ninād

ninād

Contents

EDITORIAL

The word *music* comes from the Greek word *mousikê,* meaning any of the arts or sciences governed by the Muses. Of the nine Muses, the daughters of the God Zeus, only Euterpe, Terpsichore and Polyhymnia can be related to what we broadly understand as music. In India, *Sangeet* leads to the word Samgeet which means singing together. It is said that the Hindu Rishi Narada used the Veena as an accompaniment. In the Vedas, music began as a choral chant while older communities each had their own music. If one looks at the whole panorama of music in the world the solo performance would constitute only a miniscule portion of it. A significantly larger portion is occupied by collective performances.

For a prospective future one must look back at the past, into the origin of music. Some relate the origin with the theory of intentionality of music. It is said that the ability to reflect about the past and the future is a necessary ingredient of intentionality and the age when humans manifested this ability is between 60,000 to 30,000 years ago when they started creating art forms.

Those who look at the origin of music as an evolutionary necessity suggest at least four primary needs e.g. survival, territorial, sexual and motherese or mother-infant talk. This last one may push back the origin of music to the lullaby of Neanderthal mothers, about 300,000 years ago.

Though these parameters are no mean food nourishment for pundits and scholars, as of now these evolutionary necessities are extinct. Now music has only one common primary need, **entertainment** (no aspersion intended for music therapy). This entertainment is inclusive of peace and happiness of mind. This need is very culture specific. In this era of globalization, every culture is under a considerable strain to survive. In the national context, India being a country with extensive diversity, every community has its own culture and own music. All of them, starting from *Bhadur Gaan* from a Bengal village to Classical Music in India, are fighting to survive.

If we take a cue for survival strategy from the theory of evolution of life, every culture has to adapt to the changing environment, change brought about through interaction with other powerful cultures. These intercourses need not be seen negatively with suspicion. In fact, this is the only way to evolve. The cultures that do not interact with others are in great danger of extinction (the Jarawa community in the Andamans is a classic example). So is with the cultures which shut their eyes blindly to modernity, the main pillars of which is science and technology. Tradition too is an evolutionary phenomenon. For evolving, not only for survival, it is very necessary to analyse the environment transparently without bias and make necessary course correction. Modern civilization provides a plethora of knowledge tools in disciplines from psychology to information technology. They need to be used appropriately. Ninăd would endeavour to remain as a print as well as web based peer reviewed journal for this in the area of music.

Prof. Ashok Kumar Datta
Dr. Ranjan Sengupta
Joint Editors

OBJECTIVE ANALYSIS OF THE TIMBRAL QUALITY OF SITARS HAVING STRUCTURAL CHANGE OVER TIME

Anirban Patranabis, Kaushik Banerjee, Ranjan Sengupta
and Dipak Ghosh
Sir C V Raman Centre for Physics and Music,
Jadavpur University, Kolkata 700032

Abstract

Sitar is one of the long lasted classical musical instruments in India. From its evolution, through the ages, its structure has been changing continuously by means of development in shape and sound. All these were done by the then extraordinary talented craftsmen with the imaginative and emotional attachment of legendary music maestros. The primary objective for the structural modification of sitar were improvement of sound quality and easy to play all the required idioms and express best all the emotions of that related musician. Since the introduction of microphone to its growing sophistication, a number of structural modifications occurred in sitar to produce suitable sound as per musician's perception and changing taste of audience. In this paper we emphasize on the variation in the acoustical features occurred due to the structural changes of sitar from early to contemporary recordings. Timbre and Spectral features are the key to explain scientifically the changing profile of sitar which was based on perception of the instrument makers and the players during 20th century.

Key words: Sitar, origin, need for structural changes, timbre changes, acoustical features, relationship between structure and timbre

Introduction

The present form of Sitar (a plucked string instrument in Hindustani music) what we see today is the product of a long journey from Tritantrivina to Kachchapivina or Kurmivina to Kachua-sitar. [1] This transformation probably happened since Mughal Reign [2]; though B.C. Deva said that it was indigenous and modified from a fretted lute depicted in the sculpture of 12th century [3]. There is a lot of controversy about the inventor of sitar and when it was exactly named. One view is that it was a Persian instrument called Sehetar originated by Amir Khushro of 13th century [4]; while the other opinion is it was existed in India much earlier than Khushro. The origin and naming of sitar is perhaps less important here than its structural modifications and its outcomes to the acoustical perception. From the detail description C.R. Day [5] and Sir S.M.

Tagore [1] we come to know that, upto the end of 19th century Tritantrivina, Kachapi (tortoise) vina or Kachua sitar along Tarafder sitar were existed. There was no standard size and shape of sitar until early of 20th century. And It was Kanailal & Brother, who first standardized sitar both in size and Shape; but as in those days microphone was not a part and parcel in most of the concerts, so, both the structure and the tuning of sitar was much different from today's body of sitar used to be very thin, and bridge was very oval so that by a single puff it started buzzing [8]. With the increased use of microphones and its gradual development, a number of structural modifications occurred in sitar to produce suitable sound as per musician's perception and changing taste of audience. The genesis of new sitar tone was introduced by Noderchand Mullick and Hiren Roy thereby revolutionizing the sitar making industry. Noderchand used to make only Ravi Shankar's sitar and he introduced hooks on sitar frets so as to prevent noise at the time of Jhala and at the same time it will produce cord effect; made Dandi and Tabli thicker than Kanailal's sitar and slightly closed the joari [9]. Hiren Roy made 8 stringed sitar of Nikhil Banerjee which was both Kharaj-Pancham and Gandhar-Pancham (later discarded) and its Dandi and Tabli were very thick for heavy strokes; closed the Joari (Banda) to produce grave sound [10]. The new trend of sitar tone which was pioneered by these craftsmen with the imaginative help from the legendary sitar players are still continuing by today's makers, only deer horn has been replaced by teflon or fibre and the wooden pegs by machine keys.

The present study is directed to relate the structural development of sitar by analyzing the acoustical features extracted from early to contemporary recordings. Timbre and Spectral features might be helpful to explain scientifically the changing profile of sitar which was based on perception of the instrument makers and the players during 20th century.

Timbre Parameters: Timbre is defined in ASA (1960) as that quality which distinguishes two sounds with the same pitch, loudness and duration [13]. This definition defines what timbre is not, not what timbre is. It seems to be a multi-dimensional quality. Research has shown that timbre includes the irregularity of the amplitude of the partials. Other perceptive attributes of timbre are brightness, tristiumulus1, tristiumulus2, tristiumulus3, inharmonicity and odd/even relation of the partials. Spectral irregularity/spectral smoothness basically show the irregularity of a signal where the local mean is compared with the current amplitude value. Smoothness of a spectrum is an indicator fozr partials belonging to a same sound source and a single higher intensity partial is more likely to be perceived as an independent sound. It is also useful in revealing complex resonant structures of string instruments [11]. An alternative version of the conventional spectral irregularity algorithm, called Spectral Smoothness where the power of the spectrum is highlighted by the nonlinear

square operator [13]. Brightness [13] is correlated with the subjective quality of perception and closely related to sharpness. The tristimulus [13] is used to analyze the transient behavior of musical sounds. Other uses of the tristimulus include the classification and the analysis of source spectrum of musical instruments. Tristimulus measures the energy in the fundamental (Tristimulus 1), the first three harmonics (Tristimulus 2), and the higher harmonics (Tristimulus 3) in relation to the whole energy. Odd/Even relations are the measure for the energy distribution on even and odd harmonics and are related to the subjective sensation of fullness of a sound [12]. Inharmonicity [12] is an attribute to characterize pitched sounds with partial frequencies deviating harmonic frequencies. Those are also described as quasi-harmonic, which implies that the partial frequencies can be either stretched, or compressed. Inharmonicity is an attribute of stiff strings. This is visualized best by dividing the frequency of the overtones by their partial tone number resulting in a straight line for perfectly harmonic sounds and in a curve for quasi harmonic sounds. In the present paper all these parameters have been studied and correlation with the structure of the instruments has been attempted.

Experimental Details

Six samples of sitar played by different maestro at different period (1905-1986) were collected from archive of ITC-SRA, Jadavpur University and personal collections. Each of these sound signals was digitized with sample rate of 44.1 kHz, 16 bit resolution and in a mono channel. Only Alap part (45 second length) of each signals were cut and normalized to 0 dB. Fourier transform with FFT point 1024 and with hamming window was performed for spectral analysis. Timbre analysis is done from the long term average spectra (LTAS). Hereafter we shall refer the six samples used for analysis as: sample 1 (Imdad Khan-1905-10), sample 2 (Inayet Khan-1935), sample 3 (Laxmam Bhatacharya-1945-54), sample 4 (Nikhil Banerjee-1970), sample 5 (Vilayet Khan-1985) and sample 6 (Ravi Shankar-1986).

Results and Discussions
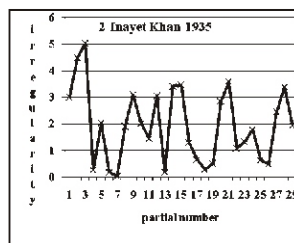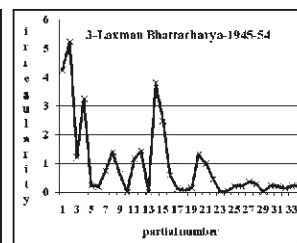
Figure 1   Figure 2   Figure 3
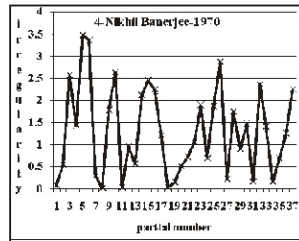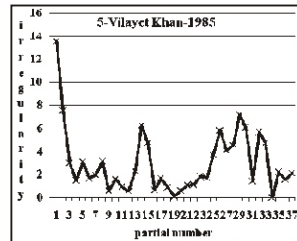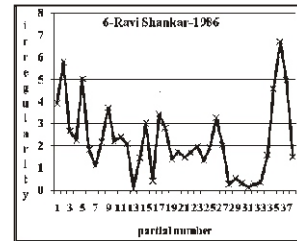
| Figure 4 | Figure 5 | Figure 6 |

Figure 1-6 shows the Spectral Irregularity against the partials for all the six samples. Irregularity among partials of Imdad Khan's sitar is the least and that of Vilayet Khan and Ravi Shankar's sitar is the most in comparison to the others. Since tumba of the instrument is responsible for resonance, we may presume that tumba of Imdad Khan's sitar was smaller in comparison to the others while that of Vilayet Khan and Ravi Shankar's sitar were larger.
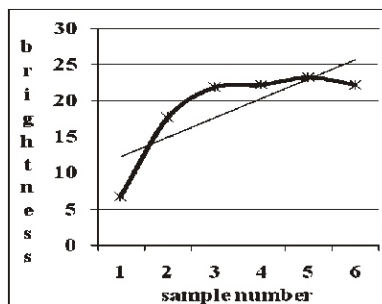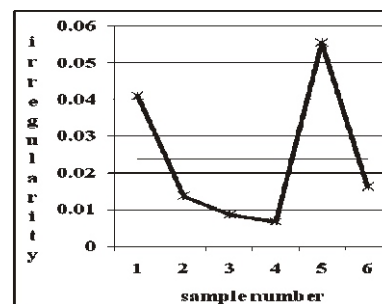
| Figure 7 | Figure 8 |

Fig. 7 and 8 shows the spectral brightness and Irregularity against each samples. We see from fig. 7 that brightness achieves saturation from sample 3 onwards after a steady increase. This might reflect that the structure of sitar became standardized after middle of 20th century. From fig. 8, we may say that a high value of irregularity for Imdad Khan and Vilayet Khan's sitar agrees with the fact that their sitars were different in size than the others.
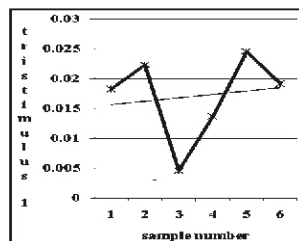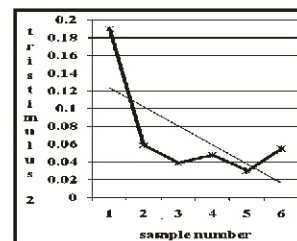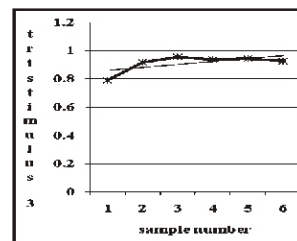
| Figure 9 | Figure 10 | Figure 11 |

Figs. 9,10 and 11 presents the tristimulus T1,T2 and T3 against the sample

number. A low value of T1 (Laxman Bhattacharya and Nikhil Banerjee) signifies strong fundamental while high T1 (Inayet Khan and Vilayet Khan) signifies a weak fundamental i.e. the harmonics are stronger than the fundamental. Tristimulus of Laxman Bhattacharya's sitar has both strong fundamental and strong higher harmonics.
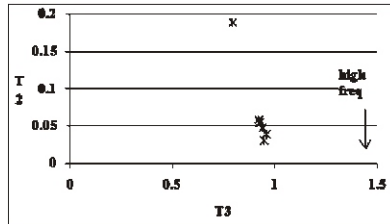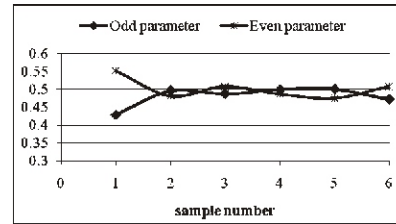
Figure 12   Figure 13

In fig. 12, where tristimulus 2 is a function of tristimulus 3 where the three corners of the low left triangle denote strong fundamental, strong mid-range, and strong high frequency partials. T2 vs T3 graph shows the dominance of high frequency band energy for the other sitars except Imdad Khan's sitar. This ensures that only the relevant fundamental tone and partials are amplified. Larger difference of odd and even relations in fig.13 showed for Imdad Khan's sitar.

In the following graphs (Fig. 14-19) along X axis partial index and along Y axis frequency/partial index (Hz) is plotted.
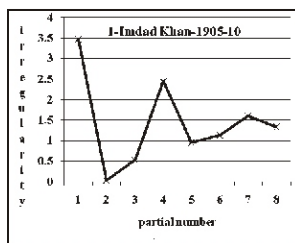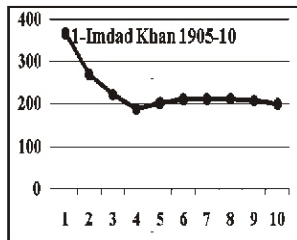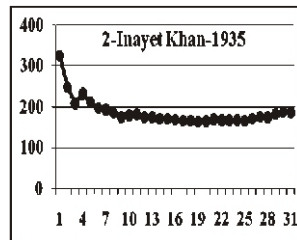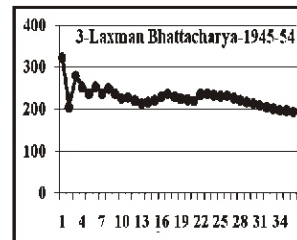

Figure 14   Figure 15   Figure 16
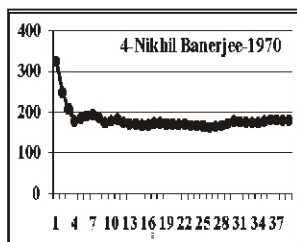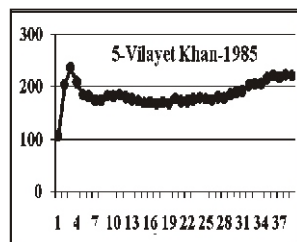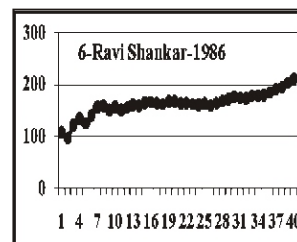
Figure 17   Figure 18   Figure 19

It can be observed from the nature of the curves in figs.14, 15 and 18 that these

sound samples are quasi harmonic having higher level of inharmonicity which is evident in Imdad Khan, Inayet Khan and Vilayet Khan's sitar. Fig 16 and 17 shows straight line i.e. they are harmonic sounds and these sounds exhibiting a constant frequency difference in all pairs of its neighboring partials. Inharmonicity increases at high frequency region for Vilayet Khan and Ravi Shankar's sitar.

From the spectral view it has been clear that the energy among the partials varies high in the modern era than the earlier days. Spectral view for sitar of Imdad Khan and Inayet Khan's sound samples shows that the information lies below 6 kHz. Spectral view for the other sound samples shows that the information lies only within 9 kHz.

Conclusion

Fig.1 and 14 reveals that there is a loss of information in Imdad Khan's sound sample either due to inadequate tool for recordings and filtration process for noise removal in the archive during that time. Also Fig. 7, 9, 11 and 12 reveals that his sitar had smaller tumba causes less resonant structure. Fig. 6 and 19 reveals that acoustic features of Ravi Shankar (1986) and Nikhil Banerjee's (1970) sitar were different from all others. From Fig. 3, 8, 9, 10, 13 and 16 it is clear that sound of Laxman Bhattacharya's (1945-54) sitar differs from the others. Cause of such difference is due to its structure of Tumba (made of wood), tabli and neck.

Reference

1. Yantra Kosha or A treasury of the musical Instruments of Ancient and Modern India, and of verious other countries: S.M. Tagore, Calcutta (1875), Sarmila Prakasani 1983, Calcutta.

2. Ragadarpan (in Mansingh aur Manqutuhala): Faqirullah (17th century), Harihar Niwas Dwivedi (transl), Gwalior, Vidya Mandir Prakasan (1953).

3. An Introduction to Indian Music: B.C. Deva, Publications Division, Ministry of Information And Broadcasting, Government of India, Third revised edition August 1993, New Delhi.

4. Geet Baddyam: Laxmi Narayan Ghosh, Pub: Pratap Narayan Ghosh, 1st edition, Calcutta 1975.

5. Music and musical instruments of Southern India and the Deccan: C. R. Day. Published by: Novello Ewer & Co. & Adam & Charles Black, London 1891, MDCCCXEI.

6.   The Musicians of India Part-1: H.K. Roy Chowdhury, Kuntalin Press (1929), Calcutta.

7.   Rag-Anurag: Ravi Shankar, edited by S.Bhattacharya, Ananda publishers, Calcutta (1999).

8.   Note: interview of M.Adhikari (nephew of Kanilal) and G.Adhikari (s/o Kanilal) on 10.06.2004, Kolkata. Their claim had been accepted by almost all the musical instrument makers in India.

9.   Note: interview of S.Laha (guitar player and close relative of Naderchand Mullick) on 04.07.2004. Kolkata.

10.   Note: interview of Barun Roy (son of Hiren Roy) on 23.06.2004. Kolkata.

11.   The temporal character of timbre, Diploma Thesis by Miha Ciglar miha.ciglar@irzu.org.

12.   Towards Automatic Musical Instrument Timbre Recognition by Tae Hong Park, A Dissertation presented to the Princeton University for PhD degree, November 2004.

13.   Hybrid Perception by Kristoffer Jensen and Georgios Marentakis Datalogisk Institute, University of Copenhagen Universitetsparken1, 2100 Copenhagen {krist, babis}t@diku.dk.

CROSS-MODAL AFFECTIVE PRIMING WITH MUSICAL STIMULI: EFFECT OF MAJOR AND MINOR TRIADS ON WORD-VALENCE CATEGORIZATION

Frank Ragozzine

Psychology Department, Youngstown State University, One University Plaza, Youngstown, OH 44555 USA (phone 330-941-1612; e-mail: frragozzine@ysu.edu).

Abstract

Abstract--Despite well-known effects of music on affect, musical stimuli have rarely been included in affective priming studies. In the present experiments, the effects of major and minor triads on judgments of the valence of visually-presented target words were examined. On each trial, participants heard either a major triad (positive valence) or a minor triad (negative valence), followed by a target word having either positive or negative valence. Participants categorized the target as either positive or negative. In Experiment 1, an analysis of target-categorization errors revealed a significant interaction: Fewer errors were made in valence-congruent conditions (major-positive and minor-negative) than were made in valence-incongruent conditions (major-negative and minor-positive). In Experiment 2, an analogous interaction was obtained in an analysis of response latencies: Participants categorized targets significantly more quickly in valence-congruent conditions than they did in valence-incongruent conditions. Thus, these experiments demonstrate cross-modal affective priming using major and minor triads as primes.

Index Terms — Affective priming, major, minor, triad.

I. Introduction

In a typical affective priming experiment, a prime word having either a positive or a negative valence is briefly presented, followed a short time later by a target word having either a positive or a negative valence. On some trials, the valences of the prime and target are congruent with each other, whereas on other trials, the valences are incongruent. Participants are asked to categorize the valence of the target word as either positive or negative. A common finding is that participants categorize the target-word valence more quickly in valence-congruent conditions than they do in valence-incongruent conditions, due to affective priming [1]-[3]. One would expect analogous results when analyzing errors: Affective priming should lead to more target-categorization errors in

valence-incongruent conditions than in valence-congruent conditions. However, interactions involving errors have been less frequently demonstrated than interactions involving response latencies, perhaps because of research designs that maximize the chances of observing response latency effects while at the same time reducing the chances of observing any significant pattern of results with respect to errors. For example, many researchers have used a somewhat small number of targets (e.g., approximately 20 targets are typically used in many experiments), or have preselected stimuli for each individual participant, leading to too small a number of errors for meaningful analysis.

In many studies of affective priming, the prime and target are presented in the same sensory modality (e.g., two visually-presented words). However, some researchers have demonstrated cross-modal affective priming effects. For example, researchers have used odors as primes and visually-presented words as targets, and have found that congruent valences lead to faster categorization of the target [4]; an analogous effect was obtained in a study using flavors as primes, with visually-presented words as targets [5]. Others have found affective priming effects using spoken words as primes and abstract visual images as targets [6].

One stimulus category that is well-known for its ability to convey affect is music. The emotional impact of music has been recognized and discussed since the time of the ancient Greeks [7]. Yet, systematic research on music and emotion has, historically, largely been neglected by psychological researchers, in contrast with the large body of published work dealing with music cognition [8] (see also [9] for a discussion of historical reasons for this disparity). Likewise, although there are multiple published examples of conceptual priming with musical stimuli (e.g., [10], [11]), affective priming studies that employ musical stimuli are scarce. In one published study of cross-modal affective priming using musical stimuli, Sollberger and colleagues [12] presented consonant or dissonant musical chords as primes, followed by visually-presented target words that participants categorized as either positive or negative. In music, consonance has positive valence, whereas dissonance has negative valence [13], [14]. Consistent with other research on affective priming, these researchers found that participants responded more quickly in valence-congruent conditions (consonant chord followed by positive word or dissonant chord followed by negative word) than they did in valence-incongruent conditions (consonant chord followed by negative word or dissonant chord followed by positive word). Additionally, these researchers were among the few to have demonstrated an effect with errors as well as with response latency. Specifically, they found that fewer errors were made in valence-congruent conditions than in valence-incongruent conditions.

In addition to consonant and dissonant chords, other sorts of musical stimuli convey affective meaning. Of particular interest, Hevner [14], [15] (see also [16]) found that, for Western listeners, music in minor keys has negative valence, whereas music in major keys has positive valence. These conventional valence judgments of major and minor are reliably obtained, and are made even by musically untrained listeners [15]. Several researchers have found that even children perceive the conventional valences of major and minor [17]-[19]. The perceived valences of major and minor are not limited to extended musical passages; they are also perceived with just a single major or minor chord. For example, Crowder [20] found that single major or minor chords are reliably categorized as having positive valence or negative valence, respectively, by adult listeners.

The purpose of the present experiments was to examine whether cross-modal affective priming occurs when major or minor triads (three-note musical chords) are paired with visually-presented positive or negative target words. Specifically, on valence-congruent trials, participants experienced either a major triad prime followed by a positive target word, or a minor triad prime followed by a negative target word. On valence-incongruent trials, participants experienced either a major triad prime followed by a negative target word, or a minor triad prime followed by a positive target word. It was hypothesized that there would be significant interactions between triad type and target-word valence such that fewer target-categorization errors would be made, and shorter response latencies would be obtained, in the valence-congruent conditions than in the valence-incongruent conditions.

II. Experiment 1

A. Purpose

Experiment 1 was conducted in order to examine whether there was any evidence for a cross-modal affective priming effect of major and minor triads on the perceived valence of the target-words. A large set of target words (50 positive targets and 50 negative targets) was chosen in order to allow for meaningful examination of target-categorization errors. It was expected that there would be an interaction between triad type and target valence such that there would be fewer errors and shorter response latencies in the valence-congruent conditions than in the valence-incongruent conditions.

B. Method

1) Participants: Twelve undergraduate students (three men and nine women, ages 18 to 27, M = 19.75) from Youngstown State University received extra credit in their psychology courses for their participation. All participants were

native English speakers. Two participants indicated that they had received more than four years of musical training; the remaining participants had received little or no musical training. The two musically trained participants were the only ones who were able, when asked at the conclusion of the study, to describe a difference between major and minor in music, and who were aware of the conventional valences of major and minor. However, none of the participants were able to correctly guess the hypothesis of the study.

2) Musical Stimuli: The triads were generated on a computer using the Csound synthesis program. Specifically, each complex tone used to create the triads consisted of four harmonically-related sinusoidal components. The relative amplitudes of the components were as follows: The fundamental frequency (f0) was set at 1, the second harmonic (2f0) was set at .75, the third harmonic (3f0) was set at .50, and the fourth harmonic (4f0) was set at .25. Using these complex tones, two root-position triads based on middle-C were created: The C-major triad consisted of the notes C4-E4-G4, and the C-minor triad consisted of the notes C4-Eb4-G4. Only triads based on C were used, in order to avoid confounds with pitch height, which has also been shown to affect valence [21]. For each triad, the three individual notes making up the triad sounded simultaneously. Each triad had a duration of 1,000 ms. Onset and offset amplitude envelopes lasting 10 ms were used in order to avoid potential unwanted transient noises. After the triads had been generated with Csound, the sound files were saved as .wav files on a PC. During the experiment, the triads were presented at a comfortable listening level of approximately 69 dB SPL.

3) Target Words: One hundred English target-words (50 positive and 50 negative) were taken from a paper published by Meier and Robinson [22]. These words were chosen because they had been used in previously published studies [22], [23] in tasks requiring word-valence categorization. These researchers had also pilot-tested these words and found that the positive and negative words lists were comparable on measures such as word length, but differed significantly in terms of positive or negative affect. Additionally, this set of words was sufficiently large to allow for a meaningful examination of error data.

4) Materials and Apparatus: The audio signal was routed from a PC through an M-Audio Delta 66 interface to a Mackie 1604-VLZ Pro mixer, and then into a Furman HA6-AB headphone amplifier. Participants listened to the triads through Sennheiser HD600 headphones. A Quest Technologies model 2900 sound level meter was used to determine the sound level of the triads when presented through the headphones. The target words were presented on a 17"

LCD computer monitor in black 28-point Tahoma regular font against a white background. The target words appeared in all lower-case letters, and were centered on the screen. SuperLab 4.0 software was used to control all stimulus presentation and to record participant responses. Participants indicated their responses using a Cedrus RB-834 8-button response pad with one large button labeled 'Positive' and another large button labeled 'Negative'. A questionnaire was created in order to determine basic demographic information as well as the extent and nature of musical training, and to determine the participants' knowledge of major and minor in music.

5) Procedure: In order to pair the words with the triads, the positive words were first grouped into subsets of words having the same number of letters. Then, half of the words in each subset were paired with a major triad, and the remainder were paired with a minor triad. The same procedure was used to pair triads with negative words. The result was that the list of positive words paired with major triads had the same number of letters as the list of positive words paired with minor triads, and likewise for the lists of negative words. The order of the 100 trials (25 major-positive, 25 major-negative, 25 minor-positive, and 25 minor-negative) was then randomized, and all subjects experienced the same randomized order of trials.

For purposes of counterbalancing, two versions of the experiment were created. Version A was created as described in the previous paragraph. In Version B of the experiment, words that had been paired with a major triad in Version A were instead paired with a minor triad, and words that had been paired with a minor triad in Version A were instead paired with a major triad. An equal number of participants were randomly assigned to Version A and to Version B of the experiment. Data from both versions were averaged for data analysis.

Participants were tested individually in a quiet room, and testing took approximately 30 minutes. Participants were seated approximately .75 m from the computer monitor such that they could easily read the target words on the monitor and comfortably reach the buttons on the response pad.

On each trial, participants first heard a triad lasting 1,000 ms, during which time a blank white screen appeared on the computer monitor. Participants were told that this sound was a cue that indicated that a target word was about to appear. Immediately following the offset of the triad, a target word appeared on the computer monitor, and this word remained on-screen until the participant responded. Participants were told to indicate whether the target word had a positive meaning or a negative meaning, and to indicate their response by

pressing the corresponding button on the response pad. Participants were instructed to respond both quickly and accurately. There was a 2 s intertrial interval (ITI) after each trial, during which time no sound was presented and a blank white screen appeared on the computer monitor. Participants received 12 practice trials before the experiment began, and none of the target words presented during the practice trials appeared again during the experiment. At the conclusion of the experiment, participants completed the questionnaire and were debriefed.

C. Results and Discussion

A 2x2 repeated-measures ANOVA was used to analyze target-categorization errors, with target-word valence (positive or negative) and triad type (major or minor) as factors. Before running the analyses, an arcsine transformation was applied to the proportion of errors in each condition for each participant, and analyses were conducted using these transformed scores [24]. However, for clarity, descriptive statistics are reported using the raw number of errors [24]. As can be seen in Figure 1, a significant interaction was obtained, such that the number of target-categorization errors was greater in the valence-incongruent conditions (minor-positive [M = 1.67, SE = 0.33] or major-negative [M = 2.08, SE = 0.57]) than in the valence-congruent conditions (major-positive [M = 0.67, SE = 0.28] or minor-negative [M = 1.33, SE = 0.40]): $F(1, 11) = 16.73$, MSE = .01, $p = .002$. This effect was a large one, $\eta p2 = .60$. A specific planned comparison showed that participants made significantly more errors in the valence-incongruent conditions than they made in the valence-congruent conditions, $F(1, 11) = 11.12$, MSE = .01, $p = .007$, $\eta p2 = .50$. There were no significant main effects of either target-word valence or triad type: $F(1, 11) = 1.76$, MSE = .02, $p = .211$; and $F(1, 11) = .26$, MSE = .02, $p = .622$, respectively.

An additional 2x2 repeated-measures ANOVA was used to analyze the response latency data, again with target-word valence and triad type as factors. Errors were excluded in this analysis (5.75% of judgments). Additionally, consistent with previous affective priming research (e.g., [4]), outliers were defined as reaction-times greater than 1,500 ms or less than 250 ms, and these outliers were excluded from the analyses. Unexpectedly, there was no significant interaction between target-word valence and triad type, $F(1, 11) = .64$, MSE = 578.21, $p = .441$. However, there was a significant main effect of target-word valence, such that positive words (M = 760.94, SE = 28.34) were categorized more quickly than were negative words (M = 810.83, SE = 23.97): $F(1, 11) = 15.00$, MSE = 1991.64, $p = .003$, $\eta p2 = .58$. There was no significant main effect of triad type, $F(1, 11) = .01$, MSE = 1659.87, $p = .945$.

As predicted, a significant affective priming effect with respect to target-categorization errors was obtained, such that fewer errors were made when the
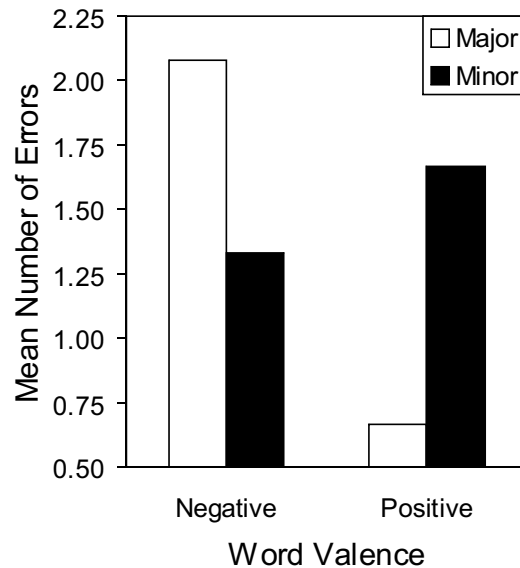
Fig. 1. Mean number of target-categorization errors as a function of target-word valence, shown separately for target words preceded by major triads (white bars) and target words preceded by minor triads (black bars).

valences of the prime and target were congruent than when they were incongruent. Inconsistent with most previous affective priming research, an analogous pattern of results was not found in analyses of the response latency data. However, there were several methodological differences between the present experiment and most other affective priming experiments that resulted in a lack of power to detect any effects in the response latency analysis. Indeed, a-posteriori observed power for the interaction with respect to response latency was only .11 in the present experiment, whereas the observed power for the interaction with respect to errors was .96. Thus, unlike most affective priming experiments in which significant effects are obtained for response latencies but not for target-categorization errors, the paradigm used in Experiment 1 was useful for detecting effects with the error data but was not sensitive to analogous effects with respect to response latencies. Therefore, an additional experiment was performed in which various methodological changes were made in an effort to enhance the power of the experimental paradigm with respect to response latency.

III. Experiment 2

A. Purpose

The purpose of Experiment 2 was to determine whether a pattern of results for response latency analogous to those obtained with target-categorization errors in Experiment 1 could be obtained by using a more powerful paradigm. In Experiment 1, there was wide variability across participants in terms of the amount of time it took to classify some of the words. Specifically, the variability across participants in the response latencies to categorize each target word was analyzed, and this analysis revealed that the standard errors ranged from 22.74 ms to 290.85 ms. In other words, the time it took to classify the target word was quite consistent across participants for some of the words, but varied dramatically across participants for other words. This variability in the response-latency data associated with particular target words contributed to error variability that may have been masking any effects of the triads on target-word categorization. Thus, a word-selection experiment was first conducted in order to choose a more suitable word list for which this potential source of error variability was reduced.

In Experiment 1, the stimulus-onset asynchrony (SOA), which is the length of time between the onset of the prime and the onset of the target, was 1,000 ms. However, some previous research indicates that affective priming effects on response latency are diminished or nonexistent when the SOA is greater than 300 ms (e.g., [2]). Therefore, in Experiment 2, the SOA was shortened to 250 ms. In addition, several other methodological modifications were made to Experiment 2. Specifically, in Experiment 1, participants were told that the triads were a cue that a word was about to appear, in order to encourage them to attend to the triads. Whether a participant attends to the primes has been shown to play a role in some affective priming studies [25], [26]. However, in Experiment 1, the ITI remained constant across trials (2 s). Thus, participants had an additional temporal cue that a word was about to appear, perhaps resulting in less focused attention to the primes over the course of the experiment. Therefore, a randomly-varying ITI was used in Experiment 2 in order to make the triads serve as a more effective cue, and thus encourage greater attention to the triads. Additionally, in Experiment 1, both speed and accuracy were emphasized in the experimental instructions. Therefore, in order to make it more likely to detect any potentially significant effects when analyzing response latency, accuracy was strongly emphasized in the experimental instructions for Experiment 2. Finally, in Experiment 1, the triads were generated using Csound, which resulted in musical stimuli that sounded unlike what is typically heard in most music. In Experiment 2, synthesized Steinway piano tones were used to create more natural-sounding musical

stimuli. With these changes, it was hypothesized that there would be a significant interaction between target-word valence and triad type when analyzing response latency data in Experiment 2.

B. Method

1) Participants: Ten undergraduate students (seven men and three women, ages 18 to 24, M = 20.30) participated. All participants received extra credit in their course for their participation.

All participants were native English speakers. Three participants indicated that they had received more than four years of musical training; the remaining participants had received little or no musical training. When asked at the conclusion of the study to describe what they knew regarding major and minor, only two of the musically trained participants were able to provide a response; however, none of the participants mentioned anything regarding the valences of major and minor in their responses. None of the participants were able to correctly guess the hypothesis of the study.

2) Word-Selection: In order to obtain a set of words for use in Experiment 2 for which error variability in response latencies was minimized, a word-selection experiment was first conducted. Participants were 22 undergraduate students, all of whom were native English speakers. None of these students participated in either Experiment 1 or in the main portion of Experiment 2. These participants were presented with the same set of 100 target words used in Experiment 1, in a random order. However, in this word-selection experiment, the target words were not preceded by a triad. Participants were asked to categorize each word as positive or negative. Words that had been incorrectly categorized by any participant were then eliminated, as were words that, averaged across participants, led to response-latency standard errors greater than 100 ms. From the set of remaining words, in order to create comparable positive and negative words lists of equal length, the 14 positive words and 14 negative words that resulted in the lowest standard errors in the response-latency data for were chosen for inclusion in the experiment. Specifically, the positive words chosen were candy, champion, clean, ethical, garden, gentle, gracious, kiss, neat, nurse, satisfying, sweet, trust, and truthful. The negative words chosen were cancer, clumsy, cruel, dead, devil, enemy, fraud, liar, nasty, obnoxious, poison, rude, sloppy, and vulgar. There were no significant differences between these positive and negative word lists in terms of the mean response latencies in categorizing the words, or in the number of letters in the words in the two lists (in both cases, p > .05, independent-samples t-test, two-tailed).

3) Musical Stimuli: The C-major triad consisted of the notes C4-E4-G4, with all notes occurring simultaneously. The C-minor triad consisted of the notes C4-Eb4-G4, with all notes occurring simultaneously. Finale 2007 was used to create these two root-position triads. Specifically, the preset Steinway piano sound from the Garritan Personal Orchestra 2.0 Finale Edition was used, and was played through the Native Instruments VST without reverberation. Adobe Audition 2.0 was then utilized in order to verify that the overall amplitudes of the two triads were the same. The duration of each triad was 250 ms. Onset and offset amplitude envelopes lasting 10 ms were used to avoid any potential unwanted transient noises. Each triad was saved as a .wav file on a PC for use in the experiment.

4) Materials and Apparatus: The same materials and apparatus used in Experiment 1 were used in Experiment 2.

5) Procedure: The following procedure was used to create Version A of the experiment: The words in the positive list were rank-ordered from low to high based on the mean response latencies obtained in the word-selection experiment. Then, the odd-numbered words in this ranked list were paired with a major triad, and the even-numbered words in this ranked list were paired with a minor triad. The same procedure was used to pair words in the negative list with major and minor triads. The result of this procedure was that the target words in the four conditions were equivalent, based on the word-selection experiment data, in terms of the response latencies to categorize the targets. In Version B of the experiment, words that had been paired with a major triad in Version A were paired with a minor triad, and words that had been paired with a minor triad in Version A were paired with a major triad. Thus, two counterbalanced versions of the experiment were created, each with 7 major-positive trials, 7 minor-positive trials, 7 major-negative trials, and 7 minor-negative trials. Data from both versions were averaged for data analysis.

Participants were tested as in Experiment 1, except that they received 4 practice trials before beginning the experiment, and testing took approximately 15 minutes. Half of the participants were randomly assigned to receive Version A of the experiment, and the remaining half received Version B of the experiment.

At the beginning of each trial, there was a either a 2, 4, 6, 8, or 10 s ITI, the length of which was determined randomly on each trial, during which participants heard nothing and saw a blank white screen while they waited for the stimulus presentation. Following the ITI, participants heard a triad lasting 250 ms. Triads were presented at a comfortable listening level of approximately 69 dB SPL. Participants were told that the purpose of this sound was to serve as a

cue to indicate that a word was about to appear, and by using a randomly-varying ITI, the triad did in fact serve as the only reliable cue that a word was about to appear. A blank white screen was presented on the computer monitor while the triad sounded. Immediately following the offset of a triad, a target word appeared on the computer monitor and remained on-screen until the participant responded. Participants were instructed to indicate whether the word had positive meaning or negative meaning by pushing the corresponding
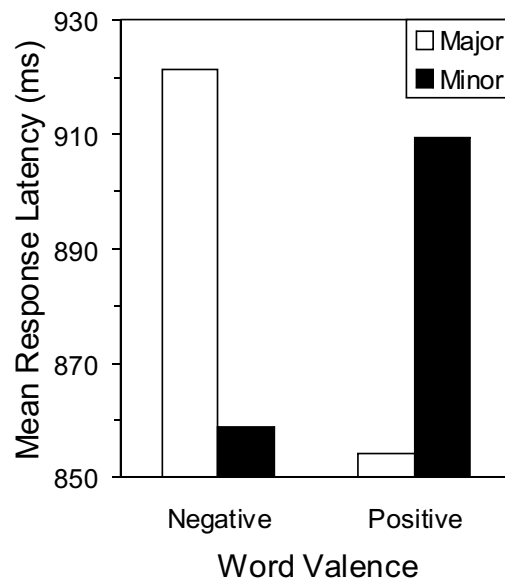
Fig. 2. Mean response latency, measured in milliseconds, as a function of target-word valence, shown separately for target words preceded by major triads (white bars) and target words preceded by minor triads (black bars).

labeled button on the response pad. Participants were further told to avoid making any errors, and that if they did make an error, they would see an error message that would remain on screen for 4 s that would inform them that their response was incorrect. Participants experienced 28 experimental trials, and the order of trials was randomly determined for each participant. At the conclusion of the experiment, participants completed the questionnaire and were debriefed.

C. Results and Discussion

A 2x2 repeated-measures ANOVA, with triad type (major or minor) and target-word valence (positive or negative) as factors, was used to analyze the response latency data. Trials on which errors were made would have been excluded in the analyses, but given the strong emphasis on accuracy as well as

the fact that the words used were those on which no errors had been made in the word-selection study, there were no errors made during the experimental trials. Outlier response latencies greater than 1,500 ms or less than 250 ms were excluded from the analyses.

As predicted, there was a significant interaction between triad type and target-word valence: $F(1, 9) = 7.67$, MSE = 4556.86, $p = .022$. This effect was a large one, $\eta p2 = .46$. As can be seen in Figure 2, participants responded more quickly in the valence-congruent conditions (major-positive [M = 854.07, SE = 83.26] or minor-negative [M = 858.64, SE = 49.50]) than they did in the valence-incongruent conditions (major-negative [M = 921.41, SE = 92.80] or minor-positive [M = 909.53, SE = 74.75]). A planned comparison showed that there was a significant difference in the expected direction between the valence-congruent conditions and the valence-incongruent conditions, $F(1, 9) = 6.58$, MSE = 1880.12, $p = .030$, $\eta p2 = .42$. There was no significant main effect of triad type: $F(1, 9) = .01$, MSE = 17165.95, $p = .932$. Likewise, there was no significant main effect of word valence: $F(1, 9) = .13$, MSE = 5245.63, $p = .728$. Thus, by using an experimental paradigm that was more consistent with that used in most affective priming research, affective priming effects were obtained such that target words were evaluated more quickly when preceded by valence-congruent primes than they were when preceded by valence-incongruent primes.

IV. General Discussion

Cross-modal affective priming was obtained using major and minor triads as primes, with visually-presented words as targets. In contrast to within-modality affective priming, there are comparatively fewer studies in the literature in which cross-modal affective priming has been observed, particularly with musical stimuli. Additionally, unlike most affective priming studies, an effect was obtained with target-categorization errors (Experiment 1) in addition to the more typical finding of an affective priming effect for response latencies (Experiment 2). Due to speed-accuracy tradeoff issues, finding effects with both errors and response latencies provides more compelling evidence for a priming effect.

Quite recently, after the present experiments had been conducted and presented at national conferences in the USA, another group of researchers published a study in which they reported an affective priming effect using major and minor chords [27]. However, there are several notable differences between [27] and the experiments reported in the present paper. As in the present paper, the authors of [27] obtained a priming effect with response

latencies; however, unlike the present experiments, the researchers in [27] were unable to demonstrate a priming effect with errors. As discussed above, due to the speed-accuracy tradeoff issue, evidence for a priming effect is more convincing when such effects are demonstrated with both response latencies and errors. Additionally, the authors of [27] did not report the size of the priming effect they obtained. As reported in the present paper, the effect sizes of these priming effects are quite large, with values for ηp2 of .60 in Experiment 1 and .46 in Experiment 2. Another key difference between these studies is that the researchers in [27] used chords based on all 12 pitch classes within an octave range. However, even though pitch class was included as a variable in their experimental design, they did not include pitch class as a factor in their analyses. In contrast, in the present experiments, all triads were based on a single pitch class (middle-C) in order to avoid any influence of pitch height, which has been shown to affect perceived valence [21]. It is not possible to determine the extent to which the results in [27] were influenced by pitch height. Finally, the researchers in [27] used four-note chords in which the root pitch was doubled, in contrast to the triads used in the present experiments. Thus, the present experiments demonstrate affective priming with even more minimal stimuli than those used in [27].

A variety of explanations for the perceived valences of major and minor have been proposed (see [28] for a review). One theory, proposed by Helmholtz [29], is that minor triads are inherently more dissonant than are major triads. However, Helmholtz acknowledged that this was more likely to occur with just-intonation tuning than with the equal-temperament tuning used in modern times (and in the present experiments). An alternative theory is that the valences of major and minor are learned during childhood, and most of the evidence supports this view. For example, children from approximately 6 to 8 years of age perceive the conventional valences of major and minor, whereas younger children do not [17], [18] (although, see [19]). In contrast, the perceived valences of consonance and dissonance are present in infancy, with 6-month old infants showing a preference for consonance over dissonance [30], [31]. One researcher has found that even two-day old infants of deaf parents showed the same preference for consonance over dissonance as did two-day old infants of hearing parents [32], suggesting an inborn basis for the perceived valences of consonance and dissonance. Thus, it seems unlikely that the present findings are simply due to the same mechanisms as those involved in the perceived valences of consonance and dissonance. Additionally, the differences between the major and minor triads used in the present experiment are arguably more subtle than the differences between the consonant and dissonant chords used in [12]. Specifically, in the present experiments, the major and minor triads differed by a single note, and this note differed by only one semitone (i.e., E in

the major triad, and Eb in the minor triad).

Affective priming is thought to occur quickly and automatically, rather than as a result of a conscious and deliberative process [1], [2]. The present experiments show that the valences of major and minor are activated even with a single briefly-presented triad, as opposed to lengthier musical passages. Additionally, no participants in the present experiments were able to accurately guess the experimental hypotheses. Furthermore, when asked to describe their knowledge of major and minor in music, only two participants (both in Experiment 1) mentioned the conventional valences of major and minor, despite the fact that they had just listened to a large number of major and minor triads in a valence-related task. Indeed, only a few participants indicated that they had even noticed a difference in the 'sound cues'; most of those that did notice a difference did not think it was particularly relevant to the experiment. These findings lend further support to the idea that affective priming is the result of an implicit process.

Finally, because the primary purpose of the present experiments was to demonstrate cross-modal affective priming effects with musical stimuli for both errors and response latencies, there were numerous methodological differences between Experiment 1 and Experiment 2. However, it may be worthwhile in future research to conduct a series of experiments in order to isolate which particular aspects of the paradigm are more conducive to finding an effect with errors, and which are more conducive to finding an effect with response latencies.

Acknowledgment

References

1.   J J. A. Bargh, S. Chaiken, P. Raymond, and C. Hymes, "The automatic evaluation effect: Unconditional automatic attitude activation with a

pronunciation task," Journal of Experimental Social Psychology, vol. 32, no. 1, pp. 104-128, Jan. 1996.

2. R. H. Fazio, "On the automatic activation of associated evaluations: An overview," Cognition and Emotion, vol. 15, no. 2, pp. 115-141, Mar. 2001.

3. R. H. Fazio, D. M. Sanbonmatsu, M. C. Powell, and F. R. Kardes, "On the automatic activation of attitudes," Journal of Personality and Social Psychology, vol. 50, no. 2, pp. 229-238, Feb. 1986.

4. D. Hermans, F. Baeyens, and P. Eelen, "Odours as affective processing context for word evaluation: A case of cross-modal affective priming," Cognition and Emotion, vol. 12, no. 4, pp. 601-613, Jul. 1998.

5. M. G. Veldhuizen, A. F. Oosterhoff, and J. H. A. Kroeze, "Flavors prime processing of affectively congruent food words and non-food words," Appetite, vol. 54, no. 1, pp. 71-76, Feb. 2010.

6. K. L. Duckworth, J. A. Bargh, M. Garcia, and S. Chaiken, "The automatic evaluation of novel stimuli," Psychological Science, vol. 13, no. 6, pp. 513-519, Nov. 2002.

7. N. Cook and N. Dibben, "Musicological approaches to emotion," in Music and Emotion: Theory and Research, P. N. Juslin and J. A. Sloboda, Eds. New York: Oxford University Press, 2001, pp. 45-70.

8. P. N. Juslin and J. A. Sloboda, J. A, "Music and emotion: Introduction," in Music and Emotion: Theory and Research, P. N. Juslin and J. A. Sloboda, Eds. New York: Oxford University Press, 2001, pp. 3-20.

9. I. Peretz, "Listen to the brain: A biological perspective on musical emotions," in Music and Emotion: Theory and Research, P. N. Juslin and J. A. Sloboda, Eds. New York: Oxford University Press, 2001, pp. 105-134.

10. J. J. Bharucha and K. Stoeckig, "Reaction time and musical expectancy: Priming of chords," Journal of Experimental Psychology: Human Perception and Performance, vol. 12, no. 4, pp. 403-410, Nov. 1986.

11. S. Koelsch, E. Kasper, D. Sammler, K. Schulze, T. Gunter, and A. D. Friederici, A. D, "Music, language and meaning: Brain signatures of semantic processing," Nature Neuroscience, vol. 7, no. 3, pp. 302-307, Mar. 2004.

12. B. Sollberger, R. Reber, and D. Eckstein, "Musical chords as affective priming context in a word-evaluation task," Music Perception, vol. 20, no. 3, pp. 263-282, Spring 2003.

13. A. J. Blood, R. J. Zatorre, P. Bermudez, and A. C. Evans, "Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions," Nature Neuroscience, vol. 2, no. 4, pp. 382-387,

Apr. 1999.

14. K. Hevner, "Experimental studies of the elements of expression in music," The American Journal of Psychology, vol. 48, no. 2, pp. 246-268, Apr. 1936.

15. K. Hevner, "The affective character of the major and minor modes in music," The American Journal of Psychology, vol. 47, no. 1, pp. 103-118, Jan. 1935.

16. L. Gagnon and I. Peretz, "Mode and tempo relative contributions to 'happy-sad' judgements in equitone melodies," Cognition and Emotion, vol. 17, no. 1, pp. 25-40, Jan. 2003.

17. S. Dalla Bella, I. Peretz, L. Rousseau, and N. Gosselin, "A developmental study of the affective value of tempo and mode in music," Cognition, vol. 80, no. 3, pp. B1-B10, Jul. 2001.

18. G. M. Gerardi and L. Gerken, "The development of affective responses to modality and melodic contour," Music Perception, vol. 12, no. 3, pp. 279-290, Spring 1995.

19. M. P. Kastner and R. G. Crowder, "Perception of the major/minor distinction: IV. Emotional connotations in young children," Music Perception, vol. 8, no. 2, pp. 189-202, Winter 1990.

20. R. G. Crowder, "Perception of the major/minor distinction: III. Hedonic, musical, and affective discriminations," Bulletin of the Psychonomic Society, vol. 23, no. 4, pp. 314-316, Jul. 1985.

21. K. Hevner, "The affective value of pitch and tempo in music," The American Journal of Psychology, vol. 49, no. 4, pp. 621-630, Oct. 1937.

22. B. P. Meier and M. D. Robinson, "Why the sunny side is up: Associations between affect and vertical position," Psychological Science, vol. 15, no. 4, pp. 243-247, Apr. 2004.

23. B. P. Meier, M. D. Robinson, and G. L. Clore, "Why good guys wear white: Automatic inferences about stimulus valence based on brightness," Psychological Science, vol. 15, no. 2, pp. 82-87, Feb. 2004.

24. D. C. Howell, Statistical Methods for Psychology, 7th ed. Belmont, CA: Cengage Wadsworth, 2010.

25. J. De Houwer and T. Randell, "Attention to primes modulates affective priming of pronunciation responses," Experimental Psychology, vol. 49, no. 3, pp. 163-170, Jul. 2002.

26. J. P. Simmons and D. A. Prentice, "Pay attention! Attention to the primes increases attitude assessment accuracy," Journal of Experimental Social Psychology, vol. 42, no. 6, pp. 784-791, Nov. 2006.

27. N. Steinbeis and S. Koelsch, "Affective priming effects of musical sounds on the processing of word meaning," Journal of Cognitive Neuroscience, vol. 23, no. 3, pp. 604-621, Mar. 2011.

28. R. G. Crowder, "Perception of the major/minor distinction: I. Historical and theoretical foundations," Psychomusicology: A Journal of Research in Music Cognition, vol. 4, no. 1-2, pp. 3-12, Spring 1984.

29. H. I. F. von Helmholtz, On the Sensations of Tone. New York: Dover, 1877/1954.

30. R. G. Crowder, J. S. Reznick, and S. L. Rosenkrantz, "Perception of the major/minor distinction: V. Preferences among infants," Bulletin of the Psychonomic Society, vol. 29, no. 3, pp. 187-188, May 1991.

31. L. J. Trainor and B. M. Heinmiller, "The development of evaluative responses to music: Infants prefer to listen to consonance over dissonance," Infant Behavior & Development, vol. 21, no. 1, pp. 77-88, 1998.

32. N. Masataka, "Preference for consonance over dissonance by hearing newborns of deaf parents and of hearing parents," Developmental Science, vol. 9, no. 1, pp. 46-50, Jan. 2006.

LINKING RAGA WITH PROBABILITY

Swarima Tewari and Soubhik Chakraborty*
Department of Applied Mathematics
Birla Institute of Technology Mesra, Ranchi-835215, India
*Corresponding author's email: soubhikc@yahoo.co.in

Abstract

The note attempts to link raga with probability. We argue that the same note will have different probabilities in different ragas depending on its absence or presence and relative importance in case of presence. The most important note statistically speaking is not necessarily the one having the highest probability but one having a high probability that is maintained consistently. This concept of statistical pitch stability is different from Krumhansl's psychological pitch stability based on note duration. Both concepts can be combined to give rise to psycho-statistical pitch stability by ranking the notes differently using the two contrasting concepts and then giving the notes the average rank.

Key words : Raga, probability, entropy, {statistical, psychological} pitch stability

1. Introduction

A raga, in Indian classical music, is a melodic structure with fixed notes and a set of rules depicting a certain mood conveyed by performance. Raga and probability have an interesting connection. Pick up any raga say R and a particular musical note say n. With respect to the raga R, the note n can be either (i) *Vadi*, (ii) *Samvadi*, (iii) *Anuvadi*, (iv) *Alpvadi* or (v) *Vivadi* according as (i) it is the most important note in R, (ii) the second most important note in R, (iii) an important note in R but not *Vadi or Samvadi*, (iv) an unimportant note (but permissible) in R or (v) not to be used at all in R (otherwise R becomes *Mishra* or mixed and not *Suddha* or pure) respectively. We argue that these five classifications of a note merits probabilistic considerations immediately [1] in the sense that a note belonging to the first three groups is likely to have a relatively high probability compared to a note of the fourth group which has a small probability. A note of the last group will have zero probability. Further, the most important note in a raga statistically speaking is not necessarily the *one with the highest probability but one which has a high probability and it maintains this high probability over the instances of realization of the note*. Herein lies the concept of statistical pitch stability [2], which is different from Krumhansl's concept of psychological pitch stability based on note duration [3][4].

From entropy considerations, a note with small probability however has a

corresponding high surprise element. Since entropy actually measures the surprise element in a message (in this case the realization of a note), it is also possible to distinguish ragas based on entropy using the aforesaid link between raga and probability [5].

Most of these arguments are supported in the present paper with experimental results. For others, references are provided. For our purpose, we have investigated with the *pancham swar* (Pa) for a sequence of 100 notes of five different ragas, one for each note group as mentioned earlier, taken from a standard text [6]. The reader is encouraged to experiment with other notes.

2. Probabilistic analysis

If P(E) is the probability of an event, the information content of the event E is defined as $I(E) = -\log_2(P(E))$. Events with lower probability will signal higher information content when they occur. The choice of the base 2 of logarithm is motivated by the following case. Consider a Bernoulli random variable with $P(X=0)=P(X=1)=1/2$. We have $I(X=0)=I(X=1)=-\log_2(1/2) = 1$. The units of information content are bits. So, we gain one bit of information when we choose between two equally likely alternatives. Now, let X be a discrete random variable which takes values $x_1 x_2 x_3 \ldots \ldots x_n$ with corresponding probabilities $p_1 p_2 p_3 \ldots \ldots p_n$. Since X is a random variable, the information content of X is also random which we denote by I(X) (what value I(X) will take depends on what value X takes). When $X = x_j$ which is an event with probability p_j then $I(X)=-\log_2(p_j)$. Accordingly, it makes sense to talk about the mean value of I(X) called its **entropy**, denoted by H(X), so that we have

$$H(X) = -\Sigma\, p_j \log_2(p_j),\ \text{where the summation is over } j=1 \text{ to } n.$$

For an impossible event E, $P(E)=0$, $I(E)= -\infty$. As negative information is ruled out, it indicates the non-feasibility of ever obtaining information about an impossible event (*). Also, we shall define $plog(p) = 0$ when $p = 0$. The range of $plog(p)$ is thus $[0,)$. We emphasize here that entropy is measuring surprise (not meaning!). In music, a note with low probability of coming will have more surprise. The use of entropy in music analysis has been successfully tried in Western music (see Snyder [7]). For more on entropy, see Applebaum [8]. Our analysis is given in table 1.

Table 1: Information Content of *Pancham swar* (Pa) in five different ragas

Raga	Relative occurrence of *Pancham swar* out of first 100 notes=p	Classification of *Pancham swar* in the concerned Raga	Information Content= -log₂(p)
Bageshree	2/100 = 0.02	*Alpvadi*	5.6439
Malkauns	0/100 = 0	*Vivadi*	Infinity (*)
Kafi	18/100 = 0.18	*Vadi*	2.4739
Bhupali	19/100 = 0.19	*Anuvadi*	2.3959
Desh	19/100 = 0.19	*Samvadi*	2.3959

*sometimes a *vivadi swar* is used at a "peak point" in a performance to create romanticism (unrestricted beautification) in contrast to classicism (disciplined beautification). Well, this only proves that an event with zero probability can be possible! In such a case, the probability of a *vivadi swar*, theoretically zero, will have some non zero though very small value empirically. Accordingly, the information content will be calculated. The previous argument to take the information content to be negatively infinite fails when the event in question is possible with zero probability. It is rational to take this information content as positively infinite theoretically to indicate both the possibility and the infinite surprise. It is this theoretical infinite surprise which is empirically replaced by some finite value to make sense. For example, suppose empirically $p=1/2^{100}$, then the information content = 100. **The second author adds using his personal experience as a harmonium player that an occasional use of the *vivadi swars* Pa and Sudh (natural) Re in *Malkauns* in a typical way does create the atmosphere of another raga, namely, *Kaushi Kanada*.**

Two concepts of pitch stability: statistical and psychological

An empirical definition of probability implies that *probability is relative frequency in the long run*. It is striking that even in a one minute of raga *Pilu* recording [2] the probability of *Vadi* and *Samvadi swars* are remarkably stable over the three short pieces of 30 seconds each as compared to the overall performance. A recording of longer duration would perhaps also stabilize the relative frequency of the less important notes of the raga. There are four lessons to be learnt from this - (i) *it seems that the most important note in a raga is not necessarily one which is used most often (as this could well be a stay note or Nyas swar like Sudh Re in Yaman) but one whose relative frequency, apart from being high (though not necessarily highest), stabilizes faster in a short period of time*. Lesson (i) leads to lesson (ii) that *controversies which still exist among Indian musicians as to which notes should be Vadi and Samvadi in certain ragas like Bageshree and Shankara*

can perhaps be resolved now. In *Bageshree,* the *Vadi* and *Samvadi Swars* are *Sudh Ma* and *Sa* according to some experts whereas others say they are *Sudh Dha* and *Komal Ga.* Similarly in raga Shankara they should be *Sudh Ga* and *Sudh Ni* according to some while others say they should be *Pa* and *Sa* respectively. The other two lessons are that (iii) if one has a recording of longer duration it still makes sense to sample a small subset from it for analysis the advantage with a longer recording however is that there can be several such samples from different positions reflecting perhaps different moods of the performer in rendering the raga and (iv) we need to have a recording of sufficiently longer duration to get more precise values of probabilities of the less important notes (as they take some time to settle).

Remark: The three short pieces of 30 seconds each from the one minute Pilu recording correspond to the first 30 seconds, the middle 30 seconds and the last 30 seconds.

As mentioned earlier, this concept of statistical stability, proposed by the second author (who is a statistician) is different from the concept of psychological stability of Krumhansl based on note duration. Statistical stability is a good tool for detecting the *Vadi-Samvadi* scientifically in case of conflict while psychological stability detects the nyas swars or stay notes. Since the *Vadi-Samvadi* are generally also *nyas swars* we recommend Krumhansl's technique to reduce the search space and then investigate with statistical stability. For the first time in computational musicology in Hindustani music, both techniques have been tried in a recording of raga *Rageshree* in which there is a conflict over *Vadi-Samvadi* among musicians and musicologists. We quote a summary of our finding [9]:-

"The nyas swars experimentally detected using Krumhansl's theory are Sudh Ni, Sa, Sudh Ga and Sudh Re. Theoretically they should be Sudh Ga, Sudh Ni, Ma and Sa. Thus Ma does not have a good rank to claim psychological stability. Since the Vadi in Rageshree has to be a nyas swar, Ma loses the contest for being Vadi from psychological consideration of note stability. As we have already ruled it out from statistical considerations earlier, the techniques of Krumhansl and Chakraborty are both confirming the rejection of Ma as Vadi. Since Ma could not be Vadi, Sa is also rejected for being Samvadi! This means the Vadi-Samvadi can only be (Sudh Ga, Sudh Ni). And we select Ga as Vadi and Ni as Samvadi from statistical considerations as stated earlier.

We are, however, not ruling out {M, S}for all situations, this being a case study. But we do assert that, with the present performance, *the artist can only defend that musical school which supports {G, N}as Vadi-Samvadi.* This concludes our discussion."

3. Concluding remarks

We have successfully linked raga with probability. We argue that the same note will have different probabilities in different ragas depending on its absence or presence and relative importance in case of presence. The most important note statistically speaking is not necessarily the one having the highest probability but one having a high probability that is maintained consistently. This concept of statistical pitch stability is different from Krumhansl's psychological pitch stability based on note duration. Both concepts can be combined to give rise to psycho-statistical pitch stability by ranking the notes differently using the two contrasting concepts and then giving the notes the average of two ranks. Statistical stability is a good tool for detecting the *Vadi-Samvadi* statistically in case of conflict while psychological stability detects the *nyas swars* or stay notes. Since the *Vadi-Samvadi* are generally also *nyas swars* we recommend Krumhansl's technique to reduce the search space and then investigate use statistical stability, in case of conflict of views over *Vadi-Samvadi*, to scientifically detect which of the views is supported by a given recording. However, probability does not directly explain the decision process of the artist. It is nevertheless important from a listener's or an analyst's perspective [10].

Remarks (a) The term psycho-statistical pitch stability, which was first proposed by the second author in his private communication with Carol Krumhansl, has also been referred to by another name, namely, universal pitch stability. See [9] for an illustration.

(b) In his book **Music and Probability** [11], David Temperley has used the classical Bayesian approach to study music perception and cognition.

References

1. S. Chakraborty, R. Ranganayakulu, S. Chauhan, S. S. Solanki and K. Mahto, *A Statistical Analysis of Raga Ahir Bhairav*, Journal of Music and Meaning, Vol. 8, sec. 4, Winter 2009
http://www.musicandmeaning.net/issues/showArticle.php?artID=8.4

2. S. Chakraborty, S. S. Solanki, S. Roy, S. Chauhan, S. S. Tripathy and K.

Mahto, *A Statistical Approach to Modeling Indian Classical Music Performance:* arXiv:0809.3214v2 [cs.SD] [stat.AP]. **http://arxiv.org/abs/0809.3214 (2008)**

3. Castellano, M. A., Bharucha, J. J. and Krumhansl, C. L., *Tonal Hierarchies in the Music of North India*, Journal of Experimental Psychology, General, 1984, vol. 113, No. 3, 394-412

4. Krumhansl, C. L., *Cognitive Foundations of Musical Pitch*, 1990, Oxford University Press

5. S. Chakraborty, S. Tewari and G. Akhoury, *How do ragas of the same thaat that evoke contrasting emotions like joy and pathos differ in entropy?*, Philomusica online, 10, 2011, 1-10

6. D. Dutta, *Sangeet Tattwa* (*Pratham Khanda*), Brati Prakashani, 5th ed, 2006 (Bengali)

7. J. L. Snyder, *Entropy as a measure of musical style: the influence of a priori* Assumptions, *Music Theory Spectrum,* 12, 121-160, 1990

8. D. Applebaum, *Probability and information: An integrated approach*, Cambridge University Press, chapter 6, 1996

9. S. Chakraborty, R. Ranganayakulu, S. Chauhan, S. S. Solanki and K. Mahto, *Which Notes are Vadi-Samvadi in Raga Rageshree?:* arXiv:0812.0706v1[cs.SD].

    http://arXiv.org/abs/0812.0706 **(2008)**

10. S. Chakraborty, M. Kumari, S. S. Solanki, S. Chatterjee, *On What Probability can and cannot do: A case study in Raga Malkauns*, Journal of Acoustical Society of India, Vol. 36, No. 4, 2009, p. 176-180

11. D. Temperley, Music and Probability, The MIT Press, 2006

ON AMOUNT OF NOTES IN OCTAVE

Mykhaylo Khramov
Independent Project CommaTor
53-2, Plekhanova Street, Lugansk, 91005, Ukraine
commator@gmail.com

Abstract

The very brief description of a possibility of development and implantation portable unelectrical shruti-harmoniums with a redundant amount of fixed pitches in each octave. Such harmoniums are capable to promote symmetrical improvement of performance and written fixing qualities for both Western, and North Indian classical and other music

Key words: 53EDO, harmonium, microtone, music, shruti.

I. Introduction

At the beginning of 2009, September I was a guest of spouses Fluke. About 30 years they gather a collection of free reed instruments, mainly harmoniums [1]. They have granted me a photocopy of the interesting letter. Here its fragment with the exotic expression about enjoy music:

«…5th March 1947 … Long after Bach's W.T.K., music needs to get rid of temperament to be fully enjoyed.
Many thanks for letting me see your excellent article. It raises hopes of at least 7 extra notes in the octave.
Signed: G. Bernard Shaw …»

Mr. Shaw has given understanding, that the fullness of enjoy music may be expressed by an amount of the notes in an octave, which are used for its writing down and performance. From this point of view in India, for example, music is fuller enjoyed, than there, where since times of J. S. Bach is dominating the fallacy that 12 notes in an octave are quite enough for music. After all, in remarkable and the sufficiently new Indian monography about shruti [2] is written the following:

«… in the Indian system the octave is known as Saptak as it contains only seven major expressive intervals, called Swaras … the need of expression as well as appreciation of Indian Music requires smaller standard intervals. A very long period of development of Indian music (over 4 millennia) gave rise to a unique scale based on a

large number of basic microtonal intervals called shrutis. Thus a musician can choose any of the shrutis, which supports the interval … there had been various opinions about the number (66, 53) [8] of shrutis during the span of development. In recent times it seems that the number of shrutis is broadly agreed upon as 22 ... However there still exists some controversy over the numbers and the exact ratios of the shruti intervals supporting each swara.
All shrutis are by no means equal ...»

Indian researchers testify that in India for performance of classical music are necessary 22 notes in an octave, if not 53 or 66. Therefore, the Indian music is enjoyed almost in 2, and may be even in 6 times fuller that is possible in a system of 12 equal divisions of the octave (12EDO). This sentence border on a joke, but does not lose seriousness. In fact music for solution of international dialogue problems is the not last means.

Induced by unlimited and insuperable bent for music the researcher from that country, where the music education leans on a 12EDO system, is doomed to feel music enjoyment on many notes fuller. In the own investigations he/she will necessarily meet a need of study gaps between standard notes and inevitable will be interested to learn ancient and still alive Indian shrutis. Today is well known in any country that from ancient times in India for some reason there are 22 notes in an octave, instead of 12, or 24.

2. Approach to Shruti Scale Through 53edo System.

It is difficult to believe in the existence of something more authentic, than materials of the above mentioned monography on shruti: the book is written in India by the modern and competent Indian authors. And it is written so, that among known and conventional is a lot of new and useful to a solution of actual problems not only Indian music.

The monography does not give an occasion for doubts that music exists, as the psychoacoustical phenomenon, in the context of so-called just intonation (JI). Indian researchers have not depleted a JI system of own classical music by a prime limit 5, as are accustomed to do their colleagues in confined by the standard of 12EDO countries [3].

The ideal JI system is nonclosed, but may be not bad approximated in the closed 53EDO system [4]. As attractive feature of this system appears proximity its minimal microtone, or comma (22.642 ¢) to size of the minimal microtone of an Indian scale, which is known as nyuna shruti (22 ¢) [5]. Pramana shruti (70 ¢) [6] and purana shruti (90 ¢) [7] are accordingly close to sums of three (67.925 ¢) and four (90,566 ¢) commas of the 53EDO system.

TABLE I explains the order of use conventional and additional accidentals for written fixing of the 53EDO system by means of widespread Western five-linear notation.

The approximation in the 53EDO system of Pythagorean limma interval is equal to four its commas and it appears by analog of purana shruti. The interval, which in Western musical theory is known as diesis, is approximated in this system by two commas. Uplimma, or limma extended by comma, approximates Pythagorean apotome.

In the TABLE I:

- PCS ~ Pythagorean Chromatic Scale;
- 3LJI ~ 3-limit Just Intonation or JI of prime limit 3;
- 5LJI ~ 5-limit Just Intonation or JI of prime limit 5;
- *p*LJI ~ *p*-limit Just Intonation or JI of prime limit *p*.

Table I
Usable Accidentals for Approximation of JI by 53EDO System.

Dependence of Use on a JI System.			Sign		
			In Scores	In Texts	Name	Ordered Operation
*p*LJI. Are added diesis contained signs	5LJI. Are added comma multiple signs	3LJI. PCS	⪪	*x*	doublesharp	to sharp pitch by two uplimma/apotome
			♯♯	#	sharp	to sharp pitch by one uplimma/apotome
			♮	=	natural/exact	do not alter pitch
			♭	*Ь*	flat	to flat pitch by one uplimma/apotome
			♭♭	*ЬI*	doubleflat	to flat pitch by two uplimma/apotome
			}	*ʃ*	threeup[natural]	to sharp pitch by three comma
			)	)	twoup[natural]	to sharp pitch by two comma
			>	>	up[natural]	to sharp pitch by one comma
			<	<	low[natural]	to flat pitch by one comma
			(	(	twolow[natural]	to flat pitch by two comma
			{	*{*	threelow[natural]	to flat pitch by three comma
			⧣	*Ш*	hypersharp	to sharp pitch by uplimma/apotome and diesis
			⧧	*ǂ*	hyper[natural]	to sharp pitch by one diesis
			⧨	*J*	hypo[natural]	to flat pitch by one diesis
			⧩	*JЬ*	hypoflat	to flat pitch by uplimma/apotome and diesis

TABLE II will help to check up based on sonantometrical logic [8] suppositions that shruti are obliged to be compatible with spectral features of accompanying tanpura drone. It demonstrates an idea of construction of shruti sets from first six prime odd sonants (*:D[3], :M[5], :Q[7], :N[11], :R[13], :P[17]*)

and them subsonants (*:d, :m, :q, :n, :r, :p*), compatible with spectra of drones with tunings *S-P* (*c-g* in Western notation) and *S-M* (*c-f*).

Table II
Approximation by 53EDO System of Drone Compatible Shrutis.

JI System			Shruti Scale by Tanpura Strings				Shruti Scale by 53EDO System			
Order No.		Pitch in Cents	M	S	P	No.	No.	Cents		Name :S
ENS	EUS		Name :S	Name :S	Name :S			Pitch	Error	
2	1	1200,000		S' :T		22	53	1200,000	± 0,000	=c= :T
64	33	1146,727	N4? :nd				51	1154,717	+ 7,990	=cJ :nd
32	17	1095,045		N4 :p		21	48	1086,792	− 8,252	<H= :p
15	8	1088,269			N4 :MD		48	1086,792	− 1,476	<H= :MD
24	13	1061,427			N3 :rD	20	47	1064,151	+ 2,724	=HJ :rD
11	6	1049,363	N3 :Nd				46	1041,509	− 7,853	<HJ :Nd
16	9	996,090	n2 :dd			19	44	996,226	+ 0,136	=Hb :dd
7	4	968,826		n1 :Q		18	43	973,585	+ 4,759	>HJb :Q
12	7	933,129			D4 :qD	17	41	928,302	− 4,827	<A‡ :qD
5	3	884,359	D3 :Md			16	39	883,019	− 1,340	<A= :Md
64	39	857,517	d2 :rd			15	38	860,377	+ 2,860	=AJ :rd
13	8	840,528		d2 :R			37	837,736	− 2,792	<AJ :R
8	5	813,686		d1 :m		14	36	815,094	+ 1,408	>Ab :m
51	32	806,910			d1 :PD		36	815,094	+ 8,184	>Ab :PD
32	21	729,219	d1? :qd				32	724,528	− 4,691	'<G‡ :qd
3	2	701,955		P :D	P :ØD	13	31	701,887	− 0,068	=G= :D
16	11	648,682		M4? :n			29	656,604	+ 7,922	=GJ :n
17	12	603,000	M4? :Pd				27	611,321	+ 8,320	>Gb :Pd
24	17	597,000			M4 :pD	12	26	588,679	− 8,320	<F# :pD
11	8	551,318		M3 :N		11	24	543,396	− 7,922	=F‡ :N
4	3	498,045	m2 :Ød	m2 :d		10	22	498,113	+ 0,068	=F= :d
21	16	470,781			m1 :QD	9	21	475,472	+ 4,691	>FJ :QD
64	51	393,090	G4 :pd			8	17	384,906	− 8,184	<E= :pd
5	4	386,314		G4 :M			17	384,906	− 1,408	<E= :M
16	13	359,472		G3 :r		7	16	362,264	+ 2,792	=EJ :r
39	32	342,483			G3 :RD		15	339,623	− 2,860	<EJ :RD
6	5	315,641			g2 :mD	6	14	316,981	+ 1,340	>Eb :mD
7	6	266,871	g1 :Qd			5	12	271,698	+ 4,827	>EJb :Qd
8	7	231,174		R4 :q		4	10	226,415	− 4,759	<D‡ :q
9	8	203,910			R3 :DD	3	9	203,774	− 0,136	=D= :DD
12	11	150,637			r2 :nD	2	7	158,491	+ 7,853	=DJ :nD
13	12	138,573	r2 :Rd				6	135,849	− 2,724	<DJ :Rd
16	15	111,731	r1 :md			1	5	113,208	+ 1,476	>Db :md
17	16	104,955		r1 :P			5	113,208	+ 8,252	>Db :P
33	32	53,273			r1? :ND		2	45,283	− 7,990	=C‡ :ND
1	1	0,000	S :Dd	S :Ø	S :dD	0	0	0,000	± 0,000	=C= :Ø

From the table it is clear the shruti sets for drones S-M and S-P do not coincide. The sufficient redundancy of grades of the approximating 53EDO system allows to optimize and to fix in writing individual shruti sets for each performed piece.

Appendix section contains a drawing of a possible embodiment of TABLE II as portable button harmonium, designed for a playing manner both with vertical, and with horizontal position of the keyboard. Ten-row keyboard is represented by two coupled five-row sets of buttons, that allows to mark one set with clear to Western performer symbols and colourings. Other set may be marked so that to facilitate acquisition of skills of playing for the Indian musician. Certainly, it will make sense only for the initial performers, as the advanced ones tend to play, not looking on buttons. They without impediments can use all advantages of compact ten-row¸ generalized keyboard allowing to not change customary fingering for playing in any tonality. Tonalities is possible to transpose with a minimal step on nyuna shruti (comma) and thus precisely to find most favorable for each performance tessitura.

Acknowledgements

References

1. Pam & Phil Fluke, 6, Albert Terrace, Saltaire West Yorkshire BD18 4PS, UK https://sites.google.com/site/commator/flukes

2. Datta, A. K., Sengupta, R., Dey, N. and Nag, D., *Experimental Analysis of*

Shrutis from Performances in Hindustani Music, pp. 18-19, Kolkata: SRD ITC SRA, 2006
http://www.itcsra.org/sra_story/sra_story_research/sra_story_resrch_l inks/sra_story_resrch_pubs/sra_story_resrch_pubs_index.html

3. R. Asmussen, "Abstract". *Periodicity of sinusoidal frequencies as a basis for the analysis of Baroque and Classical harmony: a computer based study*, Leeds: The University, School of Music, September, 2001.
http://www.terraworld.net/c-jasmussen/

4. P. Barbieri, *Enharmonic Instruments and Music 1470 1900*, p. 96, Latina: Il Levante Libreria Editrice, 2008. ISBN: 978-88-95203-14-0. http://www.patriziobarbieri.it/1.htm

5. Datta, A. K., Sengupta, R., Dey, N. and Nag, D.*Experimental Analysis of Shrutis from Performances in Hindustani Music*, p. 28, Kolkata: SRD ITC SRA, 2006   http://www.itcsra.org/sra_story_research/sra_story_resrch_ links/sra_story_resrch_pubs/sra_story_resrch_pubs_index.html

6. Ibidem

7. Ibidem

8. M. Khramov, "Approximation to 7-limit Just Intonation in a Scale of 31EDO", *Proceedings of the International Symposium Frontiers of Research on Speech and Music (FRSM-2009)*, pp. 73-82, Kolkata, 2009

Appendix
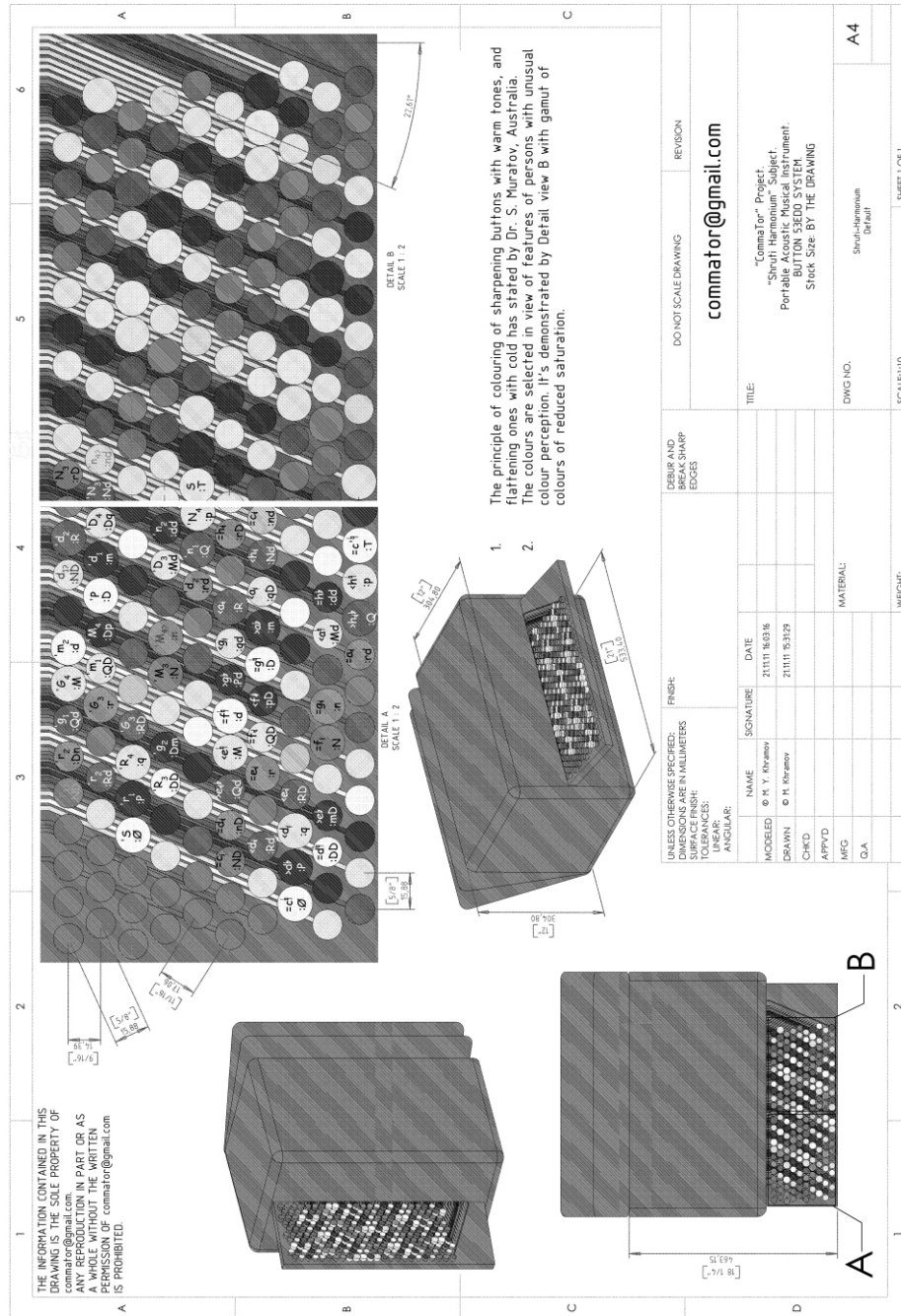
AUTOMATIC TONIC (SA) DETECTION ALGORITHM IN HINDUSTANI VOCAL MUSIC

Ranjan Sengupta[1], Nityananda Dey[2] and Asoke Kumar Datta[3]

[1]Sir C V Raman Centre for Physics and Music, Jadavpur University, Kolkata 700 032

sgranjan@gmail.com

[2]ITC Sangeet Research Academy, 1, N S C Bose Road, Tollygunge, Kolkata 700 040

[3]BOM-Public Charitable Trust, 3/3 Girish Ghosh Street, Kolkata

Abstract

Analysis of Indian classical singing performances requires notes to be detected from the extracted pitch profile of the songs. This in turn needs the knowledge of the tonic used by the singer. This is not directly obtainable and if the service of a musician is exploited the correctness would depend on the expertise of the musician. Also such a service may not be readily available. It therefore makes sense to put efforts on finding an algorithmic approach to the detection of the tonic (Sa). An expert musician can often identify the base tonic (Sa) in a singing. However how they do it is an inexplicit knowledge. It may, therefore, be presumed this to be some sort of error-feedback mechanism.

Here in this paper we have presented an algorithm for the automatic extraction of the tonic (Sa) from the steady state pitch sequences of the actual song sung by the expert musicians. The data-base consist 118 aalap performances covering four ragas by 34 male and 7 female well known singers. As there is confusion on the actual ratio system used by Indian performers, we have tested in total 3 ratio system (12 note system) has been experimented with. The detected Sa were found to be in good agreement with the measured values.

Keywords: Tonic, Indian classical vocal music, Pitch, steady state, Note

Introduction

Musicological research has long extensive existence since ancient times. However, scientific investigations in Indian music have fallen far behind those in western music. The present state of science and technology can provide ample scope for quantitative as well as qualitative approach to investigate srutis, swaras, intervals, octaves (*saptak*), consonance (*vaditya*), musical quality, rhythm etc. [1] The power available with the computing technology and the advancement in the area of artificial intelligence, approximate reasoning and natural language processing may allow us to understand the higher level tasks

like composition, improvisation and aesthetics associated with classical music.

Musical scales in India pose a confusing picture. Much of the historical commentary seems baffling with unending arithmetical calculations of ratios with little experimental support [2]. Hindustani musicians of recent days have for their own convenience taken only twelve Swarasthanas (notes) in the present system of music: namely, seven Suddha and five Vikrta Swaras. The seven suddha swaras are Sadja, Risava, Gandhara, Madhyama, Panchama, Dhaivata and Nisada.. In between these, are placed five Vikrta swaras, making a total of twelve. These are Komal Risava, Komal Gandhara, Tivra Madhyama, Komal Dhaivata and Komal Nisada. Sadja, which in short is known as Sa, and Panchama, which is known as Pa, are fixed and unchangeable notes, i.e., "achala swaras".

Analysis of musical structure of a composition with respect to its influence on human mind is an interesting topic. For this purpose one may follow the paradigm of similar research in language. The analysis of syntax and semantics play an important role in it. Drawing on this similarity one can see that that role played by the notes in a music composition plays the roles letters play in languages. In Indian classical singing syntax may be understood only in terms of notes, sometimes probably srutis, not absolute values of pitch. The extraction of notes from the pitch profile needs the knowledge of the tonic used by the singer. This is not directly obtainable and if the service of a musician is exploited the correctness would depend on the expertise of the musician. Also such a service may not be readily available. It therefore makes sense to put efforts on finding an algorithmic approach to the detection of the tonic (Sa).

An expert musician can often identify the base tonic (Sa) in a singing. However how they do it is an inexplicit knowledge. Presumably he/she tries to imitate the note sequences in his/her own voice either loudly or silently before deciding which one may be the tonic. It may, therefore, be presumed this to be some sort of error-feedback mechanism. Consequently one can use some error minimizing technique for the purpose. The basic approach taken in this study is to try out a large number of pitch values in a given range as possible tonics and to find the deviations of the actual steady states from the closest note predicted by the selected tonic. It is reasonable to assume that for an incorrect trial tonic the sum of these deviations would be large, since the intervals are known to be non-uniform. In fact, the sum of these deviations for each trial tonic is a seen to be a function of the value of the trial tonic. As we shall see later this function is generally a continuous curve in the given range. We assume that the minimum in this function will indicate the actual tonic.

The data-base consist 118 alap performances covering four ragas by 34 male and 7 female well known singers. As there is confusion on the actual ratio system used by Indian performers, in total 3 ratio systems has been experimented with. The detected Sa were found to be in good agreement with the measured values.

Method of Analysis

A. Pitch Period Extraction from Signal
A method based on Phase-Space Analysis (PSA) was used for extracting pitch periods. PSA [3] uses the basic fact that, for a repetitive function, two points of the signal having a phase difference of 2 have the same displacement and, therefore, if plotted in a 2-D phase diagram would be a straight line with a slope of /4. For a quasi-periodic signal this diagram would be very close to this straight line. For all other phase angles the curves would be wider loops. In a phase-space diagram the points representing such pairs would be lying on a straight line with a slope of /4 to the axis. In case of a quasi-periodic signal, viz., speech or music these points would lie in close very flattened loop with the same axis. As the phase between two such points increases from 0 to 2 the corresponding phase-space diagram broadens from this flat loop and forms close broader loops. It can also be shown that the root mean square value of the deviation is same as the absolute value of the difference between displacements of the two points. For quasi-periodic signals the minima in the sum of deviations occurring at a phase difference of 2n, where n is a small integer, shows comparable small minima values, usually a little larger than that for n=1. A study of a large volume of signal reveals that (a) the sum of deviations last at phase differences other than 2n, are always large and never fall below minimum deviations plus 0.2(max min) of deviations and (b) that deviations for phase differences 2n, n1, never falls below that for 2 when this value exceeds the aforesaid threshold. This allows a clear threshold and algorithmic logic for pitch extraction. Henceforth the pitch pattern files, extracted using the above method, will be referred to as 'cep file'.

B. Smoothing
The Cep file contains pitch extracted only in the quasi-periodic region of the signal. The PD algorithm above uses a predefined specific range for possible pitch values e.g., in the present case the default range is fixed between 70 Hz to 700 Hz. Three types of error were observed. One is that the determined pitch value is approximately equal to half or double of the actual pitch value. This error is typical for PSA and is usually a rare and isolate occurrence. The other is pitch occurring out side the range. Third one is spike in pitch sequence. The smoothing operation for the first type of error is simply detection of it and

doubling or halving the value as required. The two other types of error are detected through an examination of local pitch values and the erroneous values are replaced using linear interpolation from the neighbouring valid data.

C. Steady State Detection

Even in a perfectly perceptible steady note all the pitch values are never identical. There are always some involuntary small variations in pitch, which has no relevance in the perception of the corresponding value of the note. The definition of steady state, therefore, is not a trivial problem. We define a steady state in the pitch file as that subsequence of pitch data where all the pitch values lies within a predetermined band around the mean value. Furthermore as the signal considered in the present study is taken from aalap portion of the total singing we restricted our steady states to minimum duration of 60 milliseconds.

D. Tonic Detection Algorithm

We have used an error minimizing technique for the purpose. The basic approach is to try out a large number of pitch values in a given range as a possible tonic and to find the deviations of the actual steady states from the closest note predicted by the selected tonic. It is reasonable to assume that for an incorrect trial tonic the sum of these deviations would be large. In fact, the sum of these deviations for each trial tonic is a function of the value of the trial tonic. As we shall see later this function is generally a continuous curve in the given range. We assume that the minimum in this function will indicate the actual tonic.

Experimental Details

Forty one singers (Thirty four male and seven female) of Hindustani music were asked to render four ragas namely Bhairav (That Bhairav), Darbari Kannada (That Asavari), Mian-ki-Malhar (That Kafi) and Todi (That Todi), which included aalap, vistar, taan and gamaka. In most cases the notes extended on both sides of the middle octave. The F_0 range for male and female singers was 55Hz to 600Hz and 100Hz to 800Hz respectively. Direct digital recording was done in a noise proof studio having a reverberation time of 0.1 sec. via standard sound card (full Duplex PnP) in a PC to avoid possible phase distortions present in most of the analog recording [4]. The digitisation of the signal was done at the rate of 22050 samples /sec (16 bits/sample). Only the voice of the singer was recorded and the accompanying instruments were excluded. For our analysis only the aalap part of each singer was selected from each raga. Pieces of aalap for each singer for a raga were taken out from the complete aalap deleting the bandish part. These constituted the aalap signal files for a singer for each raga. For each raga a singer had one aalap signal of ~2 to 3 minutes. Total 118 aalap signal files (97 for male and 21 for female) were

thus selected for analysis, which constituted our database.

Pitch periods were extracted by using a PD algorithm based on Phase-Space analysis [3]. In the present study our requirement was to determine time periods for individual glottal cycle in the continuum of a quasi-periodic speech signal. This process on a wav file produces a .cep file, which contains pitch values and the corresponding time in the wav file. As already mentioned this file may contain erroneous pitch values, which are subsequently removed through the smoothing operation outlined in the last section. Even after all these corrections there may still be some errors which reveals themselves in the contour as sharp local fluctuations. We presume these to be irrelevant in the perception of tonality or pitch movement. To remove these we simply replace the $(i+1)^{th}$ pitch x_{i+1} by the ith pitch x_i when $| (x_{i+1} \ x_i) | > x_i *0.1 \ \& \ x_{i+1}>0$. This creates a p-file. From p-file steady state sequences are created with all consecutive pitch in a sequence, which is terminated when $| x_{i+1} \ M | > M/30$ where $M = (1/i) \ x_i$. If the duration of any sequence be less than that of a certain minimum value (60 msec. for this experiment) then the sequence is rejected. These create total 118 .std files, which constitutes our database for analysis.

For extraction of tonic two separate ranges are selected 95Hz 175Hz and 185Hz 255Hz respectively for male and female singers. Within these ranges each pitch value at an interval of 0.01 Hz was tried as a tentative tonic. For each of such tentative tonic corresponding note intervals are constructed using each of the 3 ratio systems having 12 ratios each, they are i) equi-tempered (ET),[5] ii) ratio system compiled by Lenz and Danielou (W)[5] and iii) new ratio system extracted from analyzing [6]actual songs (NS). NS is constructed by manually measuring the 12 peaks as shown in the figure 1, which depicts the frequency distribution of the duration of pitch (folded to the middle octave). Table 1 presents the above stated peak values.
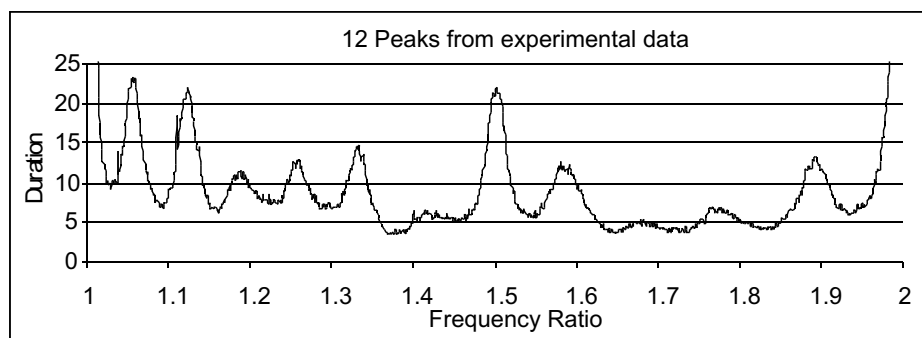

Figure 1. Distribution of the duration in 1200 bin for all signal data pooled together

Sa	re	Re	ga	Ga	Ma	MA	Pa	dha	Dha	ni	Ni
1	1.054	1.125	1.186	1.253	1.333	1.415	1.5	1.58	1.691	1.777	1.893

Table 1. Frequency ratios of the 12 peaks in figure 1 henceforth referred as new ratio system (NS)

The position of each of the steady states was determined in relation to the note structure. The deviation of each of the steady states from the closest ratio values is then found out. The sum of these deviations constitutes the error corresponding to the tentative tonic. The errors for all values of the tentative tonic are then calculated. This constitutes the error sequence. Two types of error were calculated. The first one is done without using restricted grammar and the other using restricted grammar. For each type, two different kinds of error were calculated. One is plain error and the other is the weighted error (weighted by the duration of the pitch). For each of these kinds two different methods were utilized for finding the tonic. One corresponds to least error in the error sequence (Method 1) and the other one first assumes the local minima in the error sequence as valid notes then calculate the least error corresponding to these minima (Method 2). Thus total eight values of the tonic for each of the signal files were obtained. The deviations of these eight tonics from the actual measured tonics were then found out. Average value of the deviations were calculated ratio system-wise (3 ratio systems) and sex-wise (male and female). Table 2 presents these values.

Figure 2 below presents one example each from a male and a female singer. In the legend RWL_NS represent the curve for weighted error using grammar and WWL_NS that without using grammar for method 1 in new scale. The left hand figures show only a part of the std-file and the right hand ones show the error profiles for the ratio system ns and weighted error. It may be seen that the pitch profile rarely presents a smooth steady length of region. The curvilinear rising and falling portion represent glissandos adjoining rendered notes. The zeroes generally indicate short interruption due to consonantal obstructions. The x-axis in the error profile is the value of the trial tonic and y-axis gives the corresponding error. The error profiles happen to be relatively smooth oscillating curves with prominent extrema.
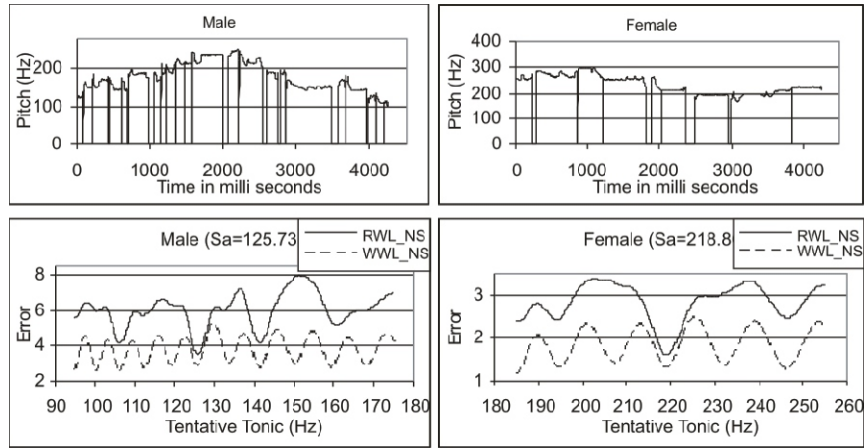
Figure 2. Example of Pitch contour and error profile in tonic extraction

Results and Discussions

Table 2 presents averages of absolute deviations of tonic detected using all the 8 procedures, as outlined in earlier. As is seen, the data is pooled sex-wise (female and male) and ratio system-wise (3 ratio systems, as discussed earlier). The figures in the brackets indicates the count of corresponding signal data files after rejecting those showing absolute deviation of error greater than sum of mean and S.D. in the respective category. Consequently, the data in the table is actually the modified average after rejecting the above-stated outliers. Average in the last row takes into account of these counts.

As is observed, weighted error using restricted grammar gives the best results. Although the ratio system NS is the best, others give very close results. Method 1, which uses the least error, gives appreciably better results compared to Method 2.

Sex		Without Grammar				Using Grammar			
		Plain Error		Weighted Error		Plain Error		Weighted Error	
		Method 1	Method 2	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
F	ET	11.89(17)	16.18(18)	11.85(18)	15.52(18)	0.66(20)	13.97(17)	0.66(20)	13.35(17)
	W	9.24(17)	7.55(17)	9.87(16)	8.27(18)	0.78(20)	9.37(17)	0.75(20)	7.84(17)
	NS	11.29(18)	7.94(18)	11.29(18)	6.95(17)	0.59(20)	11.11(17)	**0.58**(20)	11.21(17)
M	ET	12.51(77)	19.58(80)	12.08(76)	19.36(80)	2.27(84)	16.13(80)	2.46(81)	16.15(79)
	W	15.68(79)	11.15(78)	15.44(76)	11.29(78)	2.72(80)	15.64(81)	2.58(78)	12.61(78)
	NS	14.37(76)	19.88(79)	14.65(78)	20.13(82)	2.37(83)	17.22(81)	**2.23**(80)	17.39(81)
Average		13.58(284)	15.75(290)	13.5(282)	15.79(293)	2.11(307)	15.49(293)	**2.07**(299)	14.6(289)

Table 2. : Average error (in Hz) of tonic detected

Reference

1. A. K. Datta et.al., "On Scientific Approaches to the Study of Vaditya in Indian Music, Proc. Fifth Int. Workshop on Recent Trends in Speech, Music & Allied Signal Processing, Thiruvananthapuram, India, December, 2000.
   A. K. Datta, et.al , "Pitch Analysis of Recorded Vocal Performances in Hindustani Music: Evidence of a Personal Scale", J. Acoust. Soc. Ind., XXV, VI-1.1, 1997.
   A.K. Datta, et.al ,"Multiple States in a Note in Hindustani Music and their Relevance to Consonance", J. Acoust. Soc. Ind., XXVI (3 - 4), 394, 1998.
   R. Sengupta, "Study on some Aspects of the Singer's Formant in North Indian Classical Singing", J. Voice, 4 (2), 129, Raven Press, NYC, 1990.
   A. K. Datta et.al, "Relevance of Consonance in Indian Musical Scale: Theory and Practice", J.Acoust. Soc. Ind., XXIII, 110, 1995.]
   A. K. Datta, et.al, "Studies on Identification of Raga using Short Pieces of Taan : A Signal Processing Approach", Proc. Sixth Int. Workshop IWSMSP, New Delhi, India, 19-21 December, 2001.
   A K Datta et.al, "Sruti usage by old and contemporary singers in Khyal: An objective approach", Proc. FRSM-2004, Annamalai University, Tamilnadu, January 8-9, 2004.
   A K Datta et.al, "Evaluation of srutis in Hindustani Music from recorded performances", Proc. FRSM- 2005, Utkal University, Bhubaneswar, January 6-7,2005.

2. Sarangadeva,   "Sangeet Ratnakar with Comments of Kalinath", Anandashram, edition, 1987.
   Matanga ,"Brihaddesi", Samvasiva Sastri(ed), Trivandram Sanskrit Series, 1928 Bhatkhande Pt. "Hindusthani Sangeet Padhati", Bombay, 1937

3. A. K. Datta, "Generation of Musical Notations from Song using State-Phase for Pitch Detection Algorithm", J. Acoust. Soc. India, Vol. XXIV, 1996.

4. J Laver et al, "Acoustic Waveform Perturbations and Voice Disorders", J. Voice, 6(2), 1992.

5. A K Datta et.al, "Study of srutis in Indian musical scales and relevance of schools in recent times using signal processing techniques", Proc. FRSM-2003, IIT Kanpur, Kanpur, February 15-16, 2003

6. A K Datta et.al, "Objective analysis of Srutis from the vocal performances of Hindustani Music using Clustering Algorithm" Published in EUNOMIOS An open online journal for Theory, Analysis and Semiotics of Music, http://www.eunomios.org/, dt. 25.6.2011

EFFECT OF DIFFERENT LEVELS OF TRAINING ON SINGING POWER RATIO AND SINGER'S FORMANT IN CLASSICAL CARNATIC SINGERS

Supritha Aithal, Swathi S & Rajasudhakar R
Dept. of Speech-Language Sciences
All India Institute of Speech & Hearing, Mysore-06.
supritha.aithal@gmail.com and swathi.subbanna@gmail.com

Abstract

This paper presents the two parameters such as Singing power ratio (SPR) and Singer's formant (Fs) for objectively evaluating singing talent/quality among three levels of singers. Thirty singers' (10 junior level, 10 senior level and 10 vidwath level) sung sample of 'Lambodara' song was analysed using LTAS of CSL 4500 software to extract SPR. The sample was sung at three conditions including (1) singing without background music, (2) singing with background music at listening comfortable level and (3) singing with background music at more than listening comfortable level. SPR and presence or absence of F_s was analysed among the singers at three conditions. No interaction between conditions and levels of singing was found for SPR. Fs was not observed at all the three conditions among three levels of singers. There was a gradual increment in the SPR value with the increase in the years of training. Thus, SPR would be an objective tool to measure the singing voice quality in Carnatic singers.

Keywords: *Singer's formant, Singing power ratio, Acoustic analysis, Trained singers, Carnatic singers*

Introduction

"The human voice is extraordinary. It is capable of conveying not only complex thoughts, but also subtle emotion. In an instant, it can communicate the terror of a scream or the beauty of a song"- explained by Sataloff (2005) about the specialty of human voice. Voice is not only important for human communication but also serves as a primary musical instrument. Anyone who uses their voice in order to carry out their livelihood is considered as professional voice users. One among them is singers and they are referred to as elite vocal performers.

Singing is a unique form of art which frequently requires years of training to

attain the utmost form of performance. Even for the trained ears it is difficult to judge the singing talent. Each singer has a natural, unique and clearly identifiable voice quality that listeners may describe in their own way. For example, one listener may describe a singer's voice as warm voice and the other person may call it matured voice or rich voice. However, all these terms do not quantify or objectively measure the singing voice. With the vastly developing technology, attempts have been made to find which feature of singing voice makes the listener to feel the voice as matured, warmth or rich. Singer's Formant (FS) and Singing Power Ratio (SPR) are among such measures which are used to measure voice quality acoustically (Omori, Kacker, Carroll, Riley and Blaugrund, 1996; and Lundy, Roy, Casiano, Xue and Evans, 2000).

Omori et al., (1996) introduced a parameter called singing power ratio (SPR). They measured SPR by considering the ratio between the greatest harmonic peak between 2 and 4 kHz and the greatest harmonic peak between 0 and 2 kHz. The authors calculated SPR for 37 singers and 20 non singers and concluded that, SPR represented the resonant quality of the singing voice, provided a quantitative measure for evaluating singing voice quality and it has a distinctive correlation with period of voice training. Watts, Barnes-Burroughs, Estis and Blanton (2006) found significant differences in SPR values for voices of untrained talented and non talented singers which suggested the vocal tract resonance and its effect on perceived vocal timbre or quality may be an important variable related to the perception of singing talent.

Kenny and Mitchell (2007) assessed the relationship between acoustic measurement and perceptual judgment in the identification of perceptually preferred voices and comparing the sound quality in voices using and not using the open throat technique. They compared perceptual rankings of vocal quality of expert pedagogues with rankings of acoustic measures (SPR and Energy Ratio) to assess whether these acoustic measurements matched the perceptual judgments of vocal quality. Although they found the expected significant relationship between SPR and ER, there was no relationship between perceptual ratings of vocal samples of singers based on SPR or ER. They concluded that LTAS measures are not consistent with perceptual ratings of vocal quality; such measurements cannot define voice quality and suggested future research with LTAS to address vocal quality by considering alternative measures that are more sensitive to subtle differences in vocal parameters.

Lundy et al., (2000) studied the acoustic analysis of the singing and speaking voice among singing students. Both singing and speaking voice samples were recorded from 55 singing students and analyzed for SPR and standard measures of acoustic analysis. The authors found that SPR values did not differentiate

sung versus spoken samples. The authors did not find any gender difference on SPR and no influence of years of training on SPR.

The singer's formant (FS) is described as an increase in the signal intensity between the third and fourth formants, allowing the singer to be heard without amplification over sounds of accompanying music (Bartholomew, 1934). He stated that a good operatic voice needs a concentration of energy around 3 kHz. He also mentioned that this concentration must be produced with a special resonator in the larynx or lower pharynx. Sundberg (1987) also reported the main acoustical contribution to the generation of the singer's formant which stems from a cluster of third, fourth and fifth formants. Figure 1 illustrates the singer's formant seen as a hump (acoustic peak energy) at around 2-3 kHz region for western singers.
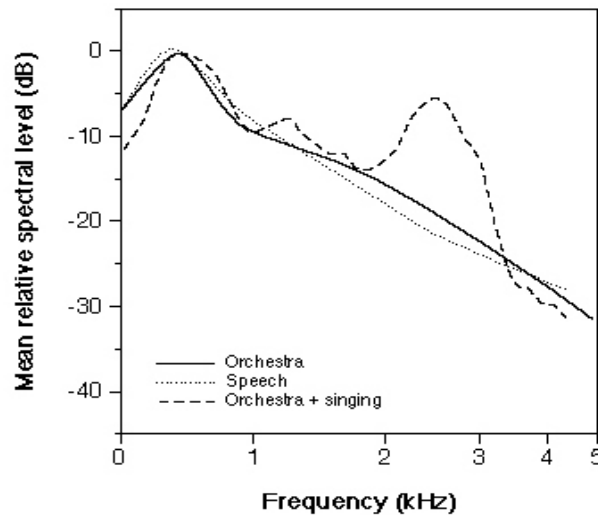
Figure 1: Singer's Formant (FS) (Taken from Sundberg, 1987)

Morris and Wiess (1997) said that the precise location of Fs varies. It might be affected by individual voice types and ranges, the vowels attempted, the pitch and the amplitude produced. All these above mentioned studies have established that FS and SPR are good parameters to judge singing quality in the western context. It is known that Carnatic music is one of the classical systems of music most prevalent in southern part of India that differs from the western music in many aspects. That is, musical progression is in terms of single note i.e. at a given time only one note or its shadow is acting. Thus, it is called homophonic or sometimes just as melodic music. On the other hand

western music is heterophonic or harmonic system as the progression is in harmony and calls for several sounds simultaneously. Hence, there is a need for objective measures like SPR and FS to quantify singing talent in Carnatic singers. Few studies were conducted to see the usefulness of SPR and singer's formant in Indian classical music.

Sujatha (1989) analysed the singing voice of 10 trained (4-10 yrs of training) and 10 untrained singers and found that, most of the trained singers had more energy concentration between the frequencies 2500 to 2969 Hz. Whereas, the energy concentration was more in 2300- 2500 Hz region in untrained singers. Sengupta (1990) studied some aspects of singer's formant (Fs) in North Indian Classical singing. Eight singers (4 males and 4 females) served as the subjects. The author found singer's formant and the centre frequency of Fs increased with raising pitch. The bandwidth was also found to increase with increase in fundamental frequency as like in western singers. Chayadevi (2003) compared Singer's formant between Carnatic music and Hindustani music. The author analysed the sung samples of /a/ vowel in twenty Carnatic and twenty Hindustani vocalists who had minimum 10 years of formal training. The results of the study indicated no significant difference in the parameters measured across two styles of singing except for the bandwidth of Fs which showed significant difference between male singers of two styles of singing. The author observed a shift in the centre frequency of Fs in both the styles of singing. Also, clustering of higher formants was found in the region noted as Fs.

Boominathan (2004) and Mohan (2010) did not find singer's formant in Carnatic classical singers. They justified by explaining that the Indian music is homophonic and accompaniments usually shadow the singer and hence singer would not require projecting his or her voice over an orchestra. This result opposed the earlier studies (Sujatha, 1989; Sengupta, 1990; and Chayadevi, 2003). Mohan (2007) evaluated the usefulness of SPR as an objective measure of singing voice quality in untrained and trained singers, in Indian context. The author considered two groups. Group I consisted of 10 female Carnatic singers who had minimum 6 years of training (mean age: 9.9 years) and group II consisted of 10 female singers who did not receive Carnatic singing training. The participants were asked to sing the Indian national anthem and the sample was analyzed using Vaghmi software. LTAS was extracted on the sung sample and SPR was measured. The author did not find any significant difference between the two groups on SPR. The author concluded that SPR may not be helpful in evaluating the progress of a

singer's training towards the development of a perceptually rich vocal quality in Carnatic singing.

Sujatha (1989), Sengupta (1990) and Chayadevi (2003) reported the presence of singer's formant, whereas studies by Boominathan (2004) and Mohan (2010) reported the absence of it. These studies reported contradictory results and far from conclusive about the singer's formant in Carnatic singers. The present study made an attempt to simulate the heterophonic condition and calls for accompanying music simultaneously like western music and tried to address the effect of this background music at different levels on SPR and Fs across different levels of trained singers.

Aim of The Study
The aims of the study were two-folded:

1. To investigate the effect of different levels of training on singing power ratio (SPR) and singer's formant.
2. To determine the effect of background music on singing power ratio (SPR) and singer's formant across different levels of singers.

Method

Participants: Thirty singers (7 males and 23 females) participated in this study. They were further divided into three groups based on the levels of training in Carnatic music. Group I consisted of 10 junior grade Carnatic classical singers (5 males and 5 females) within the age range of 9-12 years, who had at least 3 years of training. Ten senior grade singers (1 male and 9 females) constituted group II, who were within the age range of 12-21 years, having minimum 6 years of training. Group III consisted of 10 singers of vidwath level (1 male and 9 females) within the age range of 20-35 years, who had minimum 15 years of training. All the participants had normal speech, language, hearing and communication skills and free from any upper respiratory infections at the time of the study.

Procedure: Participants were explained about the purpose of the study and written consent was taken from them. The stimuli consisted of song of Carnatic music i.e. Pillari Geethe Lambodara of Malahari raga and Roopaka tala (Raga is the mode or melodic formulae and tala is the rhythmic cycles). All the participants were seated comfortably in a noise free room. The song 'Lambodara...' was played to the participants through the Intel head phones to familiarize them with the song. Singers were given trials to rehearse the song

along with the background music (maximum number of trials being 3) before recording. The participants were instructed to sing the song 'Lambodara...' in their best singing voice, assuming that they were singing in a concert with the listeners seated at a distance of 20 feet. Recording was done in three different conditions. That is, the participants were asked to sing the song 'Lambodara...' without background music (Condition 1), sing along with background music (Condition 2) at their comfortable listening loudness level, and sing along with background music played at slightly higher than their comfortable loudness level (Condition 3). Background music was extracted from Karaoke software and played to participants through headphones (at ear level) for condition 2 and 3. Olympus (WUS-550M) digital voice recorder was used for recording the singing sample at a sampling rate of 44.1 kHz. A distance of 8 to 10 cm was maintained between the voice recorder and mouth of the participants.

Analysis: The recorded samples were transferred to the computer and down sampled to 8 kHz using Adobe Audition 1.0 software. The second stanza of the song was considered for the acoustic analysis. Computerised Speech Lab (CSL 4500 model) from KayPENTAX, New Jersy, USA was used for the purpose. The duration of the selected sample was 16 seconds. SPR was extracted from Long Term Spectrum Analysis (LTAS) using Hamming window. Energy peaks between 0-2 kHz and 2-4 kHz were measured from LTAS and SPR was calculated by subtracting the amplitude of the strongest peak between 2-4 kHz from 0-2 kHz. Singer's formant was evaluated through visual inspection of acoustic spectrum by considering a boost of energy peak between 2-4 kHz. Presence or absence of the singer's formant was noted across three groups of singers and across three conditions. The data obtained were compared between different levels of singers and across conditions and further, subjected to statistical analysis using SPSS 17 software.

Results

The results of the study are discussed under three sub headings; (i) SPR, (ii) Fs and (iii) Level of background music.

i)   **Singing Power Ratio (SPR):** The mean and standard deviation of SPR was calculated for the three groups of singers using descriptive statistics. Table 1 shows the mean and standard deviation of SPR values for three groups of singers across different conditions.

Table 1: Mean and S.D for SPR values in dB for three different conditions

Conditions	Condition 1 (WBM)	Condition 2 (BM-CL)	Condition 3 (BM-MCL)
	Mean (S.D)	Mean (S.D)	Mean (S.D)
Group I (Junior)	-11.28 (2.90)	-10.00(2.52)	-9.87(3.44)
Group II (Senior)	-6.34 (3.08)	-7.10(2.10)	-7.17 (2.92)
Group III (Vidwath)	-3.67(2.39)	-5.17 (2.68)	-4.06(2.22)

[WBM Without background music , BM CL Background Music at Comfortable level , BM-MCL-Background Music more than comfortable level]

The mean SPR score was least for junior grade singers and highest for vidwath grade singers in all three conditions. The overall mean SPR across three conditions didn't show much difference. Figure 2 illustrates the mean SPR values for three groups of singers across three conditions.
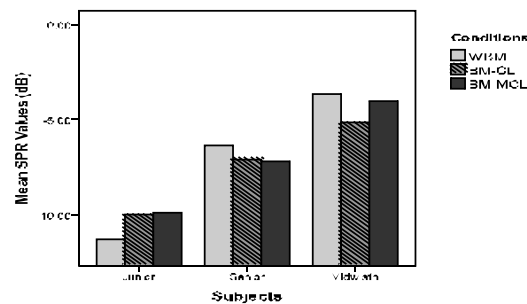
[WBM Without Background Music , BM CL Background Music at Comfortable level , BM MCL Background Music more than Comfortable Level]

Figure 2: Mean SPR for three groups of singers across three different conditions.

Mixed ANOVA was done for conditions as within subject variable and grade as between subject factor which revealed no statistically significant difference across the conditions and no interaction between conditions and grade. But, grade wise significant differences was noticed (F (2, 27) = 20.33, p<0.001). To determine the grade wise significant difference, Duncan's post hoc test was administered. Duncan's post-hoc results showed that the three grades of singers were different from each other in condition 1. Table 2 shows the results of Duncan's test for condition 1. That is, the three levels of singers (junior, senior and vidwath) are different from each other on SPR in condition 1 (without background music).

Table 2: Duncan's test result for condition 1

Groups	N	Sub-set		
		1	2	3
Group I	10	+		
Group II	10		+	
Group III	10			+

('+' in different column indicate statistical significant difference between the groups)

Table 3: Duncan's test result for condition 2

Groups	N	Sub-set	
		1	2
Group I	10	+	
Group II	10		+
Group III	10		+

('+' in different column indicate statistical significant difference between the groups and '+' in same column indicate no statistical difference)

Table 4: Duncan's test result for condition 3

Groups	N	Sub-set		
		1	2	3
Group I	10	+		
Group II	10		+	
Group III	10			+

('+' in different column indicate statistical significant difference between the groups)

Table 3 shows the results of Duncan's test for condition 2. That is, junior level singers are significantly different from other two (senior and vidwath) group of singers on SPR and among senior and vidwath, there is no significant difference found in condition 2 (background music at comfortable level). Table 4 shows the results of Duncan's test for condition 3. That is, the three levels of singers (junior, senior and vidwath) are different from each other on SPR in condition 3 (background music more than comfortable level).

(i) Singer's Formant: On visual inspection, singer's formant was not prominent in all three conditions in all the three groups. There was an overall increase in the amplitude of the acoustic spectrum when the

loudness of the background music was increased but no effect on singer's formant was found.

(ii) Level of background music: It can be inferred from figure 2 that the SPR value decreased from condition 1 to condition 3 in junior level singers. The SPR value increased from condition 1 to condition 3 in senior level singers and as such clear trend was not noticed in vidwath grade singers.

Discussion

In classical singing training one of the fundamental goals would be developing good singing voice quality by improving the resonance. An objective tool to quantify this quality which can be used as a tool to measure the progress would be of great use. This study compared the SPR, which represents the rich resonant quality of singing voice, across 3 levels of Carnatic vocalists. There was a gradual increase in the SPR values as the number of years of training increased. This clear trend represents a training effect in the development of good singing voice. The results are in agreement with the study done by Omori et al., (1990) who found that SPR of sung /a/ was significantly greater in singers when compared to non- singers. The results of this study is not in consonance with previous investigation done by Mohan (2007) who reported no significant difference between trained versus non-trained singers on SPR values. The difference in the findings can be attributed to the difference in the methodology, study population, song employed, years of training in Carnatic music and instrumentation variations.

SPR was different in all the three levels of singers in condition 1 and condition 3. In condition 2, the group II singers (seniors) performed similarly as group III (vidwath singers). That is, SPR value differentiates three levels of singers without background music (condition 1) and background music more than comfortable level (condition 3). In condition 2, junior singers are different from other two levels of singers and there is no significant difference between senior and vidwath singers based on SPR values. The obtained result is beyond the assumptions of the present study and the reason behind this remains unsolved.

The other parameter investigated in this study was singer's formant across three different conditions to simulate a condition which is similar to western context where the singers sung in presence of loud background accompaniment. Through visual investigation, it was found that singer's formant was not prominent even in condition 2 and 3 as expected. The results are not in agreement with the earlier findings of Chayadevi (2003), Sengupta (1990) and Sujatha (1989) who reported presence of singer's formant and

support the findings of Boominathan (2004) and Mohan (2010) who reported absence of singer's formant. It can be attributed to the fact that the Carnatic vocalists are not trained/learnt to sing unlike western singers against loud accompaniment, as Carnatic music is melodious and homophonic.

Overall, there was increase in the amplitude of the spectrum when the loudness of the background music was increased. That is, there was increase in the amplitude of the spectrum in both 0-2 kHz range as well as in 2-4 kHz range. This overall increase in the spectral amplitude would be because of Lombard effect.

Conclusions

Vidwath level singers (group III) had the highest SPR values when compared to senior level singers. Also, junior level singers (group I) had the least SPR values among the other two groups. This indicates, longer the training, higher the SPR value. Higher SPR value in all the three conditions in vidwath group reflected greater energy in the region of 2-4 kHz. It further hints that vidwath singers are indeed able to manipulate the vocal tract to tune the voice spectrum for the production of a target voice quality. This process in training needs to be gradually achieved by juniors and senior singers during the course of training. Hence, SPR would be considered as an objective tool for monitoring the training progress and measure the singing talent. Also, SPR analysis can be consider in the routine clinical voice evaluation for singers.

Under condition 1, 2 and 3, singer's formant was not prominent in Carnatic vocalists. Even in presence of loud background music (condition 3), the trained singers' physiological system does not adapt to improve the resonance at higher frequencies unlike western singers. Generalization of the results of the present study should be made with caution because of small sample size. To conclude, the study has found a gradual increase in SPR measure from junior level to vidwath level of training. Also, the background music has invariable effect on SPR value among the singers which need to be addressed. More number of singers and vocalic portion of the sung sample can be considered in future studies in the similar line.

References

Bartholomew, W. T. (1934). A physical definition of "good voice quality" in the male voice. Journal of the Acoustic Society of America, 6 (25), 25-33.

Boominathan, P. (2004). Do Carnatic singers have Singer's Formant? Proceedings of Frontiers of Research in Speech and Music.

Chayadevi, N. (2003). Singer's Formant in Indian Classical Singers-Carnatic vs. Hindustani. Students research at AIISH: Articles based on Dissertation, Vol I.

Kenny, D.T., & Mitchell, H.F. (2007). Vocal quality in female classical singers: The role of acoustics, perception and pedagogy. Proceedings of the third Conference on Interdisciplinary Musicology (CIM07).

Lundy, D.S., Roy, S., Casiano, R.R., Xue, J.W., & Evans, J. (2000). Acoustic analysis of the singing and speaking voice in singing students. Journal of Voice, 14(4), 490-493.

Mohan, R. (2010). The Effect of Training on the Voice of Carnatic Classical Singers. Students research at AIISH: Articles based on Dissertation, Vol II.

Mohan, R. G. (2007). The Singing Power Ratio: An Objective Measure of Singing Voice Quality in Untrained and Trained Singers, Usefulness in Indian Context. Proceedings of Frontiers of Research in Speech and Music.

Morris, J., & Weiss, R. (1997). The Singer's formant Revisited: Pedagogical Implications Based on a New Study. Journal of Singing, 53(3), 21-25.

Omori, K., Kacker, A., Carroll, L. M., Riley, W. D., & Blaugrund, S. M. (1996).Singing power ratio: Quantitative evaluation of singing voice quality. Journal of Voice, 10(3), 228-235.

Ram Mohan, G. (2007). The Singing Power Ratio: An Objective Measure of Singing Voice Quality in Untrained and Trained Singers, Usefulness in Indian Context. Proceedings of Frontiers of Research in Speech and Music.

Sataloff, R. T. (2005). Professional voice: The science and art of clinical care. 3rd edition, San Diego, CA: Plural Publishing; 3-4.

Sengupta, R (1990). Study on Some Aspects of the "Singer's Formant" in

North Indian Classical Singing. Journal of Voice, 4(2), 129-134.

Sujatha (1989). Analysis of singing voice. Students research at AIISH: Articles based on Dissertation, Vol II.

Sundberg, J. (1987). The science of the singing voice. DeKalb, Illinois: Northern Illinois University Press. 1-194.

Watts, C., Barnes-Burroughs, K., Estis, J., & Blanton, D. (2006). The singing power ratio as an objective measure of singing voice quality in untrained talented and nontalented singers. Journal of Voice, 20(1), 82-88.

ABOUT THE CONTRIBUTORS

Anirban Patranabis is a Post Graduate in Physics from Jadavpur University, Kolkata, India. He is a PhD student working in the field of Musical Signal Processing at Sir C V Raman Centre for Physics and Music, Jadavpur University, Kolkata, India. He has published some papers in national and international journals and has attended many national and international conferences and presented papers.

Kaushik Banerjee did his masters degree in Instrumental Music (Sitar), with Vocal Music as a special paper from Rabindra Bharati University, Kolkata. Thereafter he completed his PhD on the Contribution of Ustad Alauddin Khan & his disciples in the development and internationalization of Indian Classical Music, from Rabindra Bharati University, Kolkata. Trained both in Esraj and Sitar from Senia Maihar Gharana is a regular performer of both. Recipient of Junior fellowship (2004-2006) from the Department of Culture, Government of India on " A Historical study of musical (string) instrument makers, since 19th Century" . Received fellowship from National Film Archive of India (NFAI) on: "Social awareness and the changing profile of Bengali film music in the context of globalization". He is associated with Sir C V Raman Centre of Physics and Music, Jadavpur University as a Senior Guest Researcher continuing post-doctoral works on "changing characteristics of the sound (timbre) of Sitar & Sarod due to the development of microphones and related equipments a signal processing approach.

R Sengupta, a PhD in Physics, is associated with Sir C V Raman Centre of Physics and Music, Jadavpur University, Kolkata as a Scientific Consultant. He was the Senior Research Scientist and Head of the Department of Scientific Research, ITC Sangeet Research Academy, till 2010. He has done extensive research on speech and musical acoustics and contributed many research papers to national and international journals.

Frank Ragozzine received his B. A. in psychology, with a minor in music, from Youngstown State University in Youngstown, Ohio, in 1990. He then earned his M.A. (1993) and Ph.D. (1998) degrees in experimental psychology from the University of California, San Diego, in San Diego, California, where he specialized in the areas of music perception and music cognition. From 1999 until 2006, he was a member of the faculty at Missouri State University in the Psychology department, first as an Assistant Professor, and later as an Associate Professor. From 2006 to the present, he has been a faculty member in the Psychology department at Youngstown State University in Youngstown, Ohio. Much of his research has focused on perception of octave-related complexes and the tritone paradox. He has also investigated issues involving music and emotion, and has studied pedagogical issues. He has had work published in journals such as Music Perception and Teaching of Psychology. Additionally, he has presented his work at national meetings of the Association for Psychological Science (APS); the Acoustical Society of America (ASA); and the Auditory Perception, Cognition, and Action Meeting

(APCAM). Dr. Ragozzine is a member of the Acoustical Society of America as well as the Association for Psychological Science.

Soubhik Chakraborty, a PhD in Statistics, is currently an Associate Professor in the Applied Mathematics Department of B. I. T. Mesra, Ranchi, India. He has published several papers in the areas of algorithms, statistical computing and music analysis and authored one book. An acknowledged reviewer, he is also a harmonium player and the principal investigator of UGC major research project in his institute on the topic Analyzing the structure and performance of Hindustani classical music through Statistics.

Swarima Tewari, a B.E. in Electrical Engineering and an M.Tech. in Scientific Computing, is working as a Project Fellow in the UGC major research project on music analysis detailed above. She has a keen interest in music and has participated in several music related cultural events during her college days. She has published a few papers on music analysis with this research team and intends to pursue a PhD in a related area.

Khramov Mykhaylo Yuri is an Electromechanical Engineer from the Machine-Building Institute, Voroshilovgrad, Soviet Union, presently, Lugansk, Ukraine. He is also a specialist in Robotic systems from People's University, Voroshilovgrad, Soviet Union, presently, Lugansk, Ukraine. He is the founder of an Independent Project Commator, Lugansk, Ukraine. He is engaged in the analysis and correction of music scores of world-famous pieces for the performance in alternate tunings by using his own developed discipline "Sonantometry" and Synthesis of new music pieces with using of alternate tunings. He has also done publication on a Web of current outcomes for obtaining of listener opinions. He has written and used some plugins for use in "Sibelius" software. He has produced and distributed Audio CD "Another Tuning" with outcomes of his own experiments in years 2004 2006. He has executed more than 500 projects and tasks on computer 3D design of items from wrought iron. He is directing and training company teams on using PC in all related spheres of LAN administration and maintaining. Due to his special interest in music he has presented and published many papers in International conferences in music analysis.

N Dey is a Research scientist attached to the Scientific Research Department of ITC Sangeet Research Academy. He has authored a number of research papers that have been published in national journals on Speech and Musical Acoustics.

A K Datta is attached to the Department of Scientific Research, ITC Sangeet Research Academy as a Consultant Scientist. He was Professor and Head of the Department of Electronics and Communication Science Unit of Indian Statistical Institute, Kolkata. He has many publications in the area of speech and music in national and international journals. He is associated as a Senior Guest Researcher with Sir C V Raman Centre of Physics and Music, Jadavpur University as well as

with the BOM charitable trust, Kolkata.

Ms.Supritha Aithal, a B.Sc (Speech & Hearing) Internee at All India Institute of Speech and Hearing (AIISH), Mysore. She has her interest in the field of language, speech and music acoustics. She has a publication in online journal at www.LanguageinIndia.com. She is also a Classical bharathanatyam artist. She has given her dance performance in national and international concerts. And she has won several awards in state and national level competition to name a few kala aparanji sanman, kishora prathibe, swarna nattya mayuri and yuva prathiba scholarship and she is a doordarshan artist.

Ms. Swathi S is an Intern at All India Institute of Speech and Hearing, Mysore. She is undergoing training in rehabilitation of individuals with communication disorders at her esteemed institution. Her areas of interest include auditory perception, acoustic analysis, musical sciences and voice pathology. Few of her research studies include- "Assessment and management of voice problems in young and prospective professional voice users" presented in Clinical Conference at AIISH, Mysore, also "Effect of different levels of training on singing power ratio and singer's formant in classical carnatic singers" at Frontiers of Research in Speech and Music- 2012, an International conference held at Gurgaon. She was awarded with "Best clinical conference 2010-11" at AIISH, Mysore.