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations. *J. Appl. Physiol.*, 1996, **80**(5), 1448–1457.
11. Shi, Y., Correlations of pitches in music. *Fractals*, 1996, **4**(4), 547–553.
12. Gunduz, G. and Gunduz, U., The mathematical analysis of the structure of some songs. *Physica A Stat. Mech.*, 2005, **357**(3–4), 565–592.
13. Buldyrev, S. V., Goldberger, A. L., Havlin, S., Peng, C. K., Stanley, H. E. and Stanley, M. H. R., Fractal landscapes and molecular evolution: analysis of myosin heavy chain genes. *Biophys. J.*, 1993, **65**(6), 2673–2679.
14. Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. H. A., Havlin, S. and Bunde, A., Detecting long-range correlations with detrended fluctuation analysis. *Physica A Stat. Mech.*, 2001, **295**(3–4), 441–454.
15. Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Bunde, A., Havlin, S. and Stanley, H. E., Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A Stat. Mech.*, 2002, **316**(1–4), 87–114.
16. Serranoa, E. and Figliola, A., Wavelet Leaders: a new method to estimate the multifractal singularity spectra. *Physica A Stat. Mech.*, 2009, **388**(14), 2793–2805.
17. Oswiecimka, P., Kwapien, J. and Drozdz, S., Wavelet versus detrended fluctuation analysis of multifractal structures. *Phys. Rev. E*, 2006, **74**(1), 016103–016109.
18. Huang, Y. X., Schmitt, F. G., Hermand, J. P. and Gagne, Y., Arbitrary-order Hilbert spectral analysis for time series possessing scaling statistics: a comparison study with detrended fluctuation analysis and wavelet leaders. *Phys. Rev. E*, 2011, **84**(1), 016208–016215.
19. Su, Z. Y. and Wu, T., Multifractal analyses of music sequences. *Physica D: Nonlinear Phenom.*, 2006, **221**(2), 188–194.
20. Jafari, G. R., Pedram, P. and Hedayatifar, L., Long-range correlation and multifractality in bach's inventions pitches. *J. Stat. Mech. Theory Exp.*, 2007, **4**, 04012–04019.
21. Chakraborty, S., Krishnapriya, K., Loveleen, Chauhan, S. and Solanki, S. S., Analyzing the melodic structure of a north indian raga: a statistical approach. *Electron. J. Musicol.*, 2009, **XII**.
22. Bhatkhande, V. N., *Hindustani Sangeet Paddhati Kramik Pustak Malika*, Sakhi Prakashan, New Delhi, 1990.
23. Norden, E. H. *et al.*, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A Math. Phys. Eng. Sci.*, 1998, **454**(1971), 903–995.

ACKNOWLEDGEMENT. We thank the Department of Higher Education, Government of West Bengal, for logistics support.

Received 29 August 2015; revised accepted 26 December 2015

doi: 10.18520/cs/v110/i9/1817-1822