Proceedings of National Conference on Communications (NCC) 2015, IIT Bombay, India
27 February - 1 March, 2015

Discrimination of Melodic Patterns in Indian Classical Music

Kaustuv Kanti Ganguli and **Preeti Rao**
Department of Electrical Engineering
Indian Institute of Technology Bombay, Mumbai 400076, India.
Email: {kaustuvkanti, prao}@ee.iitb.ac.in

***Abstract*—The melodic phrases of a raga are an important cue to its identity. Artists, however, incorporate considerable creative variation within a raga phrase during performance while still preserving its identity in the ears of the listeners. It is of interest therefore to explore the boundaries of this categorization of phrase identity, given the space of musical variations in the tonal interval and duration dimensions. Such an endeavor can help better model musical similarity for music retrieval and pedagogy applications. In this work, we carry out melodic shape manipulations on a selected prominent phrase of raga Deshkar to study the subjective responses of musicians in comparison with non-musicians in terms of perceived discrimination of the controlled variations. A method is presented for deriving musically consistent synthetic stimuli for listening. Subjective responses on the discrimination and identification tasks are presented along with a discussion on possible perceptual mechanisms at play.**

I. Introduction

We enjoy the same musical piece every time we listen to it, unlike the case of speech. This is because the purpose of a musical message is more towards the affective function compared to normal speech communication. The inter-relationships of the acoustic attributes and subjective response to music span vast areas such as hearing perception, musicology, aesthetics and psychology. There is much research activity in music cognition in the Western world where the psychological correlates of tonality, pitch and rhythm have been explored. Considering the prominent differences between Indian and Western music, there is a need for similar studies dedicated to Indian music. In this work, we address one aspect of this subject, that is related to a fundamental building block of Indian classical genres, namely the raga-characteristic phrases.

In a typical Hinduatani music concert, an artist executes spontaneous realizations of the raga characteristic phrases that represent the raga identity. The characteristic phrases of a raga (lit. pakad) are typically referred to in terms of notation but are fully described only when the continuous pitch contour or melodic shape is presented. The artist or performer uses his knowledge of the raga grammar to interpret the notation when it appears in a written composition in the specified raga. The shape of a recurring melodic motif within and across performances of the raga shows variability in terms of one or more of the following aspects: pitch interval, relative note duration and shape of alankars (ornaments), if any, within the phrase [1]. It is interesting to explore the space of musical variations systematically in terms of the relationships between acoustic variability and perceptual distance. We focus on the first two dimensions of variability in the present study, namely pitch intervals and duration, both pertaining to the steady notes that appear in a phrase. We expect the 'perceived distances' to be different in musicians (having learned to classify the melodic shapes into categories) and non-musicians (likely to be operating with raw limits of human hearing only). We carry out listening experiments with independent manipulations on the two aspects using a suitable listening task for each: a discrimination task for pitch interval, and a raga labeling task for the relative note duration. Pitch interval and tonal duration manipulations have been used in experiments with Western musician listeners leading to the emergence of a theory of 'categorical perception' in music as reported in [2]–[4].

In the next section, we review previous literature on the categorical perception in general, and related works in music. With the melodic shape of a phrase being a gestalt (i.e., it is perceived and recognised holistically) in Indian classical music, as opposed to the individual discrete note values and duration values that form the building blocks of Western melody, we expect that the design of the experiments needs to be carefully considered. These aspects are discussed in the sections following the next. The experiments are described and the results obtained as averaged subjective responses of a set of musician and non-musician listeners are discussed.

II. Background

A. Categorical Perception

Categorical Perception (hereafter CP) is the experience of percept invariance in sensory phenomena that can be varied along a continuum [5]. CP is revealed when an observer's ability to make perceptual discriminations between things is better when things belong to different categories rather than the same category, controlling for the physical difference between the things. Sensory signals that could be linearly related to physical qualities are warped in a nonlinear manner, transforming analog inputs into quasi-digital, quasi-symbolic encodings. CP is an important phenomenon in cognitive science because it involves the interplay between humans' higher-level conceptual systems and their lower-level perceptual systems [6]. William James noted, "We live in a world of orderly, discrete objects, rather than ambiguous, amorphous percepts, and categorical perception is a mechanism that may underlie this experience" [7]. The generation of CP (enhanced within-category similarity and enhanced between-category differences) by perceptual learning has been described as the 'acquired similarity (or difference) of cues' but no mechanism has been proposed to explain how or why it occurs [8]. In short, the same entity

978-1-4799-6619-6/15/$31.00 © 2015 IEEE

gets a different identity when the perspective changes [9]. CP's origins are also tied to the 'motor theory' of speech perception, according to which the similarities and differences between speech sounds are determined for our ears by how our mouths would have produced them [10]. Categorical perception has been of major importance in speech-perception research since it was first named and systematically studied by Liberman et. al. [11]. He demonstrated CP by generating a continuum of equally spaced consonant vowel syllables with endpoints reliably identified as /be/ and /ge/, using voice onset time (VOT) as the feature by varying the second formant transition. As the same concept is intuitively valid for most speech-like stimuli, musical pitch categories may be thought of as being examples of CP effects that arise primarily as a result of learning. Music indeed is a very interesting test case, because it is less likely than speech to have inborn feature detectors already 'prepared' by evolution. Yet there exist parallels; absolute pitch, rhythm and harmony are among the variables to investigate [10]. Instrumental tone recognition by timbre perception also involves 'pattern matching' of templates and is unanimously agreed to be categorical in nature [12]. Study of CP has been researched in other perceptual domains as well. We shall now briefly look at the previous attempts made to explore CP in music.

Related work on CP in music

A strong motivation for studying CP in pattern recognition problem is that the irrelevant variations within clusters can be greatly deemphasized. It has been argued that experimental demonstrations of CP are strongly influenced by the way a task is presented [13]. The two main tasks involved in CP experiments in speech that are applied to music are Identification and Discrimination. Among the first attempts of studying CP in music, Burns & Wards [2] exploited ascending melodic intervals to design a three-category identification problem with an experimental setup that closely conforms to the speech perception experiments. Additionally in the discrimination task, the subjects were presented two successive melodic intervals and were asked to judge which interval was wider. In a similar study of CP of tonal intervals, Siegel & Siegel [3] remarked that "musicians can't tell sharp from flat": even trained musicians who are great experts in identifying standard musical manipulations (tone, duration) on a pattern, perceive a melodic entity holistically and become deaf to minute changes. All the above cases involved repetition and randomization of stimuli in different trial blocks to ensure consistency of the subjects' response. In the study of CP in musical patterns, Fiske [4] argued that there can be only three decision categories defining the degree of 'difference' of any inter-pattern relationship: (i) same, (ii) derived, and (iii) distinctly different. The above study addressed a subjective test for the discrimination task by incorporating tonal-duration manipulation on 40 pairs of tonal-rhythmic patterns of electronically synthesized tones with flute-like timbre. Note that there is a difference in the time-scale of the stimuli in this survey as opposed to the aforesaid ones. To summarize, there has been attempts of imparting 'warping' in both pitch and temporal dimensions to the original stimulus in order to observe the perceptual-intellectual demands of the listener in better understanding and recognizing musical patterns.

B. Relevance to Indian music

This paper aims to the study of CP in raga music, one of the highly practiced components of Indian art music tradition. A raga performance can be thought of as a sequence of melodic motifs or characteristic phrases. The precise phrase intonation is so crucial that it acts as a major cue to raga identification by listeners and is well accepted as the foundational unit of a raga in the pedagogical tradition as well [9]. Though a characteristic phrase (lit. pakad) of a raga often holds a unique canonical form, considerable variability is observed among the instances of the same phrase in a raga performance. This variation usually involves multiple dimensionalities, such as pitch, time, timbre, energy dynamics etc. [14]. It is implied that these phrases are still highly recognizable by trained listeners [15]. In two dimensions (pitch vs. time), the captured similarity among phrases is either local or global: there can be microtonal variation on a particular note, or the relative tonal duration structure may vary as well. Repeated use of the same melodic motif involves its inherent variability, but it is difficult to estimate the boundary of this variability space from the concert audio data. We propose a methodology to be able to gauge the fine limits of variability allowed for a melodic phrase within a raga framework. We aim to define an acoustic distance measure for melodic similarity metrics suited for Indian music audios. This can be evaluated against a baseline that uses state-of-the-art distance measures with empirical thresholding.

We exploit musicological knowledge to choose the stimuli for this experiment. We choose a phrase which is characteristic of the raga by itself without any further context. Presuming that perception of melodic phrase is categorical, it would be interesting to select a raga which has a parallel in terms of note sharing. Hence we take one such raga Deshkar which contains the same notes as in raga Bhoopali. There exist some note sequences that are common between them, though with contrasting behavior of the detailed phrase intonation. In brief, 'Re' of raga Deshkar is prescribed to be very short in duration whereas for raga Bhoopali, it is a nyas svara (long held steady note). The precise intonation of 'Ga' in raga Deshkar is higher than that of Bhoopali, whose Gandhar ('Ga') quite resembles the just intonation tuning [16]. We impart controlled distortions in terms of pitch and temporal 'warping' on a phrase sub-segment and hypothesize that the perception would emerge to be categorical. The time-scale of perception also plays a major role, because the neural processing for the recognition of entities with different time-scales is quite different [12].

III. Stimulus Creation

In any perceptual and experiment-based study, the stimuli play the most important role for the validity of the experiment. Characteristic melodic phrases of average duration of 5-7 seconds are recorded by a performing musician, mono 16 kHz sampled. We compute the pitch and energy of the vocal harmonics at every 10 ms interval, by the algorithm proposed by Rao & Rao [17]. Next, the pitch data is tonic normalized to obtain a time series (cents vs. time). The stimulus creation consists of three parts, viz.:

A. Pitch contour stylization

As mentioned in Section II-B, we incorporate controlled distortions onto the pitch contour of a melodic phrase, separately in temporal and pitch dimensions. Now it is important to identify the irrelevancies (unintended micro-prosodic variations) and deemphasize them in order to prevent their scaling during the warping operations. Thus it is very crucial to model a phrase in order to achieve a reasonably good and minimal representation. There have been different approaches employed, to stylize the pitch contour of a speech or music data by piecewise polynomial approximation. We employ a similar approach with slight adaptation to minimally represent a melodic phrase.

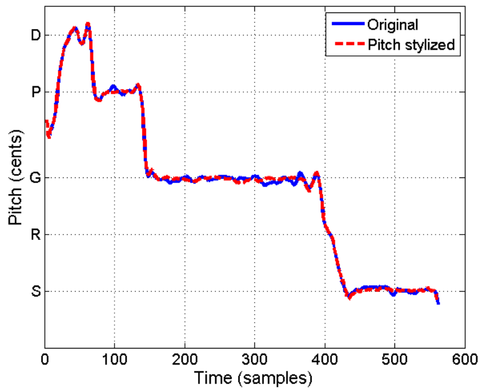

Fig. 1. Pitch contours of the recorded phrase 'Dha-Pa-Ga-Re-Sa' in raga Deshkar: before and after pitch contour stylization.

The pitch contour of a melodic phrase can be thought of as a chain of two events: (i) a pseudo-steady segment closely grazing a raga note, and (ii) a transitory segment which joins two steady segments smoothly. Generally the latter is often referred to as an alankar or an ornament which includes meend (glide), andolan (oscillation), kan (touch note) etc. To locate the pseudo-steady note segments, we first discretize the continuous pitch contour to the raga notes obtained from the pitch histogram. Then we discard or merge too short segments that do not qualify as independent steady notes, based on some heuristics. The note sequence obtained from the pitch contour as shown in Figure 1 is 'Dha-Pa-Ga-Re-Sa' which is a characteristic phrase of raga Deshkar. Lower threshold duration of 800 ms is applied to the found fragments to discard those that are considered too short to be perceptually meaningful as held notes. Next we fit a 3^{rd} degree polynomial to the remaining transitory segments. We finally add a measured noise (simulating vocal jitter) to the steady segments to make it sound more natural and less MIDI-like. Hence we generate a white noise with the same mean and standard deviation as that of the original pitch samples of the steady note segment. Next, the noise samples are smoothened by low pass filtering with necessary modifications to retain the standard deviation of the recorded one. There is an inherent temporal dependence of the perturbation patterns around a steady note: general observation is that the oscillation is larger around the onset and offset of the steady note-segment. To simulate an equivalent scenario, we preserve a few samples from the original pitch data around the onset and offset of a steady segment to result in a smooth transition into and away from the steady note. Figure 1 shows the pitch contours of the recorded phrase: before and after pitch contour stylization.

B. *Synthesis of musical stimuli*

To remove bias of the artist's voice and to get rid of the temporal evolution of the timbre as a cue to identification, we synthesize a voice-like tone with a constant timbre consisting of 5 harmonics. The relative harmonic weights with respect to the fundamental are 0, 3, 5, -6 and -20 dB respectively. These weight values have been empirically chosen to produce a perceptually pleasing timbre. 5 point median filtered vocal energy contour is used for the synthesis to retain the natural intensity dynamics. Figure 2 shows the spectrogram of the synthesized audio from the stylized contour.

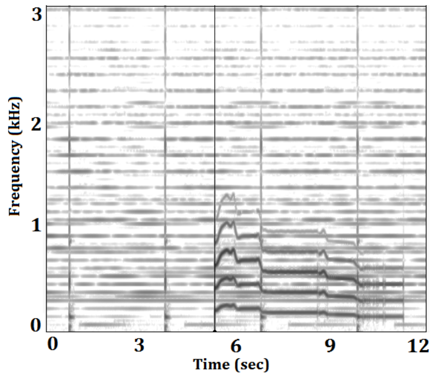

Fig. 2. Spectrogram of the synthesized audio from the stylized contour as shown in Figure 1. The relative harmonic weights with respect to the fundamental are 0, 3, 5, -6 and -20 dB respectively. Feeble horizontal lines denote the tanpura harmonics. Vertical lines denote the metronome clicks.

As there is no one-to-one correspondence of a melodic phrase to a raga without the tonic, we add a good quality recorded tanpura track in the background. Based on the feedback obtained from a pilot survey to judge the quality and intelligibility of the synthesized phrase, we add a longer 'tanpura pause' at the start and metronome at two of the note onsets to provide melodic as well as rhythmic anchor to the listener. As shown in Figure 2, clicks are coinciding with the onsets of 'Ga' and 'Sa' respectively.

C. *Melodic pattern manipulation*

The motivation behind melodic pattern manipulation is to artificially create distorted versions of the canonical form of the same melodic phrase and analyze user-response to estimate the space of allowed variability within the raga framework. This would have been impossible to obtain directly from concert audio recordings, as those do not contain the extremas of all possible variations of the same melodic phrase. The manipulation stage comprises of two independent steps, viz.:

1) Pitch shifting: Indian raga music uses micro-tonality. Hence, even within the pseudo-steady note segments, the standard deviation of pitch values is quite high that is evident from the wide peaks in the pitch histogram. Still, the histogram for different raga audios shows a considerable difference in the precise peak locations of the same note. This confirms that the exact note intonation (lit. shruti) is different across ragas. We deploy a shift of the pitch values of a steady note segment to generate the pitch shifted contour. Figure 3 shows a -25 cents shifted 'Ga' segment of the same pitch contour with respect to the one shown in Figure 1.

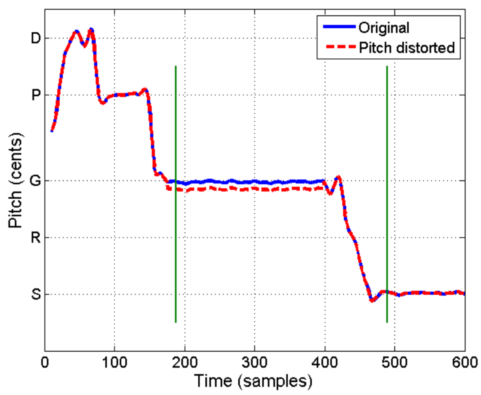

Fig. 3. Pitch contours of the same phrase as shown in Figure 1. The dotted contour corresponds to the pitch-shifted phrase, after shifting the 'Ga' segment by -25 cents. The two vertical lines correspond to the onsets of 'Ga' and 'Sa'.

2) Temporal warping: MIR research often needs dynamic time warping (DTW) distance as a similarity metric to tackle time stretching between 'similar' motifs. Previous work [15] describes a method to learn the warping paths to obtain the global and Sakoe-Chiba constraints. It has also been reported [14] that the time scaling is non-uniform and is absorbed mostly by the pseudo-steady pitch regions. But no investigation has ever focused on discovering the extent of time stretch allowed in each phrase sub-segment, which still sounds 'musical' and conveys the raga identity as agreed by musicians.

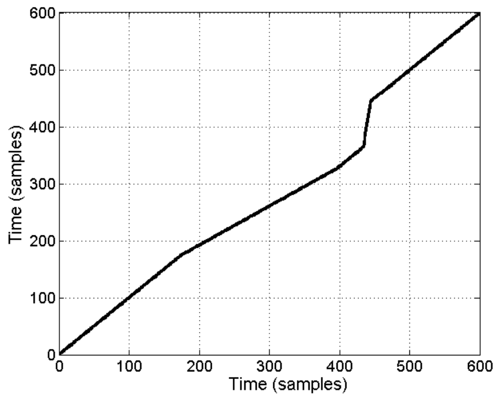

Fig. 4. The piecewise linear mapping function, that is used as a DTW path to generate the time-warped pitch contour. The path for the 'Re' segment has a slope of 8, corresponding to the warping factor of the pitch and energy contour. The 'Ga' segment is compressed accordingly.

We adopt the reverse strategy to define a warping path and generate a time-warped phrase from the given pitch contour. The resultant phrase is obtained by a point transformation of the input pitch data with respect to the given function. The following equation shows the mapping function, where a_i denotes the slope of the i^{th} segment extending from x_i to x_{i+1} and y_{x_i} is the intercept at the coordinate (x_i, y_i).

$$f(x, y) = a_i(x - x_i) + y_{x_i} \quad ; \qquad 0 \le i \le n - 1 \qquad (1)$$

The segments are the same as obtained from the stylization algorithm discussed in Section III-A. The values of the slopes a_i are chosen arbitrarily to expand (or contract) a certain segment with the factor a_i ensuring that $f(x, y)$ is a strictly monotonically increasing function, to guarantee possibility of inverse mapping. It follows from equation (1) that a_n (path slope of the last segment) cannot be a user input; hence there is a constraint while choosing the reference phrase and the concerned segment for time warping. But here we take a slightly different strategy to maintain the onset location of the last note 'Sa', hence the 'Ga' section is compressed accordingly. Figure 4 shows the generated warping path with 'Re' warping factor 8 and Figure 5 shows the corresponding pitch contour. Note that the synthesis of this warped pitch contour uses an interpolated energy contour with the same warping function. As both the pitch (cents) and energy (dB) are in logarithmic scale, the point transformation used is a linear interpolation.

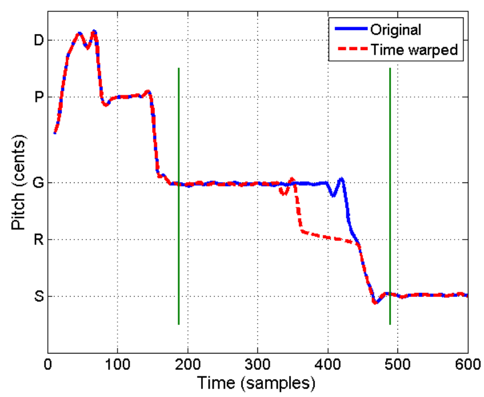

Fig. 5. Pitch contours of the same phrase as shown in Figure 1. The dotted contour corresponds to the time-warped phrase, after expanding the 'Re' segment by a factor of 8. Accordingly the 'Ga' segment is compressed. The same piecewise linear mapping function as shown in Figure 4 is used.

IV. Experiment and Results

This study critically involves a subjective experiment. We design a well-structured and controlled set of experiments to confirm our hypotheses in the form of an online survey, open for both musicians and non-musicians. A pilot study is conducted to ensure that listeners are comfortable with the quality of the stimuli and are able to identify the raga from the excerpt. The main survey contains two sets of audios for the discrimination and identification tasks, with randomized and repeated clips in each set to ensure consistency. There is no different trial blocks as the survey is not targeted to a large group of participants. Thus we consider each second response of the same subject as an independent entry. After pruning all inconsistent and incomplete responses, we report the results obtained from the analysis of 12 musicians' and 8 non-musicians' response. Certain additional clauses for each musician (genre of expertise, years of training, whether they teach music etc.) are recorded but have not been used for the analysis of the present study. In principle, most researchers [2], [3] have used melodic interval as the common feature for both Discrimination and Identification functions for studying CP in music, whereas Fiske [4] proposed tonal duration as a feature for the Discrimination task. Here we are not exactly following the procedure as suggested by [2], [3] per se, but using the two features independently for the two tasks that conforms to the concepts of Indian raga music and is in line to the listeners' perceptual cues. Now we shall define the functions of the two major tasks as follows:

A. *Pitch interval discrimination*

In this experiment subjects are supposed to choose between two categories: same and different. A total of 15 clips are presented, each being a pair of phrases: one reference and the other pitch warped, randomly ordered, with a continuum of the metronome clicks and a 'tanpura pause' in between. The different levels of pitch shift employed to the 'Ga' segment are: -30, -25, -20, -10, -5, +5, +10, +20, +25, and +30 cents.

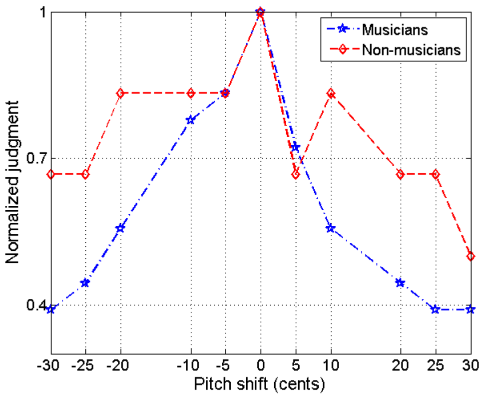

Fig. 6. Average response obtained for the Discrimination task, separately for musicians and non-musicians. The judgment is binary: categories 'same' and 'different' correspond to the levels 1 and 0 on the vertical axis.

Figure 6 shows the average response separately for musicians and non-musicians. The categories 'same' and 'different' correspond to the levels 1 and 0 on the vertical axis. From musicians' response it is evident that pairs are marked as same, if within a range of [-10, +5] cents shifting using 70% cut-off. This clearly reveals the fact that human ears perceive an entity holistically and tend to ignore small variations. There is also an asymmetry in the musicians' curve about 0 cents warping which can be interpreted as follows: the mean pitch of the 'Ga' segment in the original phrase (393 cents) is higher than the just scale 'Ga' intonation (386 cents). Therefore, lowering the 'Ga' might be within the listeners' comfort region whereas up-shifting causes it sound 'out-of-tune'. For non-musicians, no definite trend is observed. An outlier at +5 cents shifting is noticed which may have the possible explanation as follows: the mean pitch of 'Ga' (398 cents) for this phrase is very close to the equi-tempered 'Ga' location (400 cents) and may have caused a roughness in presence of the tanpura, due to the interaction of higher order harmonics of 'Ga' with that of the tonic 'Sa'.

B. *Tonal duration based identification*

In this experiment subjects (only musicians this case, because they are expected to have raga knowledge for the labeling task) are supposed to make a choice between three categories: whether the given phrase is (i) strongly suggestive, (ii) somewhat suggestive, or (iii) not at all suggestive of the given raga Deshkar. A total of 15 clips are presented, each having a temporal warping on the 'Re' segment with a factor $\geq$ 1. We restrict to this choice to keep the experiment controlled in dimension, otherwise interpretation of the response would have been non-trivial. The warping factors used are: 1, 1.5, 2, 3, and 8 which conforms to the hypothesis, giving a larger resolution in the neighborhood of the original phrase. This enables us closely delimit the first two categories.

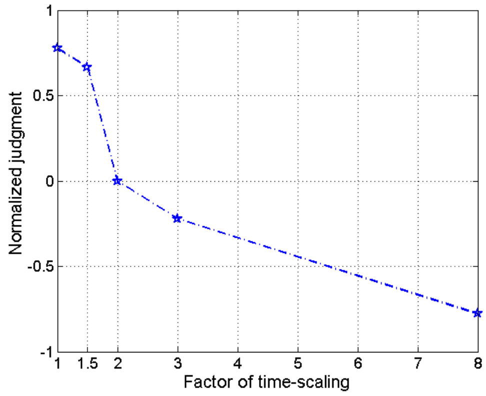

Fig. 7. Musicians' average response obtained for the Identification task. A steep fall of the curve is observed around the category boundaries.

The ideal response as discussed in the literature is a step-like function: a larger jump at the category boundary is expected. Figure 7 shows the average response of 12 musicians which somewhat resembles the expected behavior, but would have been better with uniform resolution of the warping factors and more responses. Ideally the three aforesaid categories would correspond to the levels +1, 0, and -1 on the vertical axis. We can observe clear category boundaries between warping factors (1.5, 2) and $>$ 3. An extra comment box is provided for user feedback if the user feels the given phrase to be suggestive of any other raga. Interestingly in most cases, we find a mention of raga Bhoopali for factors $\geq$ 3. This clearly confirms raga perception being categorical for trained musicians. Another significant outcome of this experiment on tonal-duration manipulation is that it shows the relevance of relative nature rather than absolute values, of the note durations in raga identification by 'pattern matching' scheme in the human brain.

C. *Discussion*

In any perceptual experiment it is necessary that all subjects are presented with the same conditions to ensure inter-subject consistency. But in this study, the experiment we deploy is in the form of an online survey. We have no control over certain details (like use of headphones instead of speakers, headphone quality, hardware settings of the audio playback device etc.) which might have considerable impact on the subject during the survey. Another crucial issue in such survey is the choice of baseline. In the present study, we compare musicians' response with that of non-musicians which is obviously not the best way. Effect of training, experience of teaching, expertise in a different genre etc. would have been interesting parameters to exploit to construct a series of baselines. It is also often visible from the user response that an initial assumption carries away the attention to an irrelevant aspect and hence creates a bias to the survey data. There might as well be data dependence in the outcome of the survey, hence interpreting the results correctly, is another crucial issue in such perception based studies.

V. Conclusion and Future Work

This work aimed to explore the relevance of Categorical Perception in Indian music through a raga based listening experiment. Categorical Perception is tied to reducing redundancy so that we simulate the neural 'minimum effort way'. The 'selective information loss' is obtained by stylizing the melodic contour to obtain a minimal representation, discarding irrelevant pitch excursions. The listening experiments results are interesting in the following aspects:

- Musicians and non-musicians differ in the pitch-interval discrimination task, with the musicians showing finer discrimination. More interesting is the observed asymmetry in the direction of the pitch interval difference in the case of the musicians. This was explained based on the underlying raga grammar and therefore suggests categorization by the musicians based on their training and raga knowledge.
- As for relative tone duration manipulations, musicians indicate that the raga phrase is recognizable up to as large as 50% change in relative note duration. Beyond this, there is a sudden drop in identification score, which again is suggestive of categorical perception.

In summary, our goal is to understand the effects of certain valid musical manipulations on music perception. We speculate that there is a possibility of a phenomenon similar to CP at work in the musicians' responses. A more rigorous study is needed to replicate discrimination experiments that demonstrate the clear existence of a boundary between categories by shifting the region of discrimination along a continuum as reported in [2]–[4], [6], [11].

Insofar in MIR for Indian music, researchers have been applying traditional distance metrics directly, with slight adaptations. It is also crucial to consider the time-scale of psychoacoustic relevance and adapt our 'windowing analysis' accordingly. It is a well-founded fact that featuring a 'good' and 'interesting' representation of a melodic motif is a branch of musicology or behavioral science studies. It is, at the same time, very difficult to obtain an exemplar performance which is unanimously agreed to be considered as ground truth. But ideally a technology 'product' should be able to 'score' a raga performance and provide qualitative comments, as been practiced in the pedagogical structure of Indian music training. It would also be interesting to further investigate different aspects of music perception and observe the effect of music training in better understanding music, not taking 'learning-to-listen' for granted. This work can be extended to develop a category-learning model and define a new objective similarity (or distance) measure best suited for Indian raga music that conforms to musicians' perception as well as music theory. Also, there have been studies showcasing the proficiency of different brain hemispheres for labelling and discrimination tasks. It would be interesting to design some psychoacoustic experiments to validate the same in context of Indian raga music as a proof-of-concept. We would like to further investigate whether fitting a fuzzy membership function is a better approach to predict the category boundaries. Yet, to make an impact of the study, it is required to conduct a large scale survey in a controlled way which is posed as a future work.

Acknowledgment

This work received partial funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement 267583 (CompMusic). The authors would also like to thank all the participants who took their valuable time and patience to take the online survey.

References

[1] P. Rao, J. C. Ross, and K. K. Ganguli, "Distinguishing raga-specific intonation of phrases with audio analysis," *Ninaad*, vol. 26-27, no. 1, pp. 59–68, Dec. 2013.

[2] E. M. Burns and W. D. Ward, "Categorical perception – phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals," *Journal of Acoustic Society of America (JASA)*, vol. 63, no. 2, pp. 456–468, Feb. 1978.

[3] J. A. Siegel and W. Siegel, "Categorical perception of tonal intervals: Musicians can't tell sharp from flat," *Perception & Psychophysics*, vol. 21, no. 5, pp. 399–407, 1977.

[4] H. E. Fiske, "Categorical perception of musical patterns: How different is 'different'," *Bulletin of the Council for Research in Music Education*, no. 133, pp. 20–24, 1997.

[5] "Categorical perception," last accessed: October 22, 2014. [Online]. Available: "http://en.wikipedia.org/wiki/Categorical_perception"

[6] R. L. Goldstone and A. T. Hendrickson, "Categorical perception," *Cognitive Science: Wiley Interdisciplinary Reviews*, vol. 1, no. 1, pp. 69–78, Feb. 2010.

[7] S. Harnad, *Categorical Perception*. Nature Publishing Group/Macmillan, 2003, ch. Encyclopedia of Cognitive Science.

[8] S. Harnad, S. J. Hanson, and J. Lubin, "Categorical perception and the evolution of supervised learning in neural nets," in *Working Papers of the AAAI Spring Symposium on Machine Learning of Natural Language and Ontology*, 1991, pp. 65–74.

[9] K. K. Ganguli, "How do we 'see' & 'say' a raga: A perspective canvas," *Samakalika Sangeetham*, vol. 4, no. 2, pp. 112–119, Oct. 2013.

[10] "Advanced topics: Categorical perception (cp)," last accessed: October 22, 2014. [Online]. Available: "http://users.ecs.soton.ac.uk/harnad/at.html"

[11] A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, "The discrimination of speech sounds within and across phoneme boundaries," *Journal of Experimental Psychology*, vol. 54, no. 5, pp. 358–368, Nov. 1957.

[12] J. G. Roederer, *The Physics and Psychophysics of Music: An Introduction*, 4th ed. Springer, 2008.

[13] S. K. Scott and S. Evans, "Categorizing speech," *Nature Neuroscience*, vol. 13, pp. 1304–1306, 2010.

[14] K. K. Ganguli and P. Rao, "Tempo dependence of melodic shapes in hindustani classical music," in *Frontiers of Research on Speech and Music (FRSM)*, Mar. 2014, pp. 91–95.

[15] P. Rao, J. C. Ross, K. K. Ganguli, V. Pandit, V. Ishwar, A. Bellur, and H. A. Murthy, "Classification of melodic motifs in raga music with time-series matching," *Journal of New Music Research (JNMR)*, vol. 43, no. 1, pp. 115–131, Apr. 2014.

[16] K. K. Ganguli, "How *musical* is the music scale?" PhD Seminar: IIT Bombay, Tech. Rep., Nov. 2013.

[17] V. Rao and P. Rao, "Vocal melody extraction in the presence of pitched accompaniment in polyphonic music," *IEEE Trans. Audio Speech and Language Processing*, vol. 18, no. 8, 2010.