An acoustical and psychological study on contribution of lyrics in raga-based happy and sad Indian music

A Banerjee1,2*, S Sanyal1,2,3, P Gayen4, S Roy2,5, P Patnaik1,4 and D Ghosh2

1 Rekhi Centre of Excellence for the Science of Happiness, IIT Kharagpur, India
2 Sir C.V. Raman Centre for Physics and Music, Jadavpur University, Kolkata, India
3 School of Languages and Linguistics, Jadavpur University, Kolkata, India
4 Department of Humanities and Social Sciences, IIT Kharagpur, India
5 Department of Physics, Jadavpur University, Kolkata, India

*Corresponding author, e-mail address: archibanerjee7@gmail.com

Abstract. A perfect complementary relationship between the lyric and the melody can give birth to a beautiful song. The melodic expression of a song is universal but the lyrical expression is not – lyric is culture specific because of its language dependence. Can melody itself communicate the core emotions of a song? Or does the addition of a lyrical sense significantly change its emotional experience? This study looks for the answers focusing on a unique subgenre of Indian Classical Music – Ragashroyi compositions, where the melodic movements sincerely follow the Raga pathways but the lyrics explore a much deeper and wider variety of emotions compared to Raga bandishes. Recordings were collected from two eminent vocalists (1 male, 1 female), each of whom was asked to sing (with proper lyrics) and hum (without meaningful lyrics) any two Bengali Ragashroyi compositions of two opposite emotions – happiness and sadness. Hurst Exponents, obtained from robust non-linear Detrended Fluctuation Analysis (DFA) of the recorded acoustic waveforms, were compared for each song-humming pair having same melodic structure to understand the acoustic contribution of the lyrics in a song quantitatively. A comparative audience response study was also conducted where several humming and song clips were played randomly and two groups (one which understands Bengali, the other which does not), each having 30 participants, were asked to mark the emotions and the characteristic features corresponding to each clip on 5 point Likert scale, and their responses were compared for each song-humming pair. This pilot study on Ragashroyi Indian music explores in depth the contribution of lyrics in vocal music from both the perspectives of computational acoustics and audience psychology.

Keywords: Singing & Humming, Indian Classical Music, Human response, Non-linear analysis, DFA

1. Introduction
From mimicking the sounds of nature to expressing the most profound realizations of this universe in songs – vocal music has evolved through millions of years and human voice was probably the first musical instrument. Songs are generally made up of two intrinsically connected parts: poetry (which is in the form of lyrics) and melody. A very good complementary relationship between the poetry/lyric and the melody can give birth to a beautiful song, capable of communicating the complete emotion of the composition to the audience. The proper fit between these parts seems to be made by acoustic features that encompass the relationship between them, representing two fields of sonic communication: musical and verbal communication [1]. While lyrics convey semantic meaning, melody enhances its emotional intention, filling informational gaps and enhancing the significance of poetic meaning of lyrics that otherwise would be incomplete. Within a song, the melodic expression is universal but the lyrical expression is not. Lyrics grow with help of language and as we move across the globe languages vary. Without proper understanding of a language it is impossible to understand...
the semantic sense of any lyrics/poetry written in it. The individual as well as combined roles of melody and lyrics in musical emotion expression and communication are greatly intriguing to the music researcher fraternity. Normally, the audience always listen to the lyrics and melody together in a song, but what happens when the melody is hummed independently without the lyrics? Can the emotions of a song get communicated to the audience who do not understand the lyrical meaning? Or does the addition of a lyrical sense significantly change the complete emotional experience of the song? We try to look for the answers to these long withstanding questions in this present study from different perspectives.

1.1. Lyrics and melody - from Acoustical, Behavioural and Neuro-cognitive perspectives

A number of behavioural and neural studies have been done to foster information about the cognitive processing of these two components. Evidence from behavioural and electrophysiological studies [2, 3] indicates that lyrics and melody are processed separately while detecting verbal and melodic errors in song. Also, lesion studies indicate dissociation between these two components demonstrated by better performance on production of melody than on production of lyrics in non-fluent aphasic patients [4] and by preservation of the ability of recognizing lyrics but not of melody in a music-agnostic patient [5]. On the other hand, several behavioural and neuro-physiological studies have reported that neural pathways involved in processing of melody and lyrics interact at some stage of processing [6-8]. Saito et.al [9] found that during listening to songs the left posterior inferior temporal cortex is specifically activated, and this brain region may serve as an interface between verbal and musical representations in order to facilitate song recognition. Brattico et al [10] stressed on the role of acoustic cues for the experience of happiness in music and to the importance of lyrics for sad musical emotions. In this work, the main aim is to look for acoustic and behavioural cues which endeavour to analyze the contribution of lyrics in vocal music using Ragashroyi compositions – a unique genre of Indian classical music.

Music has been characterized as a language of emotions [11] as listeners are able to recognize a different basic emotions expressed by music, particularly happiness and sadness [12-20]. Many studies have already reported consistently that happy music is characterized by fast tempo and major mode, whereas sad music is typically played in slow tempo and minor mode [21-23]. In case of Indian classical music (ICM), however, the situation is somewhat different. The central building blocks in this system of music are called Ragas, each raga consisting of a sequence of five or more notes and having a structural guideline about the possible transitions between these notes. But Raga (derived from “Ranj”) means that which delights peoples’ hearts (through the soundscape) and each raga can evoke some specific cluster of subdued to profound emotions. Within the structural framework of a raga there is scope for infinite improvisations for each artist and the raga grammars are “only means and not end in themselves” [24]. A bandish is a song composed in a specific raga and follows the raga structural guidelines very strictly. Bandishes also have recently strong lexical contents. A number of studies in the acoustical, behavioural and neural level [13, 19, 20, 25-28] have investigated the emotions expressed through different parts of a raga performance and the level of arousal corresponding to each. We humans are the only species able to perform temporal synchronization to an external pulse [1]. This ability seems to be part of our nature, instead of being nurtured (i.e. developed). Acoustic feature analysis by Brattico et al [10] showed that music with lyrics differed from music without lyrics in spectral centroid, a feature related to perceptual brightness, whereas sad music with lyrics did not diverge from happy music without lyrics, indicating the role of other factors in emotion classification and behavioural ratings revealed that happy music without lyrics induced stronger positive emotions than happy music with lyrics.

1.2. Nonlinearity in music and its application to study Ragashroyi songs

The word Ragashroyi literally means that which takes shelter (Ashroy) under a particular raga. Melodic movements in a Ragashroyi composition/song largely follow the pathways of the Raga it is based on but, in contrast to Raga bandishes, these songs explore much deeper and wider variety of
emotions lyrically. Naturally, in order to give proper importance to the lyrical expression an artist sometimes takes the liberty of flouting the strict raga grammar and can mix two or more ragas to paint the appropriate colour of the song. For this study, recordings were collected from two professional vocal artists (a male and a female vocalist), each of whom was asked to respectively sing (with proper meaningful lyrics) and hum (without using any meaningful lyrics) two ragashroyi compositions, which will primarily communicate two opposite emotions – happiness and sadness. Music has a very complex structure, more so for vocal music as at every instant different musical components like pitch, intensity, intonation, rhythm are very closely linked to each other and words have their own rhythm, pause and break. Hence, it is obvious that the acoustic waveform will change considerably with the addition of lyrics. These complex properties of the music time series are peculiar of systems having a chaotic, self-similar and nonlinear nature [19, 20, 24, 29]. So, analyzing music from only linear and deterministic approaches forces us to lose critical information and modern nonlinear tools of physics are needed to get in depth information about how the introduction of lyrics affects the acoustical waveform of a melodic structure within a song. Fractal analysis is one such efficient tool which can look into the embedded geometry of the complex time series with its mathematical microscope and can quantify the scale of self similarity present in the signal. It is important to note that in this study, we consider both lyrics and melody as parts of the complete nonlinear temporal acoustic waveform of the song, and try to observe the changes from the hummed version of the exactly same melodic content in the acoustic domain. Later on these acoustical changes are expected to trigger different emotional variations among the various groups of listeners. With this approach, we look to study acoustic features from utterance of lyrics that present spectral envelope similar to the features from their corresponding melody, thus pointing to which aspects of lyric prosody influences the corresponding song melody. All the acoustic waveforms of the recorded music signals were cut into smaller pieces according to different paragraphs of the songs and these parts were analysed using a robust latest state of the art non-linear technique called Detrended Fluctuation Analysis (DFA) [30] which essentially computes the Long range temporal correlations (LRTC) present in the audio signals and the geometry embedded in signal is quantified by a scaling exponent (or Hurst Exponent). Later, to understand the acoustic contribution of the lyrics in a Ragashroyi song from a quantitative approach the scaling exponent values were compared for each pair of song-humming versions of the same melodic structure i.e., sung with/without meaningful words. Another human response study was conducted on two groups of audience (each group had 30 participants) one of which understood Bengali, the other group did not. Both the humming and the song versions of the same melodic parts, taken from the two ragashroyi songs sung by both the male and female singer, were played in a randomly shuffled order and the two groups of participants were asked to mark the emotions and the characteristic features using a 5 point Likert scale. The responses obtained from both the groups for each song-humming pair of same melodic content was compared to study the impact of meaningful lyrics on the emotional expressivity of the song.

2. Experimental details

2.1. Choice of audio signals

For this experiment music signals were recorded in the same recording studio setup from one male (a legendary maestro of ICM with experience of performing over 55 years) and one female (who is learning and have experience of performing over 15 years) native Bengali speaking vocalists. Both of them were asked to consecutively hum (without words or meaningful lyrics) and sing (with proper meaningful lyrics) two Bengali ragashroyi compositions which will primarily communicate two opposite emotions – happiness and sadness. The two songs chosen by them were – (i) Nishitho shoyone jage ankhi (based on Raga Chandrakaushiki) which is primarily expected to evoke sad emotions with the lyrics explaining the frustrations of a dejected person in a sleepless pensive night (tempo ~ 45 bpm) and (ii) Holi kheliche shyam kunjokanone (based on Raga Pilu) which is expected to evoke happy and romantic emotions where the lyrics depicts the longing between two lovers and the
optimistic possibility of their meeting during the popular festival of colour Holi (tempo ~ 60 bpm). For experimental purposes, both of the singers sung all the song and humming clips in the same pitch (B) or tonic Sa in context of Indian music. Including the song and humming versions a total of 8 audio clips were recorded from both the singers. Each song/humming clip had duration of about 60 seconds.

2.2. Details of audience response participants
An online audience response study was conducted on total 60 respondents (M=32, F=28), out of which 30 participants (first group) were native Bengali speakers and the other 30 participants (second group) were Non-Bengali speaking people from across the world who also do not understand spoken Bengali language. The participants were from different educational background and the age of the participants varied from 17 to 69 years (Mean 29.82 years, SD 10.79 years). None of the participants reported any auditory impairment or neurological and psychological disorders. Most of the participants in the psychological test had no formal musical training in Indian music.

3. Methodology
3.1. Methodology for Acoustic analysis
All the recorded music signals were normalized to 0dB. Each of these sound signals was digitized at the sample rate of 44.1 KHZ, 16 bit resolution and in a mono channel. According to the different stanza of the lyrics each song or humming clip was divided into two parts – sthayi (first stanza) and antara (second stanza) and we got total 16 or 8 pairs of experimental audio clips, where each pair was consisting of song (lyrical) and humming (non-lyrical) versions of the same melodic content. The tempo was exactly same within each pair, thus the song-humming clips within each pair had same data length. DFA technique was developed for quantifying long range temporal correlation properties in non-stationary signals (e.g., in physiological time series and many other natural processes), because long-range correlations can also come from the artifacts of the time series data [30]. In our study, each of the 16 experimental audio clips was divided into 10 equal parts using MATLAB software and DFA technique has been applied to quantify the symmetry scaling behaviour of the fluctuations in each of the post processed audio signal parts and the scaling exponents were compared for each pair of song-humming (or lyrical-non lyrical) versions of the same melodic content.

3.1.1. Detrended Fluctuation Analysis
To compute the Hurst exponent $H$ of a time series: $x_1, x_2, \ldots, x_N$ using Detrended Fluctuation Analysis or DFA technique (first conceived by Peng et al [30]), firstly integration of $x$ is done to form a new series $y = y_1, \ldots, y_N$, where $\bar{x}$ is the mean of $x_1, x_2, \ldots, x_N$.

$$y(k) = \sum_{i=1}^{k} (x_i - \bar{x})$$

The integrated series is then sliced into boxes of equal length or intervals of size $n$. In each box of length $n$, a least-squares line is fit to the data. The least square line represents the trend in that box. The coordinates of the straight line segments are denoted by $y_{n}(k)$. The root-mean-square fluctuation of the integrated series is calculated by

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (y(k) - y_{n}(k))^2}$$

where the part $y(k) - y_{n}(k)$ is called detrending. The relationship between the detrended series and interval lengths can be expressed as

$$F(n) \propto n^\alpha$$
where $\alpha$ is expressed as the slope of a double logarithmic plot $\log F(n)$ versus $\log (n)$ (as shown in representative Figure 1 (a,b)). The parameter $\alpha$ (scaling exponent, autocorrelation exponent, self-similarity parameter etc.) represents the auto-correlation properties of the signal.

![Figure 1a](image1.png)  
**Figure 1a:** DFA scaling exponent from double logarithmic plots for a) humming and b) song versions of the same melodic part.

We have applied the DFA technique following the NBT algorithm used in [31]. The scaling exponent provides a quantitative measure of long range temporal correlation (LRTC) that exists in the audio signals. When the auditory waveform is completely uncorrelated (Gaussian or non-Gaussian probability distribution), the scaling exponent values measure around 0.5, this is also called white noise. The value of scaling exponent lies between 0 to 0.5 for anti-correlated sequences, from 0.5 to 1 for long range temporal correlations and $\alpha > 1$ for strong temporal correlations that are not of a power law form. When computing the scaling of the signal profile, the resulting scaling exponent, $\alpha$, is an estimation of Hurst exponent $H$. If $\alpha$ is between 0 and 1, then $x$ was produced by a stationary process which can be modeled as an fGn process with $H = \alpha$. If $\alpha$ is between 1 and 2 then $x$ was produced by a non-stationary process, and $H = \alpha - 1$ [32].

3.2. Methodology for Audience response analysis

The two groups of respondents (one which understands Bengali, one which does not) were asked to mark the emotional intensity corresponding to each song and humming clip in a Likert psychological scale of 1 to 5, where 1 = Very Low, 2 = Low, 3 = Moderate, 4 = High and 5 = Very High. A sample questionnaire which the participants of the survey had to fill is given in **Figure 2**.

![Figure 2](image2.png)  
**Figure 2:** Sample questionnaire of survey for recording the emotional response of the audience corresponding to each song/ humming audio clip.

We have chosen 11 discrete emotions (Happy, Sad, Anger, Fear, Surprise, Disgust, Calm, Tense, Romantic, Heroic, Exciting) for this study [33] to cover the wide range of evoked musical emotions,
as much as possible, in the context of Indian music. For all the given audio clips the participants were instructed to mark only for those emotions, which they found appropriate for the song/humming audio clip they are listening to and leave the rest empty. All of the recorded song and humming clips (in both male and female voices) were arranged in a shuffled order, but it was consciously taken care of that the humming version of a particular melodic structure always appeared before the song version of the same in the sequence, so that the Bengali speaking audience did not get biased by the lyrical content while responding to the humming version. The entire survey was conducted in online mode and the respondents participated voluntarily in the study. The results obtained from the human response survey have been analyzed with rigorous statistical methods to obtain new interesting results on the perceptual processing of lyrics and melody in Ragashroyi Bengali songs among Bengali and non-Bengali speaking population.

4. Results and discussion

4.1. Results of acoustic analysis
Analyzing the recorded audio clips using the software Wavesurfer [34] and Praat [35] we attempted to compare the pitch contours and the intensity contours for each pair of song-humming clips having the same melodic construct. One pair of such sample pitch contours for the humming and song versions of same melody are given in Figure 3 (a,b).

![Figure 3 (a,b): Pitch contours and acoustic waveforms of the (a) humming and (b) song versions of the same melodic structure.](image)

The comparative study clearly shows that the pitch contours in both cases are almost identical to each other, but in case of humming the consistency of the steady note pitch contours are more prominent compared to the song version of the same melody. It is evident that this little discontinuities or small fluctuations in case of song version can be attributed to the presence of various consonants and the sudden and spontaneous pitch fluctuations related to them.

![Figure 4: Intensity contours of the humming (red lines) and the song (green lines) versions of the same melodic structure.](image)
Another sample plot featuring the comparative intensity variations in the humming and song versions of the same melodic construct is given in Figure 4. From the comparative intensity variations between the humming and song versions of the same melodic formation we can clearly see that the song version features significantly greater amount of fluctuation in the intensity contour compared to that of the corresponding humming version. In case of humming (red lines), each melodic phrase in the audio clip is sung at a go without much fluctuations in the intensity and the intensity contour undergoes a deep only during the pauses in the audio clip. On the contrary, the sound of the lyric dominate over the sound of the melody in the song version, and we observe sudden intensity fluctuations especially in the consonant regions of each word and during the pauses embedded between two consecutive words in the lyrics. From the pitch and intensity contour analysis we can safely say that the basic difference between singing and humming is caused mainly by the presence of rhythmic content embedded in the lyrics, as well as the pitch and amplitude modulation associated with it which can directly contribute to the change in the order of power law temporal correlations present in the auditory waveforms. To get the complete picture nonlinear DFA technique was applied to all the audio clips and the average DFA scaling exponent values (averaged over 10 parts of the each audio clip) were calculated for each song and humming clip. Finally, the change in symmetry scaling nature was observed for each song-humming pair having the same melodic structure for both the male (Figure 5a) and the female (Figure 5b) singer.

Figure 5 (a,b): Variation in DFA scaling exponent values for different humming-song pairs for two ragashroyi songs sung by the (a) male and (b) female vocalist.

Comparing the above figures, we can clearly observe that

i. For both of the songs of opposite emotions (sad/nostalgic – happy/romantic) within each song-humming pair (divided according to the sthayi and antara, i.e., the different stanza of the lyrics), the DFA scaling exponent value corresponding to the acoustical waveform of the humming version feature higher value compared to the song version of the same melodic content. This trend is consistent for the male (Figure 5a) and the female (Figure 5b) vocalist. For both the singers, the maximum decrement is observed from the humming version of the Holi kheliche sthayi part to the song version of the same melodic content. In other words, it is observed that the long range temporal correlation in the auditory waveform of any song depends both on the lyrical and the melodic construct and the temporal correlation is higher in the humming version compared to the song version of the same melodic structure in all song-humming pairs chosen for this experiment. This observation can be attributed to the dominant presence of lyrics in Ragashroyi songs along with the well constructed melodic modes of the Raga it is based on. Though the song Nishitho shoyone jage ankhith which is expected to evoke sad/nostalgic emotion among audience is sung at a relatively lower tempo (~45 beats per minute) compared to the song Holi kheliche shyam (tempo ~ 60 bpm) which is expected to evoke happy/romantic emotions, from the overall perspective of world music it can be safely
said that both of them tend to the slower tempo range. The scope of projecting the lyrics over
the melodic structure in the slower tempo range is always higher than in case of higher tempo
songs. Moreover, the colors of the Raga in a Ragashroyi song, if matches with the essence of
the lyrics, provides infinite scope of elaboration and intricacies to the artist. Following the
Raga pathways or phrases the artist can improvise and create a number of new alleys to express
and communicate the intended meaning of the lyrics. This projection of lyrics in the song
versions naturally incorporate the sudden and spontaneous fluctuation of rhythm, intensity and
pitch related to the lyrical aspect, which are not seen in the humming version of the same
melodic content and these fluctuations in musical perceptual parameters are the reason behind
the decrement of the long range temporal correlation from the humming to song version within
each song-humming pair.

ii. Another important observation is that for both the male and the female singers, both the
humming and song versions in different parts of the acoustical waveform of the sad song
Nishitho shoyone feature overall higher values of DFA scaling exponent compared to the happy
song Holi kheliche. This can be attributed to the melodic structure of both of the songs.
Following the nature of the Raga Chandrakaushiki and the melodic progression of the
composition Nishitho shoyone which based on the same raga, it is observed that there is
abundance of minor notes as well as continuous note to note transitions like meend (continuous
upwards or downwards gliding between two notes, especially two distant notes) and andolon
(continuous oscillations between two notes) which are associated to sad, tense emotions. Fast
movements and jumping fragmented note to note transitions are almost absent in this
composition. On the contrary, the melodic construct of the composition Holi kheliche shyam
based on Raga Pilu is dominated by short segments of taans (fast moving sequence of 4-5 or
more notes) as well as frequently occurring jumping note to note transitions and features wider
use of major notes compared to minor notes. This song also features a strong rhythmic structure
in the lyrical aspect which matches with that of the melodic counterpart. So, apart from the
small difference in tempo, the two contrast emotion evoking Ragashroyi Bengali songs largely
differ in the acoustical constitutes of the melody as well as lyrics. This observation hints
towards a possibility that the presence of slow moving, continuous and gliding transitions can
be associated with higher long range temporal correlation compared to faster movements and
jumping note to note transitions.

iii. Both for the male and female singer it is observed that overall, the song and humming versions
of the antara (second stanza) feature lower values of DFA scaling exponent compared to those
of the sthayi (first stanza) both for the two contrast emotion evoking Ragashroyi Bengali songs.
Normally, in the context of Indian music, it is very commonly observed that the melodic
constructs of the antara parts blossom more around the higher frequency ranges compared to
the sthayi part of the same song, which in turn can attribute to the lowering of long range
temporal correlation in the antara parts compared to the sthayi parts in both of the contrast
emotion evoking songs chosen for this experiment.

iv. Lastly, we can observe that the absolute DFA scaling exponent values yield relatively lower
values in case of the male artist compared to the female artist for both of the songs. Here, it
must be noted that the tonic Sa or the reference pitch were same (B) for all the humming and
song audio signals recorded from both the male and the female artist. So, the natural timbre
difference between these two voices is probably the main reason behind the difference in the
symmetry scaling nature of the audio signals generated from these two different sources.

4.2. Results of audience response analysis
Using the same song and humming audio clips we conducted an audience response survey with two
groups, each having 30 participants to determine the possible emotional changes associated with the
introduction of lyrics in these two chosen contrast emotion evoking Ragashroyi Bengali songs. The
analysis of the audience response obtained from the two groups (Participants of one group understand
spoken Bengali, but the other group does not) show prominent difference in the emotional responses for different parts of both the songs. In the audience response survey we used song and humming clips from the two songs both in male and female voice, but during analysis we observed that the responses of the Bengali as well as Non-Bengali audience were very similar for the male and female voice. So, we decided to merge the audience responses corresponding to the same humming and song parts sung by the two artists. During analysis of the emotional responses two major parameters were measured: (i) percentage of association of a particular emotion with a particular audio clip and (ii) corresponding weighted average of rating. Percentage of association means out of total 30 participants in each group how many participants associated the particular emotion to a particular experimental song and humming clip and as the participants rated their subjective arousal for that particular emotion corresponding to that particular clip in a scale of 1 (very low) to 5 (very high), the weighted average of rating measurement gives us the average emotional arousal evoked by the audio signal for that particular audience group. Table 1 and Table 2 report the overall percentage of association for each of the chosen 11 discrete emotions corresponding to each of the chosen experimental song and humming clips among the Bengali and Non-Bengali audience group respectively.

Table 1: Percentage of association of chosen 11 emotions with each song and humming clip among the Bengali audience group.

| Bengali audience | Nishitho shoyone jage | Holi kheliche shyam |
|------------------|-----------------------|---------------------|
|                  | Humming | Song | Humming | Antara | Song | Humming | Antara | Song | Humming | Antara | Song |
| Happy            | 73.33   | 63.33 | 80.00   | 63.33  | 93.33 | 80.00   | 90.00   | 83.33 |
| Sad              | 83.33   | 86.67 | 83.33   | 90.00  | 66.67 | 70.00   | 66.67   | 76.67 |
| Anger            | 60.00   | 60.00 | 66.67   | 56.67  | 60.00 | 63.33   | 63.33   | 60.00 |
| Fear             | 60.00   | 66.67 | 70.00   | 73.33  | 56.67 | 63.33   | 63.33   | 66.67 |
| Surprise         | 63.33   | 66.67 | 70.00   | 63.33  | 66.67 | 73.33   | 70.00   | 73.33 |
| Disgust          | 60.00   | 60.00 | 66.67   | 66.67  | 60.00 | 63.33   | 63.33   | 60.00 |
| Calm             | 90.00   | 80.00 | 76.67   | 80.00  | 80.00 | 70.00   | 73.33   | 70.00 |
| Tense            | 63.33   | 76.67 | 70.00   | 83.33  | 63.33 | 70.00   | 66.67   | 70.00 |
| Romantic         | 76.67   | 73.33 | 80.00   | 73.33  | 76.67 | 90.00   | 86.67   | 90.00 |
| Heroic           | 60.00   | 60.00 | 73.33   | 63.33  | 63.33 | 70.00   | 70.00   | 76.67 |
| Exciting         | 63.33   | 60.00 | 73.33   | 63.33  | 80.00 | 80.00   | 83.33   | 76.67 |

Table 2: Percentage of association of chosen 11 emotions with each song and humming clip among the Non-Bengali audience group.

| Non-Bengali audience | Nishitho shoyone jage | Holi kheliche shyam |
|----------------------|-----------------------|---------------------|
|                      | Humming | Song | Humming | Antara | Song | Humming | Antara | Song | Humming | Antara | Song |
| Happy                | 80.00   | 76.67 | 73.33   | 76.67  | 90.00 | 83.33   | 90.00   | 90.00 |
| Sad                  | 90.00   | 86.67 | 83.33   | 80.00  | 73.33 | 76.67   | 76.67   | 80.00 |
| Anger                | 76.67   | 73.33 | 76.67   | 73.33  | 70.00 | 73.33   | 73.33   | 73.33 |
| Fear                 | 83.33   | 80.00 | 73.33   | 70.00  | 73.33 | 73.33   | 73.33   | 73.33 |
| Surprise             | 76.67   | 70.00 | 76.67   | 73.33  | 70.00 | 73.33   | 76.67   | 80.00 |
| Disgust              | 76.67   | 73.33 | 80.00   | 76.67  | 70.00 | 73.33   | 73.33   | 73.33 |
| Calm                 | 100.00  | 93.33 | 80.00   | 83.33  | 80.00 | 83.33   | 83.33   | 90.00 |
| Tense                | 76.67   | 76.67 | 86.67   | 86.67  | 70.00 | 76.67   | 73.33   | 70.00 |
| Romantic             | 83.33   | 76.67 | 76.67   | 83.33  | 86.67 | 90.00   | 80.00   | 83.33 |
| Heroic               | 76.67   | 73.33 | 73.33   | 73.33  | 63.33 | 73.33   | 76.67   | 73.33 |
| Exciting             | 80.00   | 73.33 | 76.67   | 76.67  | 83.33 | 83.33   | 83.33   | 80.00 |

Table 1 clearly revealed that for Bengali audience group (i.e., who could understand the lyrical meaning of the two Ragachroyi songs) the emotional response significantly varied from the humming to the song version of the same melodic content in both of the songs. In the sthayi part of the song Nishitho shoyone, most (90%) Bengali participants associated the humming version with
the emotion calm, followed by sadness (83.33%), whereas after introduction of the lyrics, the same Bengali participants associated the song version of the same melodic structure primarily with sadness (86.67%), followed by calmness (80%). So, a clear switching happened between calmness and sadness (in terms of percentage of association) as the impact of lyrical addition. Though the humming version of the antara part of Nishitho shoyone primarily evoked sadness (83.33%), it was closely followed by happiness and romantic emotions; whereas the lyrical version of the same melody unanimously (90%) communicated sadness and also evoked tense emotion among large part (83.33%) of the Bengali audience. In both the sthayi and antara parts of the song Holi kheliche shyam, we find a prominent switching between the happy and romantic emotion for the Bengalis after the introduction of lyrics. The humming version of the sthayi part was primarily associated with happiness (93.33%), followed by calm and exciting emotions; but the song version of the same primarily contributed to evocation of romantic emotion (90%) among Bengali audience, followed by happy and exciting emotions. More prominent switchover occurred in the antara part of the same song where the Bengali audience mostly associated happiness (90%) with the humming version followed by romantic (86.67%), but felt romantic emotion (90%) more commonly than happiness (83.33%) after listening to the lyrics of the same melodic structure. From this result we can infer that the lyrical meaning has a major share in creating the entire emotional experience of a song, though the melodic component is almost an equally important underlying factor to complement and support the lyrical emotional expression. Analysis of Table 2 revealed that this sort of switchovers in the emotional association within a song-humming pair is hardly observed among the Non-Bengali audience group. Given the fact that the lyrical meaning of the songs will not be perceivable to this group of audience, this observation is just the scientific validation of the contribution of lyrics within a song. The Non-Bengali audience mostly associated calmness with both the humming (100%) and song (93.33%) versions of the sthayi part in the song Nishitho shoyone, whereas tense emotion was most commonly associated with both the humming (86.67%) and the song (86.67%) version of the antara part of the same song. In case of the sthayi part of the song Holi kheliche shyam only, similar to the Bengali audience response, a switchover between happiness and romantic emotions is observed for the non-Bengali audience too. Again in the antara part of the same song, this switchover is missing as both the humming and song versions are most commonly (90%) associated with happy and calm emotions. These observations indicate that the non-Bengali audience mostly relied on the melodic structure of the audio clip along with the sound of the lyrics (not the meaning) to mark the appropriate emotional response.

Now to get an idea about the level of arousal of each emotion corresponding to each song-humming pair, the weighted averages of the emotional responses of these two groups of audience are plotted in radar plots for each experimental audio clip. Figure 6(a-d) features the same for different parts of the song Nishitho shoyone and Figure 7(a-d) shows the variation in the weighted average of the emotional rating for the song and humming versions of the different parts of the song Holi kheliche shyam for these two groups of audiences.

Analysis of Figure 6a revealed that for the sthayi part of the song Nishitho shoyone, a prominent dip occurred in the average happiness arousal level from the humming to the song version among the Bengali audience, also a slight hike in the tension level is observed from humming to song here. Figure 6b revealed a more prominent hike in the average arousal level of both sad and tense emotion from the humming to song version of the Nishitho shoyone sthayi part in the emotional response from the Non-Bengalis, though unlike the Bengali community here the deep in the happiness arousal level is missing after the introduction of lyrics. In the Nishitho shoyone antara part, the response of Bengali audience (Figure 6c) again featured significant dip in the average arousal level of positive emotions like happy, exciting, romantic, surprise after the introduction of lyrics from the humming to song version, while a prominent increment in the average sadness arousal is observed. For the same humming-song pair, these sort of changes are absent in the response of the Non-Bengali audience (Figure 6d), rather a slight increment in the romantic
emotion arousal level is observed from the humming to song version in contrast to the emotional dimension of the lyrical meaning.

**Figure 6a**

**Figure 6b**

**Figure 6c**

**Figure 6d**

**Figure 6(a-d):** Variation in the weighted average of 11 emotions in different parts of *Nishitho shoyone*

**Figure 7a**

**Figure 7b**

**Figure 7c**

**Figure 7d**

**Figure 7(a-d):** Variation in the weighted average of 11 emotions in different parts of *Holi kheliche.*
For the sthayi part of the song Holi kheliche, the response of the Bengali audience (Figure 7a) revealed a sharp increment in the romantic emotion arousal level from the humming to the song version which can be directly attributed to the semantic contribution of the lyrics. In the Non-Bengali audience response corresponding to the same humming-song pair (Figure 7b), no such dominance is found for romantic emotion, rather increments are observed in the average arousal levels of happy and exciting emotions from the humming to song version, which can be attributed to the acoustic or phonetic contribution of the lyrics. In the Holi kheliche antara part, the emotional response of Bengali audience (Figure 7c) again featured a significant hike in the average arousal level of romantic emotion from humming to song, just like the sthayi part of the same song. No such changes are observed in the response of the Non-Bengali audience for the same song-humming pair (Figure 7d) rather in this case, the emotional arousal levels of all the 11 emotions are almost identical for both the song and humming versions of this melodic construct.

Table 3a: ANOVA results comparing emotional response data of Bengali audience corresponding to each song-humming pair from the two Ragashroyi songs.

| Bengali respondents |       |       |       |       |       |
|---------------------|-------|-------|-------|-------|-------|
|                     | DF    | Sum of Squares | Mean Square | F Value | Prob>F |
| Nishitho shoyone sthayi humming vs song | Model 6 | 101.48 | 16.91 | 54.12 | 8.74E-04 |
|                     | Error 4 | 1.25 | 0.31 |       |       |
|                     | Total 10 | 102.73 |       |       |       |
| Nishitho shoyone antara humming vs song | Model 6 | 10.23 | 2.56 |       |       |
|                     | Error 4 | 4.50 | 0.75 |       |       |
|                     | Total 10 | 14.73 |       |       |       |
| Holi kheliche sthayi humming vs song | Model 6 | 56.36 | 18.79 | 15.72 | 0.0017 |
|                     | Error 4 | 8.37 | 1.20 |       |       |
|                     | Total 10 | 25.7 |       |       |       |
| Holi kheliche antara humming vs song | Error 4 | 21.31 | 5.33 |       |       |
|                     | Total 10 | 26.73 |       |       |       |

Table 3b: ANOVA results comparing emotional response data of Non-Bengali audience corresponding to each song-humming pair from the two Ragashroyi songs.

| Non Bengali respondents |       |       |       |       |       |
|-------------------------|-------|-------|-------|-------|-------|
|                        | DF    | Sum of Squares | Mean Square | F Value | Prob>F |
| Nishitho shoyone sthayi humming vs song | Model 6 | 45.23 | 11.31 | 19.38 | 0.0014 |
|                        | Error 4 | 3.50 | 0.58 |       |       |
|                        | Total 10 | 48.73 |       |       |       |
| Nishitho shoyone antara humming vs song | Model 6 | 19.66 | 3.28 | 1.42 | 0.3837 |
|                        | Error 4 | 9.25 | 2.31 |       |       |
|                        | Total 10 | 28.91 |       |       |       |
| Holi kheliche sthayi humming vs song | Model 6 | 89.24 | 22.31 | 5.22 | 0.0371 |
|                        | Error 4 | 25.67 | 4.28 |       |       |
|                        | Total 10 | 114.91 |       |       |       |
| Holi kheliche antara humming vs song | Model 6 | 70.18 | 11.70 | 2.92 | 0.1592 |
|                        | Error 4 | 16.00 | 4.00 |       |       |
|                        | Total 10 | 86.18 |       |       |       |

A one-way ANOVA test was performed based on two different variable parameters to test the statistical significance of the results obtained from the human response analysis for both the groups of Bengali and Non-Bengali participants. Here the emotional response data from song-humming pairs have been considered as test groups and the one way ANOVA was performed keeping the Bengali and non Bengali speaking participant as constant factor. The ANOVA results for the Bengali and Non-Bengali participants are reported in Table 3a and Table 3b respectively. Comparing the results from Table 3a and Table 3b, we can clearly see that in case of the Bengali
respondents, strong statistical significance (i.e. p<0.05) is observed for most of the test groups at 95% confidence interval. This implies that in most cases the variance of the emotional response data in the two test groups are considerably high to arrive at a statistically significant conclusion. In this case, the variance between the two test groups is statistically significant for the song-humming pair of the Nishitho shoyone sthayi part, as well as for both the sthayi and antara parts of the song Holi kheliche shyam. The difference was not statistically significant only for the song-humming pair of the Nishitho shoyone antara part. In case of non-Bengali respondents, however, for the song-humming pairs from the antara parts of both Nishitho shoyone and Holi kheliche shyam, the statistical variance is not considerable enough to arrive at a conclusive inference. For the other two song-humming pairs corresponding to the sthayi part of two songs, statistically significant variances are observed.

5. Conclusions
This first-of-its-kind study attempted to analyze the acoustical as well as psychological contribution of lyrics in a unique genre of Indian music – Ragashroyi Bengali songs. This study presents unique data regarding quantification of acoustical contribution of lyrics and melody in two contrast emotion (sad/nostalgic and happy/romantic) evoking songs of this genre of music as well as their emotional impacts on the two groups of audience – one who understands the Bengali lyrical meaning and one who does not understand the same. For this purpose recordings were collected from one male and one female professional experienced artists who consecutively hummed (without any words) and sung (with proper lyrics) the two songs keeping the tempo and the melodic structure exactly same in the humming and song versions of each song.

Analysis of pitch and intensity contours of all the recorded signals revealed that though the overall pitch contours remain almost identical for the song and humming versions of the same melodic structure, sudden discontinuities and fluctuations related to the lyrical rhythm, utterance and pauses are present both in the pitch and intensity contours of the song versions, which are not found in the humming versions of the same melodic content. The main findings of the acoustic analysis using latest state of the art nonlinear Detrended Fluctuation Analysis (DFA) technique can be summarized as

i. In all parts of the chosen two contrast emotion evoking Ragashroyi Bengali songs sung by both the male and female singers, the humming version featured higher value of DFA scaling exponent or higher long range temporal correlation compared to the song version of the same melodic form. From this we can infer that the sudden and spontaneous pitch, intensity and rhythm fluctuations associated to the lexical content influence the long range temporal correlation present within the audio signal in a negative way.

ii. The different musical elements also have a great contribution in determining the amount of long range temporal correlations present in the acoustical waveform of the complete melodic structure. That is why the sad emotion evoking song with continuous, gliding transitions featured higher DFA scaling exponent values in all parts compared to the happy emotion evoking song with faster moving notes and jumping transitions between different notes.

Thus, we infer that the contributions of both lyrics and melody are equally important in determining the scale of self similarity present in the acoustical waveform of a Ragashroyi Bengali song. On the other hand, while studying the impacts of these acoustical differences (between the song and humming audio signals) on the human psychology, the main findings of the audience response analysis can be summarized as

i. In the context of Ragashroyi Bengali songs, the semantic contribution of the lyrics is highly significant in determining the complete emotional expression of a song. This is evident from the prominent switchovers found in the percentage of association of a particular emotion for Bengali audience group within each song-humming pair. On the contrary, as the lyrical meaning is indecipherable to the Non-Bengali group of participants, these switchovers are also missing in their emotional response for almost all song-humming pairs from both the sad and happy song.
ii. The lyrical contribution can also impact the emotion arousal level of a song. A resonance between the emotional expressivity of the melodic and lexical constructs of a song will always cause greater emotional arousal for the song version than the humming version of the same melody. Besides the semantic aspect of the lyrics, the acoustical cues of the lyricalics also impact the emotional expression of a song. This is evident from the changes in the arousal levels of different emotions for some of the song-humming pairs in case of the Non-Bengali audience. This is a pilot study on this unique genre of Indian music which endeavours to analyze the contributions of lyrics in vocal music from both the perspectives of computational acoustics and human psychology.

Acknowledgements:
The first author Archi Banerjee acknowledges the Department of Science and Technology (DST), Govt. of India for providing (SR/CSRI/PDF-34/2018) the DST CSRI Post Doctoral Fellowship to pursue this research work. Shankha Sanyal acknowledges DST CSRI, Govt of India for providing the funds related to this Major Research Project (DST/CSRI/2018/78 (G)) and the Acoustical Society of America (ASA) for providing the International Students Grant. Pinaki Gayen acknowledges the Ministry of Education, Govt. of India for providing the RGN Fellowship (F1-17.1/2013-14/RGNF-2013-14-SC-WES-45199) for pursuing this research.

References
[1] Fornari J 2013 Studying the intervenience of lyrics prosody in songs melodies. NICS Reports 3 20-26
[2] Bonnel AM, Faita F, Peretz I, Besson M 2001 Divided attention between lyrics and tunes of operatic songs: evidence for independent processing Perception Psychophysics 63 1201–13
[3] Besson M, Faita F, Peretz I, Bonnel A-M, Requin J 1998 Singing in the brain: Independence of Lyrics and Tunes Psychological Science 9 494–98
[4] Hebert S, Racette A, Gagnon L, Peretz I 2003 Revisiting the dissociation between singing and speaking in expressive aphasia Brain 126 1838–50
[5] Hebert S, Peretz I 2001 Are text and tune of familiar songs separable by brain damage? Brain Cognition 46 169–75
[6] Gordon RL, Schön D, Magne C, Astesano C, Besson M 2010 Words and melody are intertwined in perception of sung words: EEG and behavioral evidence PLoS One 5 e9889
[7] Koelsch S 2005 Neural substrates of processing syntax and semantics in music Current Opinion in Neurobiology 15 207–12
[8] Poulin-Charronnat B, Bigand E, Madurell P, Peereman R 2005 Musical structure modulates semantic priming in vocal music Cognition 94 B67–78
[9] Saito Y, Ishii K, Sakuma N, Kawasaki K, Oda K, & Mizusawa H 2012 Neural substrates for semantic memory of familiar songs: is there an interface between lyrics and melodies? PLoS One 7(9) e46354
[10] Brattico E, Alluri V, Bogert B, Jacobsen T, Vartiainen N, Nieminen S K, & Tervaniemi M 2011 A functional MRI study of happy and sad emotions in music with and without lyrics Frontiers in psychology 2 308
[11] Áhlberg L O 1994 Susan Langer on representation and emotion in music Br. J. Aesthet. b 69–80
[12] Altenmüller E, Schurmann K, Lim V K, and Parlitz D 2002 Hits to the left, flops to the right: different emotions during listening to music are reflected in cortical lateralisation patterns Neuropsychologia 40 2242–56
[13] Balkwill L L, and Thompson W 1999 A cross-cultural investigation of emotion in music: psychophysical and cultural cues Music Percept. Interdisciplinary J. 17 43–64. doi: 10.2307/40285811
[14] Peretz I, Gagnon L, and Bouchard B 1998 Music and emotion: perceptual determinants, immediacy, and isolation after brain damage Cognition 68 111–41
[15] Justlin P N, and Västfjäll D 2008 Emotional responses to music: the need to consider underlying mechanisms Behav. Brain Sci. 31 559–621
[16] Justlin P N and Laukka P 2004 Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening J. New Music Res. 33 217–238
[17] Khalfa S, Peretz I, Blondin J P and Robert M 2002 Event-related skin conductance responses to musical emotions in humans Neurosci. Lett. 328 145–49
[18] Krumhansl C L 1997 An exploratory study of musical emotions and psychophysiology Can. J. Exp. Psychol. 51 336–53
[19] Banerjee A, Sanyal S, Patranabis A, Banerjee K, Guhathakurta T, Sengupta R, Ghose D and Ghose P 2016 Study on brain dynamics by non linear analysis of music induced EEG signals Physica A: Statistical Mechanics and its Applications 444 110-20
[20] Ghose D, Sengupta R, Sanyal S and Banerjee A 2018 Musicality of human brain through fractal analytics Springer Singapore
[21] Peretz I, Gagnon L and Bouchard B 1998 Music and emotion: perceptual determinants, immediacy, and isolation after brain damage Cognition 68 111–41
[22] Dalla Bella S, Peretz I, Rousseau L and Gosselin N 2001 A developmental study of the affective value of tempo and mode in music Cognition 80 B1–B10
[23] Pallesen K J, Brattico E, Bailey C, Korvenoja A, Koivisto J, Gjedde A and Carlson S 2005 Emotion processing of major, minor, and dissonant chords: a functional magnetic resonance imaging study Ann. N. Y. Acad. Sci. 1060 450–53
[24] Sanyal S, Banerjee A, Patranabis A, Banerjee K, Sengupta R, & Ghosh, D 2016 A study on Improvisation in a Musical performance using Multifractal Detrended Cross Correlation Analysis Physica A: Statistical Mechanics and its Applications 462 67-83
[25] Chordia P, Godfrey M and Rae A 2008 Extending content-based recommendation: the case of Indian classical music Proc. of Int. Conf. on Music Information Retrieval (Atlanta, GA) 571–76
[26] Wieczorkowska A A, Datta A K, Sengupta R, Dey N and Mukherjee B 2010 On search for emotion in Hindusthani vocal music Advances in Music Information Retrieval, eds Z. W. Ras and A. Wieczorkowska (Berlin; Heidelberg: Springer) 285–304
[27] Mathur A, Vijayakumar S H, Chakrabarti B and Singh N C 2015 Emotional responses to Hindustani raga music: the role of structure Front. Psychol. 6 513. doi: 10.3389/fpsyg.2015.00513
[28] Sanyal S, Nag S, Banerjee A, Sengupta R and Ghosh D 2019 Music of brain and music on brain: a novel EEG sonification approach Cognitive neuromodulation 13(1) 13-31
[29] Sanyal S, Banerjee A, Nag S, Sarkar U, Roy S, Sengupta R and Ghosh D 2020 Tagore and neuroscience: A non-linear multifractal study to encapsulate the evolution of Tagore songs over a century Entertainment Computing 100367
[30] Peng C K, Buldyrev S V, Havlin S, Simons M, Stanley H E and Goldberger A L 1994 Mosaic organization of DNA nucleotides Phys. review E 49(2) 1685
[31] Hardstone R, Poil S S, Schiavone G, Jansen R, Nikulin V V, Mansvelder H D and Linkenkaer-Hansen K 2012 Detrended fluctuation analysis: a scale-free view on neuronal oscillations Frontiers in physiology 3 450
[32] Jose A R, Rodriguez E and Echerverria J C 2009 A DFA approach for assessing asymmetric correlations Physica A: Statistical Mechanics and its Applications 388(12) 2263-70
[33] Mauss I B and Robinson M D 2009 Measures of emotion: A review Cognition and emotion 23(2) 209-37
[34] Sjölander K and Jonas B 2000 Wavesurfer-an open source speech tool Sixth International Conference on Spoken Language Processing
[35] Boersma P and Van Heuven V 2001 Speak and unSpeak with PRAAT Glot International 5(9/10) 341-47