To appear in Proc. of Frontiers of Research on Speech and Music (FRSM)
15-16 December 2017, NIT Rourkela, Orissa, India.

# IMITATE OR RECALL: HOW DO MUSICIANS PERFORM RAGA PHRASES?

Kaustuv Kanti Ganguli and Preeti Rao

Department of Electrical Engineering
Indian Institute of Technology Bombay, Mumbai.
{kaustuvkanti,prao}@ee.iitb.ac.in

***Abstract-*** *A raga performance in Hindustani vocal music builds upon a melodic framework wherein raga-characteristic phrases are presented with creative variations while strongly retaining their identity. It is therefore of interest, for both music information retrieval and pedagogy, to understand better the space of "allowed" variations of the melodic motifs. Our recent study of melodic shapes corresponding to a selected raga phrase showed that variations in the temporal extent of a passing note within a characteristic phrase was perceived categorically by trained musicians. In this work, we investigate the same phenomenon via a different paradigm 'listen and imitate' wherein the results are achieved in a more natural situation for the subjects. The results demonstrate that trained musicians perform a memory abstraction of the raga-characteristic phrases and are deaf to small acoustic differences when instructed to imitate. The analyses also provides insights into the relative importance of the constituent svaras in a phrase in the raga context.*

**Keywords-** Raga-characteristic phrase, Listen and Imitate, Categorical Perception (CP).

## 1. Introduction

The pedagogy in Hindustani music repertoire demands a learner to produce and recognize raga phrases. A relevant analogy would be to imagine a phrase as a spoken word in a language that musicians understand. Given variants of a phrase, we investigate how sensitive musicians are to the differences and measure how the physically measured acoustic signal differences relate to perceived differences. Our recent work [1] investigated, through acoustic measurements followed by behavioral experiments through listening, the possibility of a canonical form or "prototype" of a raga characteristic phrase. In our context, a prototype may be considered as the phrase that serves to establish the raga around the initial phase of the performance. The case study was conducted for a characteristic phrase DPGRS in raga Deshkar. We hypothesize that a trained Hindustani musician performs a memory abstraction for the raga characteristic phrases.

Prototype and exemplar models are the two predominant approaches to modern conceptualizations of perceptual categorization, each providing a different assumption about the nature of category membership [2]. Prototype models attribute categorization to the comparison of incoming stimuli with internal prototypes, which are some form of averaged or ideal category representations. Exemplar models propose that experiences are stored in memory and that categorization is determined by the set of exemplars elicited by the incoming stimulus. Exemplars for a given category thus are specific instances of a stimulus, rather than a single, averaged representation of experienced or idealized stimuli [2]. Our recent study [3] used standard music information retrieval (MIR) tools to explore melodic structures in a data-driven way, to validate certain musicological hypotheses. Judicious use of both data and knowledge can lead to building a cognitively-based computational model that could simulate the human-judgment of melodic similarity. Apart from its pedagogical value, this approach has potential applicability as a compositional aid.

The structure of the rest of the paper is as follows. We first review previous studies on human similarity-judgment in speech and music literature. Next we discuss the experimental details and data collection. The following section describes the results for different phrase categories and subjects' groups with a discussion on possible perceptual mechanisms at play. Finally, we summarize our findings with concluding remarks about a comparison of the standard CP test design with the task at hand.

## 2. Overview of Literature

The findings from our previous work [4] suggested that trained Hindustani musicians perceive melodic phrases categorically, with less sensitivity to small changes around the prototype region. This supports the hypothesis that prototypes work as a perceptual attractor where musicians are less sensitive to small variations and they tend to perceive the phrase holistically. That perception of a raga characteristic phrase showed perceptual magnet (attractor) effect, we carried out the same set of experiments with the same GRS phrase in a context (DPMGRS) where it is not characteristic of any particular raga. The individual responses indicate the same (perceptual attractor) effect, with the constraint that the location of the 'prototype' depends on each musician's initial assumption (of the closest raga).

Vempala and Russo [5] observed that melodic contour (directions of pitch change) was an important predictor in the perceived similarity of phrases with single note alterations. Authors motivated the importance of a cognitive basis for melodic similarity. Mullensiefen and Frieler [6] showed that automated similarity measurements performed well in folk-song similarity from symbolic scores. From this work, we learn: (i) importance of using musically trained subjects, (ii) designing of stimuli, and (iii) drawing conclusions about the predicting power of various representation-cum-similarity measures. There are few distinct approaches in the methodology adopted by researchers. Marsden [7] remarks that some studies have attempted to judge similarity on the basis of some real musical activity. Studies aimed at producing measurements for use in query-by-humming systems have been based on asking subjects to sing a known melody (Hu et al., 2002; Pardo et al., 2004). Sometimes subjects are asked to deliberately vary a melody (Bernabeu et al., 2011), the variations are assumed to be more similar to the original than to other melodies.

The current experiment belongs to the `Listen and Imitate' paradigm which was tested by [8] in context of perception of low vs. high boundary tones in German language. The task was to imitate the stimulus prompt (F0 contours) as closely as possible. The general idea for an imitation test is that speakers listening to the stimuli representing a continuum between extreme values of one specific pattern, are expected to produce instances of two discrete sets of stimuli when they perceive two distinct categories. The aim of this paradigm is to test whether subjects memorize the stimulus and tend to recall from their working memory or whether they interpolate the stimulus to the closest prototype template already stored in their long-term memory (LTM) and reproduce the prototype.

Pierrehumbert and Steele [9] presented short phrases differing in peak delay to their subjects. The task was to imitate the presented stimuli as closely as possible. If the presented stimulus continuum represents a gradual change of one specific pattern, then subjects are expected to imitate the continuum correctly. Authors remark that when perception is categorical, the subjects' repetitions should fall into two clearly separable categories with respect to the manipulated feature. Furthermore, the imitations should show an explicit category boundary between the first and the second category. Authors found hints for CP of two peak categories in American English. Several other researchers demonstrated that this method seems to be quite successful for confirming the existence of intonational categories. Redi [10] reported that the subjects did not repeat all points in the continuum but they seemed to have two distinct peak delay categories in mind, which they reproduced. She also argued that this imitation task is a much better design to test for the perception of intonation than the classical one used to confirm CP, because the subjects' intuition is tested in an indirect way, without explicitly asking to identify or discriminate stimuli. This, therefore, seems to be a more natural way to receive information about the behavior and the perception of intonational categories.

## 3. Experimental Details

### 3.1. Stimuli

This experiment is conducted for Type A stimuli [1] that involves only R duration manipulation in the model space, for both characteristic DPGRS and non-characteristic DPMGRS phrases [4], as shown in Figure 1. Figure 2 shows the model space for both the phrase categories. The GRS segment (from the onset of G till the end of the phrase) for both phrase categories is exactly the same, they differ only in the precontext of the GRS segment – DP for characteristic DPGRS phrase and DPM for non-characteristic DPMGRS phrase. Note that the DPM segment for the latter is aligned to the DP segment of the former phrase category, the G onset (marked by vertical line in Figure 1) exactly coincides. The signal processing steps to obtain the melodic contour from audio is left out in the interest of space, refer to our previous publications for the pre- and post-processing steps for pitch contour stylization [11].

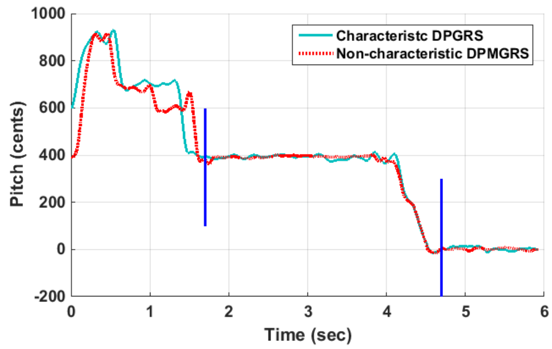


**Figure 1.** Stylized pitch contours for the characteristic DPGRS (cyan) phrase in raga Deshkar and the non-characteristic DPMGRS (red) phrase (a generic descending phrase, not typical of any particular raga). The blue vertical lines correspond to the onsets of the G and S svaras.

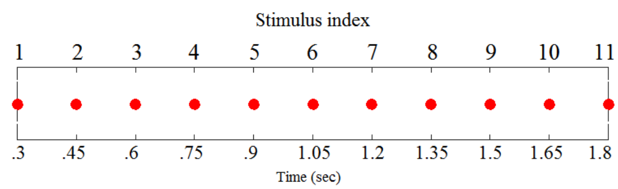


**Figure 2.** Model space: stimuli indices and corresponding duration of the R svara. The elongation of the R is compensated by shrinking of the G duration.

### 3.2. Subjects

Three categories of musicians participated in this experiment. The first category (Hindustani vocalists trained over 10 years) contained 3 participants. These subjects had formally learnt raga Deshkar. The second category (Hindustani musicians, not vocalists) contained 3 instrumentalists who learnt raga Deshkar but were not used to vocal rendition. The third category (vocalists, not formally trained in Hindustani music) contained 3 vocalists who were adept in singing commercial songs but did not formally learn raga music. This category of musicians is interesting to us, because they expertise in imitating a song which is the core task for our experiment. Without loss of generality, we combined the first two groups as 'Trained musicians' subjects' group and the last one as 'Untrained musicians' subjects' group.

### 3.3. Method

As the name 'listen and imitate' suggests, the experiment involves recording of participants' vocal rendition of the stimuli. Subjects were instructed to repeat the stimulus prompt as closely as possible, no other information about the stimulus was revealed. A total of 22 stimuli (11 each from the two phrase categories) were presented, each followed by recording of the subject's corresponding sung response, in each trial block. The stimuli within each phrase category were randomized, the phrases from the other category were used as conditioning stimuli and hence were interleaved in a randomized order. After each stimulus was played, a longer pause was provided with the tanpura drone continuing in the background. Instructions were given about keeping overall length similar to that of the stimulus. While the stimuli consisted of a metronome, no metronome was played in the subsequent pause, to ensure that subjects are not burdened with an additional constraint of aligning the G and S svara onsets with the metronome beats.

### 3.4. Recording

The recordings were carried out in a quiet environment, on a high fidelity digital recorder (Edirol R-09H) with an audio encoding of 44.1 kHz sampled 16-bit mono PCM (.wav) format. Subjects listened to the stimulus through an over-ear headphone (Sennheiser HD-180) with a moderate volume level. After the stimulus region, the background tanpura played only in the headphone, hence was not recorded on the recorder, ensuring we recorded a clean audio for better melody extraction and further analyses.

## 4. Results and Discussion

We take a semi-automatic approach of segmenting the G/R/S svaras from the pitch contours of the recorded phrases and thereby carrying out acoustic measurement (duration, intonation) on each svara. A stable svara transcription algorithm [3] segments the melodic shapes into its constituent svaras. We manually verify the precise onset and offset locations and store the duration of the stable svaras and the transient pitch segments joining them. As mentioned earlier, the only changes in the stimuli are the R and G svara durations. Figure 3 shows the correlation of the model space duration and the sung duration of these two svaras for both phrase categories and subjects' group.

The hypothesis is that if a subject listens and exactly imitates, the correlation between model space durations and sung durations would be high and hence the points would lie along the diagonal of the plot. We investigate the dependence of svara, phrase category, and subjects' training/background in the perception, hence imitation of the phrases. It has been established that the 'listen and imitate' is a more natural way of capturing presence of CP. Hence we shall interpret the plots as to how a musician perceives a phrase and performs a memory abstraction thereof.

### 4.1. Characteristic DPGRS

Observation on the data-points for the R svara for 'Trained musicians' subjects' group indicates a proper categorization in the sung duration with a sharp category boundary around the stimulus index 6 (absolute duration of 1 sec). The consistent short duration of the R svara for the stimuli indices 1 through 6 indeed corresponds to the prototype shape of raga Deshkar GRS phrase. However, the other data-points (corresponding to stimuli indices 7 through 11) had more dispersion. The distribution of the sung duration of R svara is bimodal with a sharp peak at .4 sec and a flatter long tailed peak 1.3 sec. This indicates that trained musicians, who are familiar with the raga which the stimuli belonged to, used their stored prototypes, though instructed to imitate a given stimulus.

Untrained musicians, on the other hand, followed the melodic shape almost exactly for both the svaras R and G – this is evident from the diagonal nature of the correlation plots. The slope of the (hypothetically joined) curves are complementary, showing that the subjects adjusted the G and R svara durations to keep the overall duration between the G and S svara onsets constant. Notably, all subjects, though not instructed to, tapped their hands at the given tempo of the metronome beats within the stimuli and tried to maintain it in their renditions of the phrase. We define a metric, $\beta$ as the ratio of sum of G and R durations in the model space and the musician-rendered phrases. The distribution of $\beta$ (mean=.94, SD=.3) indicates that subjects intended to maintain the tempo of the stimuli.

The case for G svara for trained musicians, however, is interesting. At one end of the curve (corresponding to stimuli indices 1 through 5 of the model space of R duration), the dispersion is low, indicating a categorization. But the G durations are, otherwise, correlated with the model space duration when the R is elongated. This indicates, when the R duration is in the non-prototypical context (vis-à-vis in the prototypical context of raga Bhupali), the trained musicians imitated the G svara duration. In such cases, the value of $\beta$ is less because the combined duration of G and R svaras are more than that of the model space (3 sec).

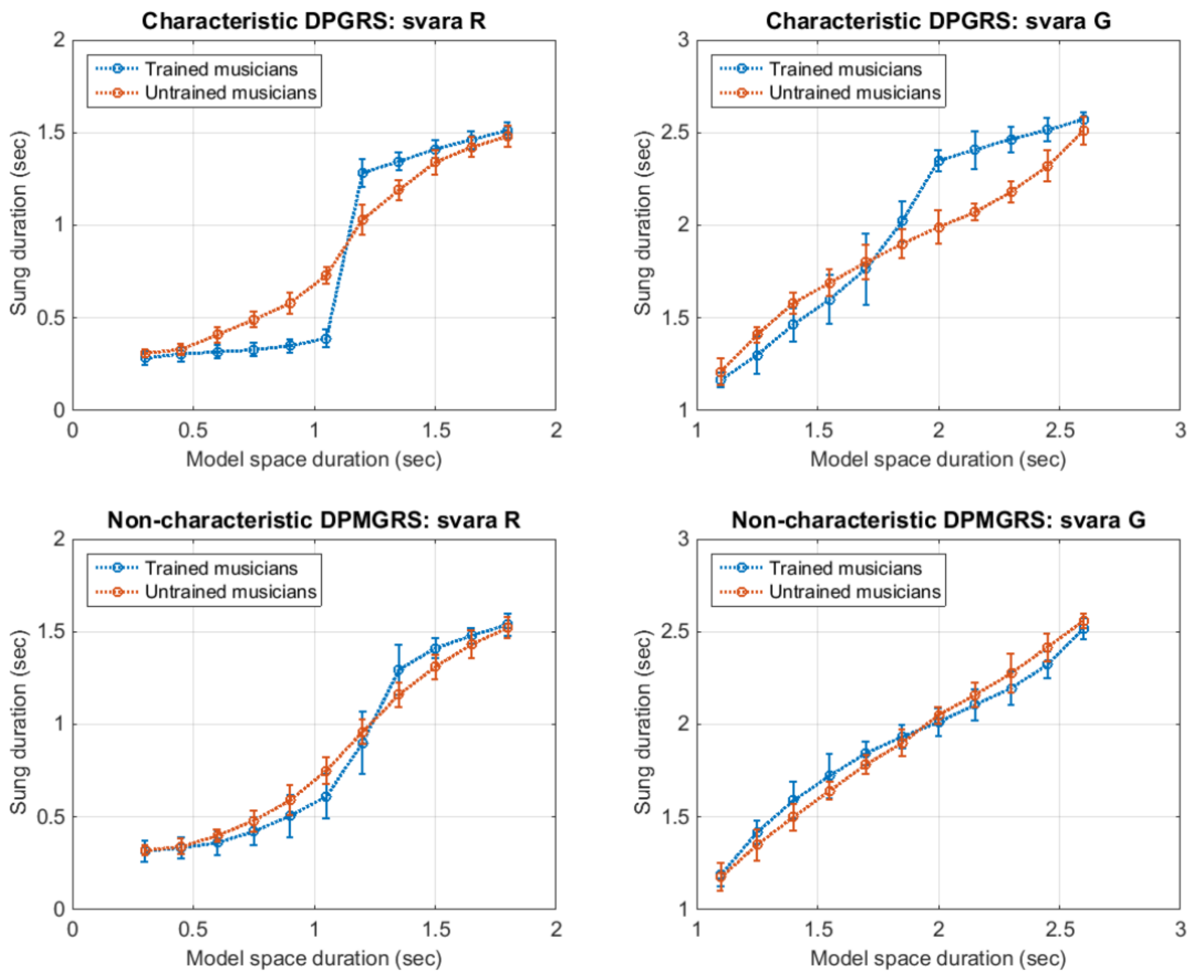


**Figure 3.** Plots of mean (with error-bar) sung duration vs model space duration of the R and G svaras for the two phrase categories and two subjects' group.

The S svara durations, for both subjects' group, have a high dispersion. As there is no expected event thereafter, musicians might not have paid much attention to the S duration. Sometime, instrumentalists, who are not accustomed to singing long phrases, went out of breath and hence shortened the S duration. In general, other than certain pitching errors (around the onsets), instrumentalists were as good as vocalists and hence we combined their responses.

### 4.2. Non-characteristic DPMGRS

The case for non-characteristic DPMGRS is interesting, because there is no expected prototypical context for this phrase for the trained musicians also. As expected, for both R/G svaras and both subjects' groups, the curves almost lie along the diagonal. The untrained musicians' response is no different, they almost exactly followed the stimulus melodic shape. The interesting observation for the trained musicians' response for the R svara is the high dispersions along the mid-range of the model space. These large error-bars are the result of differences in individual responses. All of these subjects disclaimed that they assumed the phrase to belong to a particular raga (e.g. Yaman, Sudh Kalyam, Vachaspati, and Maru Bihag). Interestingly, the DPMGRS phrase, with the different R svara durations, can fit in either of these ragas where there is no prescribed R duration. The distribution of β (mean=.96, SD=.2) for this phrase category indicates that the phrase is more closely imitated, which is evident from Figure 3 as well.

## 5. Concluding Remarks

In general, the imitation task design which does not tell the listener what has to be imitated and how closely, seems to be a good alternative to a standard CP test design, because the results seem to be achieved in a more natural situation for the subjects. Our findings suggest that trained Hindustani musicians perceive melodic phrases categorically, with less sensitivity to small changes around the acquired prototype region. Overall, the results indicate presence of CP in trained musician's perception and that they perform a memory abstraction in the LTM for the melodic shape of raga characteristic phrases. Skilled vocalists, yet formally untrained in the repertoire, performed the imitation task with high precision. The stimulus prompt length was kept short to conform to the human acoustic memory. One interesting finding was that all subjects attempted to maintain the tempo of the prompt, though was not instructed to. In fact, most of the responses were rendered as a continuum (of tempo) of the stimulus prompt. Similar to the perception about belongingness of the non-characteristic DPMGRS phrase to a possible close raga where it fits [4], a disjoint set of subjects (trained musicians) again showed the same phenomenon in the imitation task. This indicates that trained musicians tend to interpolate any melodic pattern to a familiar raga where it 'fits' in and thereby their perception is not acoustically, rather semantically, driven. With the convergence of the current findings with our previously reported ones [4,11], we believe that indeed there is a categorization in the musicians' perception and they, consciously or unconsciously, recall the memorized prototype while rendering a raga characteristic phrase in isolation. Now whether the same phenomenon occurs during a free improvisation in course of a raga performance is a complex question and needs further investigation to conclude.

## Acknowledgement

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).

## References

[1] K. K. Ganguli and P. Rao, "Exploring melodic similarity in Hindustani classical music through the synthetic manipulation of raga phrases," in Proc. of the Cognitively-based Music Informatics Research Workshop (CogMIR), New York, August 2016.

[2] B. E. Acker, R. E. Pastore, and M. D Hall, "Within-category discrimination of musical chords: Perceptual magnet or anchor?" *Perception & psychophysics*, 57(6):863–874, 1995.

[3] K. K. Ganguli, S. Gulati, X. Serra, and P. Rao, "Data-driven exploration of melodic structures in Hindustani music," in Proc. of Int. Soc. for Music Information Retrieval (ISMIR), August 2016.

[4] K. K. Ganguli and P. Rao, "Perceptual anchor or attractor: How do musicians perceive raga phrases?" in Proc. of Frontiers of Research on Speech and Music (FRSM), November 2016, Baripada, India.

[5] N. N. Vempala and F. A. Russo, "A melodic similarity measure based on human similarity judgments," in Proc. of Int. Conf. on Music Perception and Cognition, July 2012.

[6] D. Mullensiefen and K. Frieler, "Measuring melodic similarity: Human vs. algorithmic judgments," in Proc. of Interdisciplinary Musicology, 2004.

[7] A. Marsden, "Interrogating melodic similarity: A definitive phenomenon or the product of interpretation?" *Journal of New Music Research (JNMR)*, 41(4):323–335, 2012.

[8] K. Schneider, "The German boundary tones: categorical Perception, perceptual magnets, and the perceptual reference space," PhD thesis, Universitat Stuttgart, 2012.

[9] J. B. Pierrehumbert and S. A. Steele, "Categories of tonal alignment in English," *Phonetica*, 46(4):181--196, 1989.

[10] L. Redi, "Categorical effects in production of pitch contours in English," in Proc. of the $15^{th}$ Int. Congress of the Phonetic Sciences, 2003.

[11] K. K. Ganguli and P. Rao, "Discrimination of melodic patterns in Indian classical music," in Proc. of National Conference on Communications (NCC), February 2015.