On the perception of raga motifs by trained musicians

Kaustuv Kanti Ganguli, and Preeti Rao

Citation: The Journal of the Acoustical Society of America **145**, 2418 (2019); doi: 10.1121/1.5097588
View online: https://doi.org/10.1121/1.5097588
View Table of Contents: https://asa.scitation.org/toc/jas/145/4
Published by the Acoustical Society of America

On the perception of raga motifs by trained musicians

Kaustuv Kanti Ganguli and Preeti Rao[a)]
Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India

(Received 11 January 2019; revised 19 March 2019; accepted 22 March 2019; published online 29 April 2019)

A prominent aspect of the notion of musical similarity across the music of various cultures is related to the local matching of melodic motifs. This holds for Indian art music, a highly structured form with raga playing a critical role in the melodic organization. Apart from the tonal material, a raga is characterized by a set of melodic phrases that serve as important points of reference in a music performance. Musicians acquire in their training a knowledge of the melodic phrase shapes or motifs particular to a raga and the proficiency to render these correctly in performance. This phenomenon of learned schema might be expected to influence the musicians' perception of variations of the melodic motif in terms of pitch contour shape. Motivated by the parallels between the musical structure and prosodic structure in speech, identification and discrimination experiments are presented, which explore the differences between trained musicians' (TMs) and non-musicians' perception of ecologically valid synthesized variants of a raga-characteristic motif, presented both in and out of context. It is found that trained musicians are relatively insensitive to acoustic differences associated with note duration in the vicinity of a prototypical phrase shape while also clearly demonstrating the heightened sensitivity associated with categorical perception in the context of the boundary between ragas. © *2019 Acoustical Society of America*. https://doi.org/10.1121/1.5097588

[TS] Pages: 2418–2434

I. INTRODUCTION

The perception of similarity in music is central to music information retrieval (MIR), where pieces of music must be categorized using different musical features in order to facilitate the search (Downie, 2003). Research on musical features and the associated similarity measures has therefore remained pertinent given the rapid growth in digital music archives over the years. For instance, in the categorization of songs based on melody, global melodic features have been found to be less important to overall perceived song similarity compared to the match observed at the level of local melodic motifs (Cambouropoulos *et al.*, 2001; Volk and Kranenburg, 2012; Boot *et al.*, 2016). These findings from studies on Western folk-song tune families closely match the notion of characteristic motifs of ragas in Indian art music (Cowdery, 1984; van der Meer, 1980). The melodic organization of the Indian genre is governed by the modal system of ragas, and a musical performance is based on a selected raga. The set of characteristic phrases, or motifs, of the raga collectively embodies the raga grammar and forms the building blocks for compositions and improvisation in the raga (Rao *et al.*, 2014). Given this, training in Indian classical music traditions involves acquiring a knowledge of raga motifs and the ability to recall and render the phrases in the context of a performance. Further, an affective quality, or musical meaning, is associated with each raga that may be considered to depend on the listener's acquired knowledge of the musical idiom (Asano and Boeckx, 2015; Widdess, 2013). A musical performance begins with the alap section where the performer strives to establish the identity of the chosen raga in the mind of the listener via one or more of its characteristic phrases embedded in melodic improvisation at a slow, irregular tempo. As the performance progresses through the faster and more rhythmic sections, the characteristic phrases recur with shape variations driven by the context, but still largely retaining their recognizability.

While a raga phrase is called by its underlying notes' sequence, the durations and intonations of the notes are not restricted to quavers, semi-quavers, etc., or naturals, sharps, and flats (van der Meer, 1980). Thus, the melodic shape of a phrase is better represented by a continuous pitch curve, which implicitly encodes the intonation and duration of its constituent segments, representing expression and relative emphasis of the different pitches (Rohrmeier and Widdess, 2012). Recalling the role of intonation contours in linguistics, we note that speakers of a language associate prototypical pitch contour shapes with specific utterance modes such as statement or question (Hirst and Di Cristo, 1998). Drawing parallels, it is entirely conceivable that in the raga music context, prototypical representations of raga-characteristic motifs stored in long-term memory are recalled by both performers and listeners as a consequence of the learned schema. The existence of phrase prototypes (Ps) would influence the perception of raga-characteristic phrases by trained musicians (TMs) in a manner that can possibly be verified by behavioral experiments involving acoustic variations of phrase shape. Further, it would be interesting to contrast this with perception in non-characteristic phrase contexts or perceptual experiments of the same material with non-musicians (NMs).

The applications of this work are closely related to those of the general notion of melodic similarity, where approaches that are informed by human similarity judgements are desirable. Computational modeling of melodic similarity at the local phrase level find applications in content-based MIR, including the structural segmentation of long musical pieces

[a)]Electronic mail: prao@ee.iitb.ac.in

and subsequent classification into tune families (Mullensiefen and Frieler, 2004; Vempala and Russo, 2012; Pearce *et al*., 2010; Allan *et al*., 2007; Novello *et al*., 2006). A plausible mode of listening involves simply recognizing recurrent phrases, as opposed to apprehending all the details of the melodic shape, to eventually make global similarity judgements (Marsden, 2012). This mode clearly holds for raga music in which the phrases serve as cues to raga identity and the associated semantics, including affect. It follows that an understanding of the perceptual discrimination of melodic variants of motifs can contribute to compositional aids and to the development of automatic feedback for pedagogy apart from being more generally useful in melody-based MIR tasks.

In Sec. II, we discuss previous perception studies involving music Ps. We also review perception studies in speech prosody that aid us in the formulation of experimental paradigms for our work with melodic phrase shapes. This is followed by a brief introduction of the raga music background and choice of stimuli for the perception experiments, presented subsequently, in this paper.

II. REVIEW OF PERCEPTION STUDIES

It is known that the perceived similarity of familiar sounds is influenced not just by low-level acoustic differences transformed to sensory representation differences, but also by the potentially inferred cognitive semantic information from long-term memory (Kuhl *et al*., 1992). This phenomenon has been convincingly demonstrated for speech sound categories, most notably involving phones at the ends of the voice onset time (VOT) continuum such as /b/ and /d/ where it has been termed "categorical perception" (CP; Liberman *et al*., 1957; Goldstone and Hendrickson, 2010; Pisoni and Lazarus, 1974; Repp, 1984). First reported by Liberman *et al*. (1957), it was seen that the physical space between stimuli on a single-dimensional continuum mapped non-uniformly to the perceptual space. The extent of warping depended on the stimuli location with respect to the phoneme category boundary on the continuum. Iverson and Kuhl (1995) demonstrated the similar phenomenon in the case of isolated vowel tokens where a strong relationship was observed between the category goodness of a vowel token and discrimination between acoustic variations in its vicinity. Termed the "perceptual magnet effect" (PME), this implied that phonetic category information is stored in memory, acting like a perceptual attractor with decreased discrimination sensitivity in the physical space immediately around it.

Both phenomena, PME and CP, more generally refer to the enhanced within-category similarity and enhanced between-category differences attained in the process of category learning. Identification scores of listeners subjected to randomly ordered stimuli drawn from the acoustic continuum between the two categories show a relatively abrupt shift at the category boundary. Further, the discrimination of stimuli equally spaced on the continuum is highest near the category boundary and can be predicted from the observed identification functions (Burns and Ward, 1978). There exist categories in the context of music, too, that can be considered to be acquired by learning such as musical pitch and instrument timbre (Cutting and Rosner, 1974). As in speech perception, learning and musical experience would be expected to play a role over simple auditory coding in detecting physical differences (Aaltonen *et al*., 1987). Music constitutes an interesting test case because it is less likely than speech to have inborn feature detectors already "prepared by evolution" (Harnad, 2003). Burns and Ward (1978) investigated the perception of melodic musical intervals by musicians via identification and discrimination functions for a set of stimuli (a stimulus being a sequence of two tones representing a musical interval). Musical intervals of different widths were observed to be perceived categorically by the musicians but not by musically untrained listeners.

In a similar study of the CP of tonal intervals by Siegel and Siegel (1977), it was found that musicians who could accurately label each of three musical intervals were surprisingly poor at discrimination within a musical category, i.e., they could not reliably tell *sharp* from *flat*. A few others have noted top-down effects of musical expectancy interacting with lower perceptual processes in the context of major chords (Barrett, 1999; Acker *et al*., 1995; McMurray *et al*., 2008). Acker *et al*. (1995) found that musicians are particularly sensitive to deviations of a note position within a chord from its nominal position, leading to the observation that musical expectancy actually *narrows* a category. This phenomenon where discrimination became sharper near the P of a C major chord was confirmed by Barrett (1999) on a group of professional musician participants. However, the same experiment with amateur and NM participants demonstrated instead the PME, known to be associated with learned categories. The unexpected heightened sensitivity of professional musicians in the vicinity of the chord P was explained on the basis of their acquired auditory skill of recognizing in- and out-of-tune chords, so necessary in their work.

It may be of interest to note that the similarity of short melodic fragments, represented by note sequences, has also been studied to understand variations that are most predictive of human judgement in a task that involves short-term memory rather than category-based effects. In this context, Fiske (1997) argued for three decision levels defining the degree of difference in any inter-pattern relationship: (i) same, (ii) derived, and (iii) distinctly different. The corresponding subjective listening test incorporated tonal-duration manipulation on 40 pairs of tonal-rhythmic patterns of electronically synthesized tones with flute-like timbre (5–8 s in duration separated by 3 s). While tempo change and transposition were considered "same," melodic and rhythmic ornamentation were mostly considered "derived." Vempala and Russo (2015) tested both single-note and multi-note pitch modifications in short melodic segments via human similarity judgements of the reference and modified melody on a five-point rating scale. They found that melodic contour (direction of pitch change) was overall the strongest predictor of the human similarity judgements.

The raga motif context, with its continuous pitch contour (rather than a sequence of notes of discrete pitch and duration), perhaps relates more closely to intonation contours in speech. Further, the distinct shapes of intonation

contours are linked to different linguistic meanings, akin to the learned semantics of raga phrases. Ladd and Morton (1997) carried out identification and discrimination experiments with native listeners on short English utterances using stimuli across the continuum of normal to emphatic intonation contours. They demonstrated evidence of categories by noting abrupt shifts in identification scores, but observed only limited differences in discriminability within stimulus pairs inside and across the category boundaries. More recently, Rodd and Chen (2016) demonstrated the existence of PME in the perception of Dutch pitch accents where participants rated the "goodness" of stimuli relative to category belongingness, and also discriminated between stimuli separated by a fixed acoustic distance in a manner related to the distance of the stimuli from the category P. Schneider *et al.* (2009) considered intonation contours corresponding to question and statement categories in German with their high and low boundary tones, respectively. Perception experiments with a range of equispaced stimuli between the high and low tones demonstrated PME only in the case of the statement category but not the question category.

A different experimental paradigm based on production, rather than listening, was introduced to test for CP in speech intonation by Pierrehumbert and Steele (1989). 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 participants are expected to imitate the continuum correctly. When perception is categorical, however, the participants' repetitions should fall into two clearly separable categories with respect to the manipulated feature, which indeed was the case in the experiments on pitch accent categories in American English. A few other studies as well have used the production or "listen and imitate" method successfully to confirm the existence of intonational categories. Redi (2003), again in a pitch accent experiment, reported that the participants did not repeat all points in the continuum but they seemed to have two distinct pitch peak delay categories in mind, which they reproduced. She also argued that this imitation task is a much better design to test for the CP of intonation rather than the classical listening test because the participants' intuition is tested implicitly.

In summary, in the context of music, while tonal intervals and chords have been the subject of CP studies, there is no similar work on melodic motifs. Raga music lends itself to the latter given the existence of learned schema involving motifs. Tonal intervals presented to Western musicians elicited the PME while, in the case of chords, TMs displayed the opposite, i.e., a heightened sensitivity in the vicinity of the prototypical chord. Given the observed divergences, the perception of raga motifs by musicians trained in the genre is worthy of investigation, and the experimental methods of speech prosody perception appear well suited for this task.

III. MUSIC BACKGROUND

Melodic organization in Indian art music is governed by the system of ragas, a modal concept between a scale and a tune (Powers and Widdess, 2001). Raga grammar specifies the tonal material in terms of allowed notes and their hierarchy as well as the small set of raga-characteristic phrases. The phrases are the prescribed note sequences, each of which is associated with a distinctive melodic shape described by a continuous pitch versus time curve. The absence of symbolic notation to represent the characteristic melodic shape has not been an obstacle to the transmission of melodic compositions through the generations in this oral tradition. Learning takes place implicitly through listening and reproducing raga-specific materials presented by the teacher rather than through explicitly articulated rules (Widdess, 2013).

While there are hundreds of ragas, a musician's repertoire would typically comprise about 50 of the most popular ragas. A musical performance is based around a chosen raga, and comprises known compositions accompanied by significant improvisation. Both, strict adherence to the raga grammar and creativity in improvisation are accorded the highest importance. In view of this flexibility in practice, it is interesting to note that perceived deviation from the raga grammar is associated with at least one melodic feature of the performance being actually suggestive of a different raga (Vijaykrishnan, 2007; Raja, 2016; van der Meer, 2008; Kulkarni, 2011); that is, when a feature that is specified in the raga grammar takes on a value that is sufficiently different from its nominal value so as to be suggestive of a different raga. Certain closely spaced ragas in melodic features' space are the "allied ragas," which have identical scales but differ in other aspects such as the hierarchy of notes and characteristic phrases. Learners are introduced to members of an allied raga pair together and warned against confusing the two ragas (Kulkarni, 2011; Autrim-NCPA, 2017; Bagchee, 2006).

Raga-characteristic phrases appear repeatedly in concerts, constituting powerful cues to raga identity. While the melodic shape of the motif can also be influenced by the local context, such as the tempo, one or more relatively invariant features serve as cues, contributing strongly to its recognizability by the listener (Ganguli and Rao, 2017). In the present study, we choose a raga phrase in which controlled acoustic variations in a single feature can be introduced to create ecologically valid stimuli for our identification and discrimination experiments. As described in the remainder of this section, we consider a particular phrase of a popular raga, Deshkar. A statistical study of the variations in the melodic shape of the selected phrase extracted from an audio corpus of concerts forms the basis of the stimulus design for the perception experiments.

A. Raga grammar

Table I shows the relevant aspects of the grammar of raga Deshkar, as presented in music texts, together with that of its allied raga, Bhupali. The two ragas are the most widely known pentatonic ragas, sharing the scale of five notes (see the solfege in Fig. 1 for reference). Between the two ragas, the hierarchy of notes differs and so do the characteristic phrases as indicated by the different note sequences. We select the DPGRS, a phrase that represents a common descending movement in both ragas and constitutes a

TABLE I. Notes and phrases, separated by commas, of the two allied ragas as compiled from musicological texts (Kulkarni, 2011). Overline/underline indicates higher/lower octave.

Grammar	Deshkar	Bhupali
Tonal material	SRGPD	SRGPD
Characteristic phrases	SGPD, P(D)$\bar{S}$P D G P, D P G (R)S	R$\underline{D}$S, RPG, PD$\bar{S}$ $\bar{S}$DP, GDP, GRS

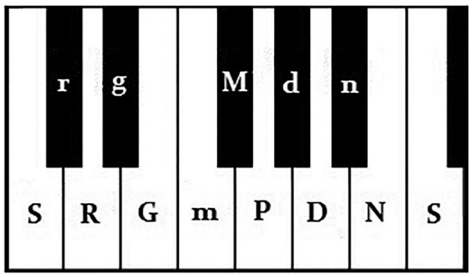

FIG. 1. The chromatic solfege of Hindustani music shown with an arbitrarily chosen tonic (S) location.

recurring component in performances. The melodic shapes of the phrase, in particular the GRS motif, however, differ significantly between ragas. The R note in Deshkar is marked by the parentheses in Table I to indicate that it is de-emphasized relative to neighboring notes G and S, each separated from R by 2 semitones (200 cents). The de-emphasis is realized by the temporal duration as signified by the word "alpa" (meaning "little") used to describe it in musicology texts (van der Meer, 1980; Kulkarni, 2011; Bagchee, 2006; Autrim-NCPA, 2017). In Bhupali, on the other hand, there is no such differential emphasis of the notes in the phrase.

Figure 2(a) shows the continuous pitch contour of a sample DPGRS phrase extracted from a Deshkar concert. We can see the R note (at 200 cents) appears like a narrow step in the transition from G to S. Figure 2(b) shows zoomed-in sub-segments corresponding to the transition from G to S extracted from multiple instances of the phrase from a single concert, indicating the limited extent of variability in R duration and the invariant nature of the transitions into and out of the R note.

B. Data-based characterization

In order to design stimuli that are ecologically valid, we carry out a corpus-based statistical study of the musical attributes of interest in the selected phrase. To capture the relative emphasis on the note R, the durations of G, R, and S are measured on a large set of DPGRS phrases obtained from manually annotated concert recordings in Deshkar and Bhupali ragas by famous vocal artists drawn from the Hindustani music dataset of the Dunya[1] corpus. The same dataset was recently used to investigate the distributional properties of the ragas (Ganguli and Rao, 2018). A trained Hindustani musician marked all occurrences of the desired phrase in each recording. The rest of the processing leading to the note duration measurements is automatic as described next.

A predominant $F0$ detection algorithm is used to detect the vocal pitch at 10 ms intervals throughout the audio recording in the automatically detected singing voice regions (Rao and Rao, 2010). Hindustani vocal music contains the constant drone as well as the percussive tabla accompaniment. The detected vocal melody is normalized from Hz to the cents scale with reference to the detected tonic, determined using a classifier-based multi-pitch approach to tonic detection (Gulati *et al.*, 2014). A post-processing step involves interpolation of short devoiced intervals (empirically chosen to be $\leq$80 ms) caused due to unvoiced consonants (specially in regions with lyrics in contrast to the melismatic vowel singing).

The musician annotated regions corresponding to the DPGRS phrase are extracted for further segmentation into the constituent note events of interest (i.e., G, R, and S) as follows. The onset/offset of a note is defined as the beginning/end of a continuous region within $\pm$35 cents of the nominal $F0$ of the note (Ganguli *et al.*, 2016). Very short duration excursions (defined as segments of less than 250 ms with $F0$ lying outside the tolerance region) are detected and included in the surrounding "stable" region. This effectively takes care of naturally occurring ornamentation, and $F0$ dips in the continuous pitch contour of the realized note due to the occurrence of voiced consonants.

Figures 3(a) and 3(b) show examples of the raga motif pitch contours with detected note events superposed. The distributions of the measured durations of the note events across 110 phrases of Deshkar and 188 phrases of Bhupali

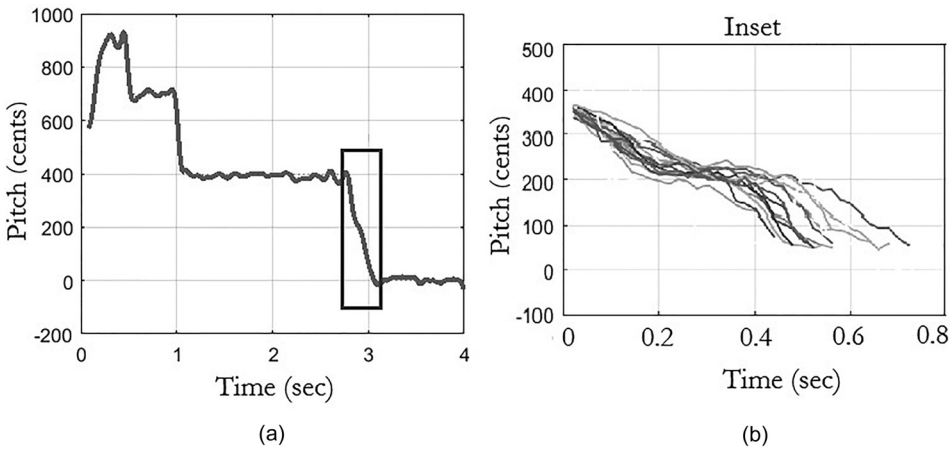

FIG. 2. (a) Representative melodic contour of a raga Deshkar DPGRS phrase (a phrase that represents a common descending movement in both ragas and constitutes a recurring component in performances; S is at 0 cents); (b) zoomed view of the inset from time-aligned G offset to the S-onset for 15 phrases from a single concert.

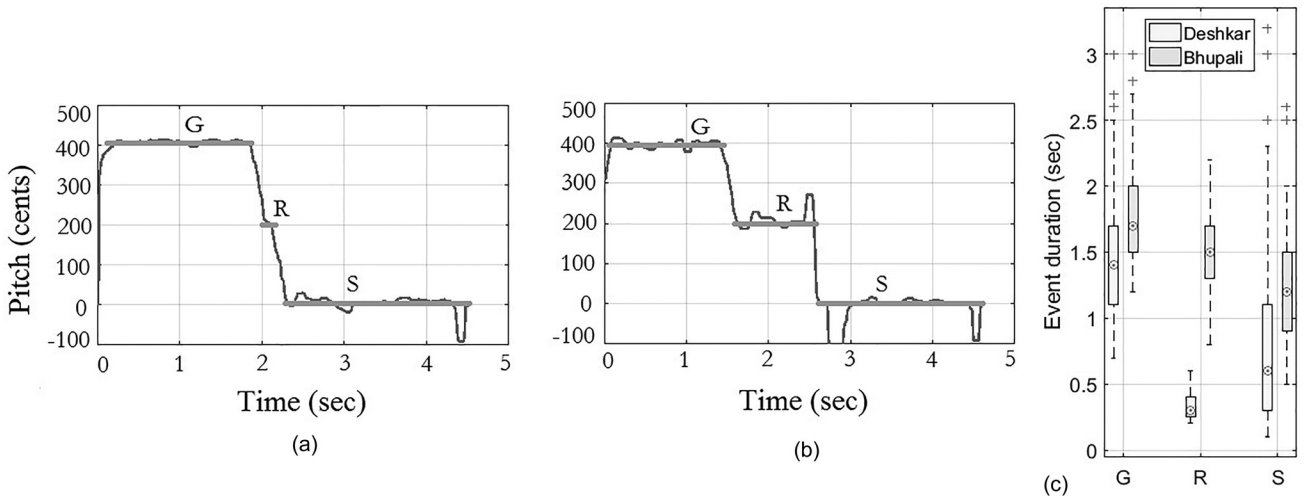

FIG. 3. Pitch contours of arbitrarily chosen GRS motifs from ragas Deshkar (a) and Bhupali (b). The horizontal lines indicate the segmented G/R/S notes; (c) distribution of constituent note durations of the annotated phrases from the corpus (110 phrases of raga Deshkar and 188 phrases of raga Bhupali).

are shown in Fig. 3(c), extracted from 12 concert recordings. All the measured phrases come from the initial to mid regions of the concerts corresponding to the slow and medium tempo regions where the recognizability of the phrases is highest (van der Meer, 1980; Ganguli and Rao, 2018). The S note is the final resting note of the phrase and hence its duration distribution is particularly wide (Rao *et al.*, 2014). The spread in the duration of a note can be attributed to the influence of local context such as the underlying local tempo and the proximity to the boundary of a rhythmic cycle (van der Meer, 1980). Even so, we observe a clear difference in the realization of the duration of the note R across the two ragas with the Deshkar R being narrowly constrained and seemingly insensitive to contextual effects. Increasing the relative emphasis of the R note by extending its duration would make the melodic shape more like that of a Bhupali phrase. Thus, the R-note duration appears to constitute a cue to the raga category, making it a suitable candidate for testing phrase perception in the acoustic continuum between categories.

In the perceptual experiments of this paper, the DPGRS phrases corresponding to ragas Deshkar and Bhupali constitute the two distinct categories. The stimuli for the experiments are drawn from the temporal continuum between these in terms of the R-note duration. Further, to study the perception of similar acoustic variations in a different melodic context, we also consider the same physical variations of R note in a different phrase context, viz., DPMGRS which occurs widely as a descending movement in many (non-pentatonic) ragas but does not constitute a characteristic phrase in any raga.

IV. GENERAL METHODS

The questions addressed in this paper concern the existence of raga phrase Ps and whether these serve as perceptual magnets or anchors, with reference to perceptual discrimination by trained listeners, of acoustic variations in a characteristic feature. As indicated by the raga grammar, and supported by the corpus-based observations, the characteristic melodic feature in the chosen phrase is the R duration, signifying relative note emphasis. The box-plots of Fig. 3(c) help us to design a stimulus set that spans the most commonly observed realizations of the chosen phrase in practice. The perceptual experiments are intended to help us to (i) verify our hypothesis about existence of a P and PME around the Deshkar motif shape, and (ii) examine whether, indeed, another perceptual category exists at the extreme end of R note elongation as implied by the box-plots of Bhupali duration features in Fig. 3(c). In this section, we present the components that are common across our subjective experiments, viz., the generation of audio stimuli, choice of participant groups, and the overall experimental procedure.

A. Stimulus creation

Given that melodic shape is the focus of the perception experiments, differences in timbre and loudness need to be suppressed in the stimuli. This is achieved by using the $F0$-contour extracted from a reference instance as the base material to synthesize stimuli for the experiments. Certain signal processing operations are needed further to prepare a suitable stimulus set for the behavioral experiments. These steps involve: (i) stylization of the reference pitch contour, (ii) imparting controlled distortions to create the required range of stimulus pitch contours, and (iii) audio synthesis from each of the pitch contours to obtain the set of natural sounding stimuli for the perception experiments.

1. Pitch contour stylization

By stylization, we retain the essential melodic shape of the motif while eliminating artist-dependent vocal variations such as note embellishments. Stylization first reduces the pitch contour of the GRS segment to a sequence of melodic events each replaced by its elemental form, e.g., constant-$F0$ segment for a note and a smooth curve for a transition. Next, controlled variations are introduced to simulate the involuntary perturbations occurring in singing that contribute to the naturalness of the sound. A more detailed description follows.

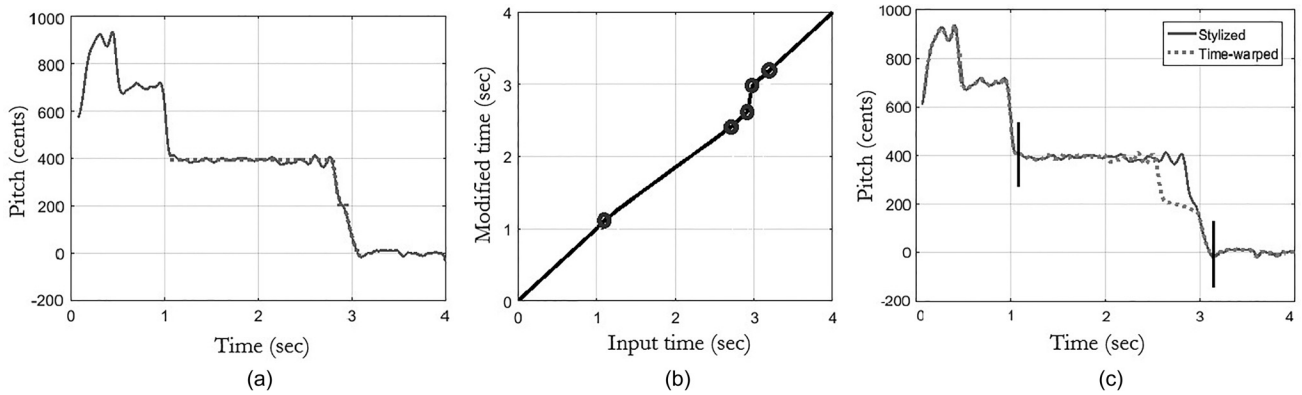

FIG. 4. Pitch contour stylization and stimulus contour generation: (a) reference phrase shape and semi-stylized contour (dashes) for the segment of interest (G-onset to S-onset); (b) piecewise linear mapping function to obtain the time-warped pitch contour; (c) stylized and modified pitch contour, after expanding the R duration by a scale factor of 2.5. Vertical lines in (c) indicate onsets of G and S notes, which coincide with the metronome locations in the synthesis (see Fig. 5).

The chosen reference phrase is an actual instance with event durations close to the medians of the distributions in Fig. 3(c). The extracted phrase is segmented into stable note and transitory segments as presented in Sec. III B. We next replace the transients with a third-degree polynomial to capture just the essential shape (Ganguli and Rao, 2015; Ganguli *et al.*, 2017). Figure 4(a) shows the G and R notes, together with the fitted transient segments GR and RS, in dotted horizontal lines over the reference contour. This semi-stylized DPGRS pitch curve is used to generate the required set of stimuli, each with a different R duration.

2. Imparting controlled distortions

The stimuli are derived from the stylized pitch contour of the reference phrase by modifying the relative durations of the G and R notes. This is carried out by a process of non-uniform time warping that keeps the durations of each of the transient segments, as well as the total duration of the G-onset to S-onset, unchanged. The piecewise linear mapping is with reference to certain anchor points in the G-onset to S-onset region.

Figure 4(b) shows a mapping from reference to modified phrase time axes that is designed to expand the R segment by a factor of 2.5. The circles indicate the following anchoring instants from left to right: G-onset and -offset, R-onset abd -offset, and S-onset. Figure 4(c) shows the resulting modified contour with the reference stylized contour, where we note the expanded R duration and the corresponding contracted G segment. The vertical bars indicate the onsets of G and S notes. Modified pitch contours are constructed similarly for the set of stimuli required for the perceptual experiments. Next, the naturalness of each stimulus is increased by introducing random perturbations in the steady regions via the addition of white Gaussian noise of the same variance as that measured in the corresponding reference phrase region. Finally, the oscillations, representing natural overshoot and undershoot, close to the G note boundaries (up to 200 ms) as extracted from the actual reference pitch contour, are restored in the synthesized contour by the overlap-add of pitch samples.

3. Audio synthesis of stimuli

Another important aspect is the audio rendering of the stimuli for presentation to the participants. A complex tone with neutral vowel-like timbre is synthesized with its $F0$ varied according to the required pitch contour. Five harmonics with relative weights chosen as 0, 3, 5, −6, and −20 dB comprise the tone with overall intensity of the tone maintained constant. We add a good quality recorded drone (tanpura) at the tonic $F0$ as background. As a timing reference, metronome clicks are added at equal intervals with a click at each of G-onset and S-onset, indicating the underlying beat structure. The spectrogram representation of the synthesized and tanpura/metronome-added reference phrase (spanning 6 s) is shown in Fig. 5. The duration from G-onset until S-onset (roughly the combined duration of G and R notes together with the connecting glides) is fixed at 2.2 s.

B. Participants and procedure

Table II describes the distinct participant groups across the three experiments, the first two of which involve listening to a prompt followed by rating it. The third experiment involves reproducing a presented prompt by singing it.

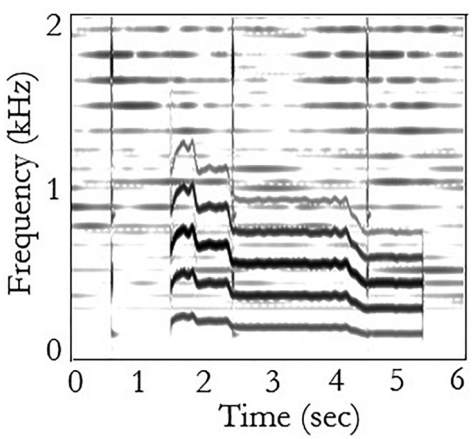

FIG. 5. Spectrogram of the synthesized audio from the stylized reference contour; horizontal lines are the tanpura harmonics, while vertical lines are the metronome clicks.

TABLE II. Summary of participant groups for the perception experiments. TM, trained musicians; NM, non-musicians; IP, Indi-pop singers.

Index	Experiment type	Trained	Untrained	
		TM	NM	IP
1 (a)	PME goodness rating	23	—	—
(b)	PME discrimination	28	15	—
2 (a)	CP identification	23	—	—
(b)	CP discrimination	23	15	—
3	Listen and imitate	24	—	24

A total of 28 Hindustani musicians (13 female) with average age of 31 years (standard deviation, SD = 4.5) and average years of training of 15 years (SD = 2.8) participated in experiment 1. Ten of them were instrumentalists, having trained in the sitar, mohan veena, or flute. All the participants had had vocal training as well, and had formally learned both the ragas, Deshkar and Bhupali. Five of them had more than 8 years of experience teaching music. A subset of 23 musicians took part in experiment 2. Additionally for the discrimination tasks, we had a category comprising NM participants. We had 15 NMs (6 female) who did not normally sing or play an instrument and had had no training in any genre (labeled "NM" in column 5 in Table II). These participants listened to popular music but did not have any exposure to Hindustani raga performances. Twenty-four Hindustani musicians (7 instrumentalists), partially overlapping with the previous set, participated in experiment 3. A newly added participants' category for this experiment is the untrained singers category, consisting of 24 singers, not formally trained in Hindustani music but known to be good singers of popular (Indi-pop) songs, which they rendered by imitation.

The participants listened to the stimuli via headphones in a quiet laboratory environment. A participant's tasks across the first two experiments were the following: (i) to listen to the presented single stimulus, and label it using the provided label set, and (ii) to listen to a pair of stimuli separated by a short interval, and indicate whether the two stimuli are the same or different. The experiment was locally hosted on the Sonic Mapper (Scavone *et al.*, 2002) software and self-paced. Sonic Mapper facilitates listen and rate perception experiments with its built-in features for stimulus randomization, experimenter-controlled timing and recording of responses based on a selected label set or rating scale. Further, any presentation of a stimulus (or stimulus pair) can be replayed by the participant before the rating step. The number of plays is recorded in the interface and can serve as a proxy for listening effort for the particular participant and stimulus (pair) combination. At the start of the session, written instructions are provided to ensure that all participants received identical information. A practice session was presented before the rating task with three stimuli (or pairs as applicable), which represent the diversity of the model space. However, there is no explicit mention of the physical differences between stimuli.

A subset of the participants in both musician and NM categories in the discrimination tasks (experiments 1b and 2b) volunteered to be control participants. The control participants rate both AB and BA pairs allowing "order of presentation" effect to be examined here. None of our participants was paid for their participation in the experiments. The third experiment tests for the existence of categories by asking the participant to vocally imitate each prompt drawn in random order from the set of stimuli.

V. EXPERIMENT 1: TESTING FOR THE PME

With experiment 1, we investigate the hypothesis of a PME in raga phrase perception. This requires first identifying a P and a non-prototype (NP) exemplar of the phrase, around each of which perceptual discrimination can then be tested. We consider the DPGRS phrase of raga Deshkar with the specific melodic feature of R-note duration. The participants are asked to rate the goodness (i.e., belongingness to raga Deshkar) of each stimulus within the range of R-duration variation. Of the stimuli that are thus determined to be associated with raga Deshkar, we identify the best and worst rated ones as the P and NP, respectively. These are used next in the construction of stimulus pairs representing a range of physical separations with respect to either the P or the NP stimulus. If indeed PME occurs, we expect to see significantly different levels of discrimination in the two conditions.

A. Method

The PME paradigm consists of a goodness rating and a subsequent discrimination task (Acker *et al.*, 1995; Schneider, 2012; Rodd and Chen, 2016).

1. Goodness rating

The task, administered to 23 TMs, was to evaluate the quality of the presented stimulus with reference to the raga Deshkar phrase on a predefined scale from −3 (very bad exemplar) to 3 (very good exemplar). The question posed is "How good is the phrase as an example of raga Deshkar?" Note that textual descriptions were provided for only the two extremes (—3, very bad and 3, very good) and the mid-point (0, neutral) of the rating scale.

During the test, the 13 stimuli of Table III were repeated twice in randomized order within a trial block. Two trial blocks were presented to each participant with a minimum gap of 1h between blocks. Listeners had to select one of the rating values before they could proceed to the next stimulus. They were explicitly asked to use the full range of the rating scale as far as possible. Each stimulus is 4 s long, which added with the extended tanpura background, amounts to 6 s. Each rating takes no more than 10 s (assuming single play, and considering the complexity of the Likert-type scale over a binary choice). This adds up to $26 \times 10 = 260$ s $\sim$ 4:30 min per trial block. If the number of plays is higher (up to three or four for confusing stimuli), the total time taken for a trial block is found to be no more than 12 min.

The goal of this experiment was to identify a P and a NP representation of the raga Deshkar motif from among the presented stimuli to use in the discrimination experiment.

TABLE III. Stimulus description in terms of index, scaling factor, and absolute duration of the R note for experiment 1. All stimuli from 1 to 13 are used in experiment 1a and the stimuli 5–11 are used in experiment 1b.

Stimulus number	Scale factor with respect to reference	Absolute R duration (s)
1	0.01	0.003
2	0.25	0.07
3	0.5	0.15
4	0.75	0.22
5	1	0.3
6	1.5	0.45
7	2	0.6
8	2.5	0.75
9	3	0.9
10	3.5	1.05
11	4	1.2
12	5	1.5
13	6	1.8

The P would correspond to the token receiving the highest goodness rating, while NP would be one that receives the lowest rating but can still be considered to belong to raga Deshkar.

2. PME discrimination

The discrimination task had 15 NMs participating in addition to 28 TMs. In this task, a pair of stimuli, separated by 500 ms as mentioned in Sec. IV B, was presented to elicit the judgement of whether the two stimuli were the same or different. One member of each pair was one of the P or NP stimulus determined via the goodness rating experiment. The other member of the pair was chosen from the stimulus set indicated by the middle box of Table III, so that the acoustic difference within the pair was between one to six steps. Using the convention of an AB pair comprising two stimuli in increasing order of stimulus index, we create six AB and six BA pairs out of the seven stimuli with reference to the P and the same number with reference to the NP. A matching total number of identical pairs (12) is then included for each of P/NP to avoid any bias. This gives us a total of 192 pairs in the context of 2 repetitions per trial block and 2 blocks per participant.

In the interest of limiting the overall test duration, only a subset of 12 TMs was administered the full set of 192 stimulus pairs. The remaining participants (of the $28 + 15$ as in Table II) received a balanced mixed of AB/BA and AA pairs with random selection as executed by the Sonic Mapper software for a limited set of 96 stimulus pairs.

Each stimulus is 4 s long, added tanpura background results in ~6 s. So each pair takes 12 s + inter-stimulus interval (ISI) $0.5\,\text{s} = 12.5\,\text{s}$. If each pair is presented only once, the time taken for each pair is 15 s (from start of play to the start of next pair). Thus, the entire set takes $96 \times 15 = 1440\,\text{s}$ or 24 min. Even if the number of plays is higher (say, up to four times for confusing pairs), the total time taken for a trial block is less than 30 min. The two trial blocks are separated by at least 1h. The non-control participants take less time.

B. Results and discussion: Goodness rating

Most participants used at least six of the proposed seven levels of the rating scale. However, four participants used only four values of the scale. Nevertheless, all participants used rating values from the upper half of the scale (1–3) showing there were enough stimuli perceived as good examples of raga Deshkar.

Figure 6 shows the mean rating for each stimulus condition across all trials of the 23 Hindustani musicians. Thus, mean for each of the 15 stimuli is contributed from $23 \times 4 = 92$ ratings. The lighter plot shows the response time (in terms of number of replays of the prompt). We see that stimulus 5 achieves the highest mean rating for goodness with a gradual fall on either side. The number of replays is lowest in the vicinity of this best rated stimulus and rises rapidly as the judged quality of the stimulus decreases. The stimulus 5 corresponds to an R absolute duration of 0.3 s (see Table III) which, incidentally, matches the median value of the R-duration distribution derived from the corpus as in Fig. 3(c). We therefore choose stimulus 5 as the P for the raga Deshkar motif. As for choosing the NP version, we face the following considerations. We require the NP to be recognizable as a raga Deshkar phrase while being separated from the P so that a sufficient number of stimuli are available between P and NP for the subsequent discrimination experiment.

From Fig. 6, the stimuli in the index range 8–11 potentially qualify as NP versions of the motif given the lower goodness ratings and relatively high number of replays (indicating the participants' confusion). Informal comments by the participants revealed that the stimuli beyond index 11 were clearly suggestive of raga Bhupali. Accordingly the stimulus 11 is chosen as the NP version of raga Deshkar. Table III, middle panel, shows the model space for the discrimination test, presented next.

C. Results and discussion: PME discrimination

We report results of stimulus pair–based discrimination where one member of the pair comprises either the P or NP

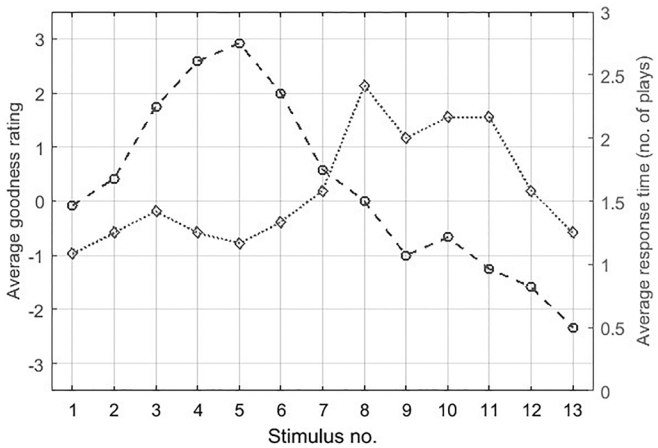

FIG. 6. Average across listeners and trials of goodness ratings by 23 trained musicians (TMs). The lighter curve shows the average response time (in terms of number of repetitions for each stimulus).

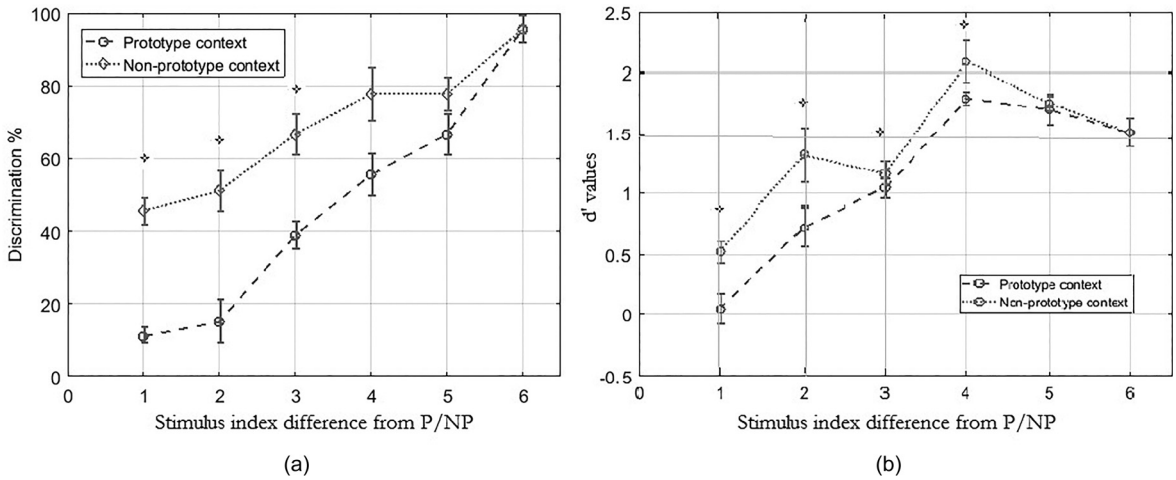

FIG. 7. (a) Percentage discrimination and (b) d' values averaged over participants and trials in the vicinity of P/NP for the TMs' group; "*" indicates that the difference between the corresponding P and NP contexts is significant at threshold of 0.01; $p = (6 \times 10^{-16}, 0.0001, 0.009, 0.048, 0.063, 0.1)$ for discrimination scores, and $(3 \times 10^{-8}, 0.0001, 0.009, 0.0006, 0.049, 0.1)$ for d'.

as determined from the previous goodness rating experiment. Both musicians and NMs participated in this experiment. The measured discrimination can serve to quantify the perceptual distance between stimuli.

1. Observations

It is important to consider the effect of order of presentation in the context of stimulus pairs. In perceptual discrimination studies with speech stimuli, pairs with the more prominent stimulus occurring second in the pair are more discriminable, where prominence is defined as produced with more effort (Batliner and Schiefer, 1987). Order of presentation dependence was studied with our control participants who did both AB and BA versions of every pair of distinct stimuli. We found that while the means of the BA (i.e., the lower R duration comes second in the pair) were the same or marginally higher than those of AB for the corresponding condition for the musicians, none of the differences across the six stimuli pairs was significant (at threshold of 0.01) in either participant category. We had $p = (0.085, 0.092, 0.062, 0.012, 0.021, 0.028)$ for P context, and (0.084, 0.081, 0.069, 0.076, 0.031, 0.028) for NP context in a two-sample t-test; Welch, 1947). Based on this observation, we combined the AB and BA pairs for a given stimulus pair to obtain averaged results.

Figure 7(a) shows the discrimination performance with increasing acoustic difference from either P or NP as selected from the goodness rating experiment by the TMs' group. The error bars are computed from all ratings (28 participants $\times$ 2 trial blocks $\times$ 2 repetitions) per stimulus. We observe that with increasing distance from either of P or NP, the mean discrimination performance improves, as expected. Further, the discrimination one stimulus step away from P is poorer ($\sim$10%) than that of NP ($\sim$50%), and the difference between the two contexts gradually decreases. There is significant difference between P and NP neighborhoods up to four stimulus steps. That is, the neighbors of P are always discriminated significantly worse than the similar neighbors of NP. From the NMs' response, on the other hand, as shown in Fig. 8(a), both P and NP neighbors show similar discriminability at the lower acoustic differences. For the higher acoustic differences, the results are less consistent.

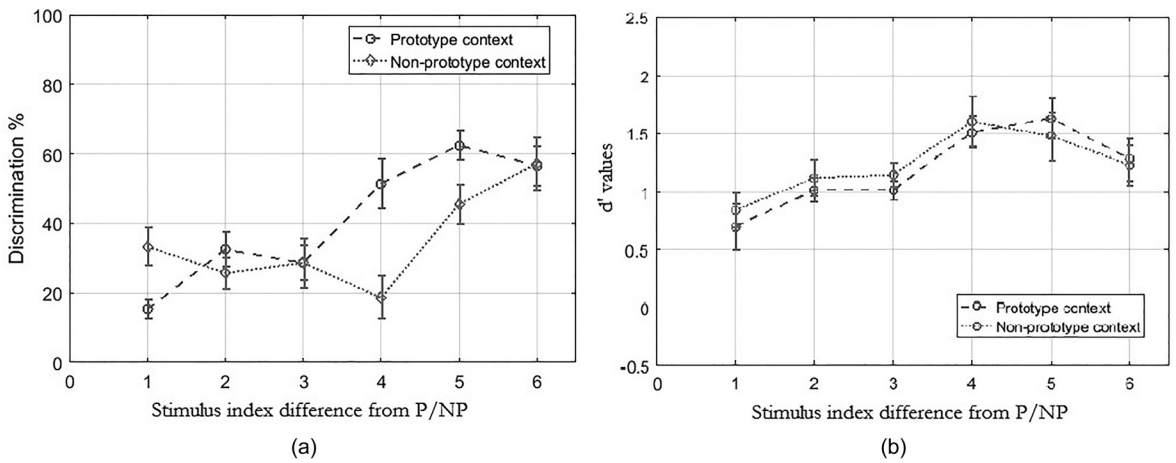

FIG. 8. (a) Percentage discrimination and (b) d' values averaged over participants and trials in the vicinity of P/NP for the NMs' group. None of the differences between the corresponding P and NP contexts were found to be significant at threshold of 0.01.

2. Discrimination sensitivity

In perception experiments, participants almost always differ from each other in identification and/or discrimination performance. According to signal detection theory (SDT; Egan, 1975; Stanislaw and Todorov, 1999), listeners who share the same perceptual precondition (identical auditory threshold) can produce different results in a perception test. This is because the percentage discrimination (or the hit rate) is influenced by two factors, viz., the participant-dependent response bias and the sensitivity or "perceptual distance" between stimuli (Wickens, 2002). It is the latter, represented by d', that we are interested in comparing across the different stimulus pair conditions and participant groups.

SDT explains participant responses as a combination of sensitivity index d' and response bias λ. On any trial in a discrimination task, the rating is "different" when the evidence for the signal (that is, the acoustic difference between the stimuli) is larger than the individual response criterion λ, and "same" when it is smaller. Unlike the percentage discrimination, the value of d' does not depend upon the individual criterion, but instead is a true measure of the internal response. The d' is calculated by taking the Z-transformations (assuming Gaussian distribution) of the hit rates (h) and the false alarm rates (f) by the following equation:

$$d' = Z(h) - Z(f). \quad (1)$$

Since there are no Z-transformation values for arguments corresponding to 100% or 0%, we modify the experimental outcomes as recommended by Schneider (2012). All values above 99.9% were fixed to 99.9% and all values below 0.1% were fixed to 0.1%.

Figure 7(b) shows the distribution of d' values obtained across trials and musician participants. Differences between the means of the P and NP contexts for the same acoustic distance are found to be significant ($p < 0.01$) for differences up to, and including, four stimulus steps. In the case of Fig. 8(b) for the NM participants, none of the differences between corresponding P and NP means were found significant. Our results thus support the hypothesis that perceptual distance is smaller for the same acoustic distance in the vicinity of the P relative to the vicinity of the NP for the TM participants. This indicates that the P serves as a perceptual attractor.

VI. EXPERIMENT 2: TESTING FOR CP

Post-experiment feedback by the participants indicated that the stimuli toward the longer end of the expanded continuum in the goodness rating task of Table III actually evoked a sense of raga Bhupali. This suggests the existence of a category at each end of the R-duration continuum. This motivates the next experiment to test for the CP of the DPGRS phrase shape. The low R duration corresponds to raga Deshkar while the high R duration corresponds to raga Bhupali. The occurrence of CP would be indicated by greater perceptual discrimination between stimuli closer to the category boundary.

The stimulus continuum is derived from the PME goodness rating stimuli as shown in Table III, but with adjusted end points and resampling the space with equal intervals between the P of raga Deshkar (scale factor = 1) to the stimulus with scale factor = 6. This range was then divided into 11 equally spaced steps in an arithmetic progression with a step size of 0.15 s, spanning the continuum from raga Deshkar to Bhupali consistent with the corpus-based observations of Fig. 3. The resulting model space is presented in Table IV.

A. Method

The CP paradigm consists of an identification and a discrimination task. Twenty-three TMs participated in both tasks. The identification task demands raga knowledge. The NM participants' group participated only in the discrimination task.

To study the effect of context on the melodic shape variations, the identical CP discrimination task is carried out with a non-characteristic phrase. A group consisting of 16 TMs (a subset of the participants of the experiment 2b in Table II) participated in this experiment. The reference stimulus now is the descending DPMGRS, which is not a characteristic phrase of any particular raga. Figure 9 illustrates its comparison with the DPGRS reference phrase. We simulated this non-characteristic context by recording a trained Hindustani musician with timing that matched the phrase segment DP of the original raga Deshkar reference phrase. Next, the stylized contour of the DPMGRS phrase was obtained by replacing the DP segment in the characteristic phrase reference contour with the new DPM segment.

1. Identification

Each of the 11 stimuli of Table IV, which differed in the R duration, was presented to the participant for two-way classification as Deshkar or Bhupali. Each stimulus within a trial block was repeated two times, and presented in a randomized order to receive a reliable set of results, which could be analyzed statistically. Another trial block was presented with the same set of stimuli but with a different randomized order. Thus, each participant labeled 44 stimuli

TABLE IV. Stimulus description in terms of index, scaling factor, and the absolute duration of the R note for experiments 2 and 3.

Stimulus number	Scale factor with respect to reference	Absolute R duration (s)
1	1	0.3
2	1.5	0.45
3	2	0.6
4	2.5	0.75
5	3	0.9
6	3.5	1.05
7	4	1.2
8	4.5	1.35
9	5	1.5
10	5.5	1.65
11	6	1.8

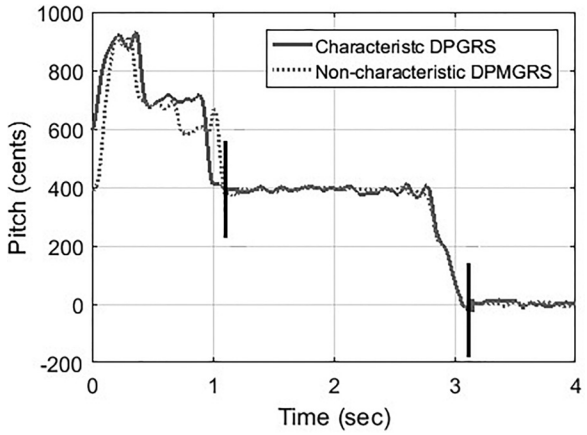

FIG. 9. Stylized pitch contours of corresponding stimuli, one from each phrase context.

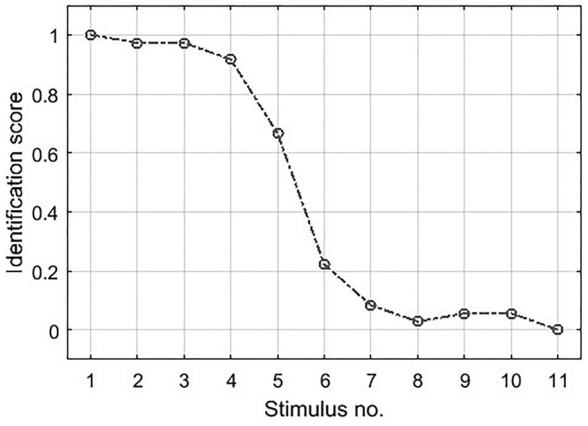

FIG. 10. Deshkar identification scores by TMs versus stimulus index.

using the Sonic Mapper interface discussed earlier. Each identification rating takes no more than 8 s (assuming single play). This accumulates to $22 \times 8 = 176$ s $\sim$ 3 min. If the number of plays is higher (up to three or four for confusing stimuli), the total time taken for a trial block is no more than 8 min. The two trial blocks are separated by 1h or more.

2. Discrimination

During discrimination, pairs of stimuli consisting of either identical (AA) or different (AB) pairs were presented. The different pairs comprised stimuli separated by one index in Table IV to get ten distinct pairs. The 12 control participants, restricted to the TMs' group, were given both AB and BA pairs in the case of the characteristic phrase context only. For the non-control participants, a balanced mix of AB and BA pairs were presented. All pairs were repeated two times in each trial block. To exclude any bias toward the same/different choice, a comparable number (40) of AA pairs was added. In total, 80 stimulus pairs presented in randomized order were to be evaluated. Each pair takes 12 s + ISI 0.5 s = 12.5 s. If each pair is played once, the time taken for each pair is 15 s (from start of play to the start of next pair). Thus, the whole set for control participants takes $80 \times 15 = 1200$ s or 20 min. If the number of plays is higher (usually up to three times for confusing pairs), the total time taken for a trial block is less than 25 min. The two trial blocks are separated by at least 1h.

B. Results and discussion: Identification

Figure 10 shows the scores in terms of the fraction of stimuli identified as Deshkar, averaged across the musician participants. We observe an S-shaped curve with a steep crossover between Deshkar and Bhupali choices around the stimulus range 4–7. Individual differences in the category crossover were observed to be confined to the same narrow region occurring between four and seven, indicating that similar criteria were used across participants. The exact location of the category crossover is computed as the interpolated point in the continuum at which the identification function passes (50%) or chance level (Schneider, 2012; Stanislaw and Todorov, 1999). Averaging across participants, the crossover between the categories was found to be between stimulus numbers 5 and 6 at 5.13, which corresponds to an R duration of ~0.9 s. Thus, a GRS phrase presented with a R duration of 0.9 s is equally likely to be recognized as raga Deshkar or Bhupali.

C. Results and discussion: CP discrimination

1. Characteristic phrase context

The effect of presentation order in the stimulus pairs was studied via the responses of the TM control participants. It was found that there was no significant effect of order at threshold of 0.01. For the ten stimuli pairs, we had $p = (0.015, 0.055, 0.071, 0.051, 0.1, 0.038, 0.1, 0.003, 0.064, 0.032)$. We see that only in the pair 8–9, the BA is significantly better discriminated ($p = 0.003$). There was no apparent explanation for this isolated deviation from the absence of the order effect. We therefore averaged the scores across all trials of the 23 TM participants to obtain the percentage discrimination in Fig. 11(a). The maximum discrimination accuracy for all but three participants was 100%. As for the NM group, one participant marked all pairs as same in the discrimination task. The ratings of this participant were excluded from the analyses, reducing the number of NM participants to 14. Figure 11(a) also shows the average discrimination function of these 14 participants, in contrast to the 23 musician participants.

We map the percentage discrimination to perceptual distance as presented in Sec. V C 2 to obtain Fig. 11(b). We see a clear peak in TMs' averaged d' occurring at the stimulus pair 5–6 falling off very rapidly on either side. The peak value itself at around $d' = 2.2$ is indicative of significant discrimination. The NM group, on the other hand, exhibits a narrow range of relatively low d' values across the set of stimulus pairs, indicating their similar perception of the acoustic difference irrespective of its location with respect to the category boundary.

To test for CP, the correlation of the individual crossover points and corresponding discrimination peaks was examined (Schneider, 2012; Pisoni and Lazarus, 1974; MacKain *et al.*, 1981). For each of the 23 participants the crossover was separately computed. A linear regression

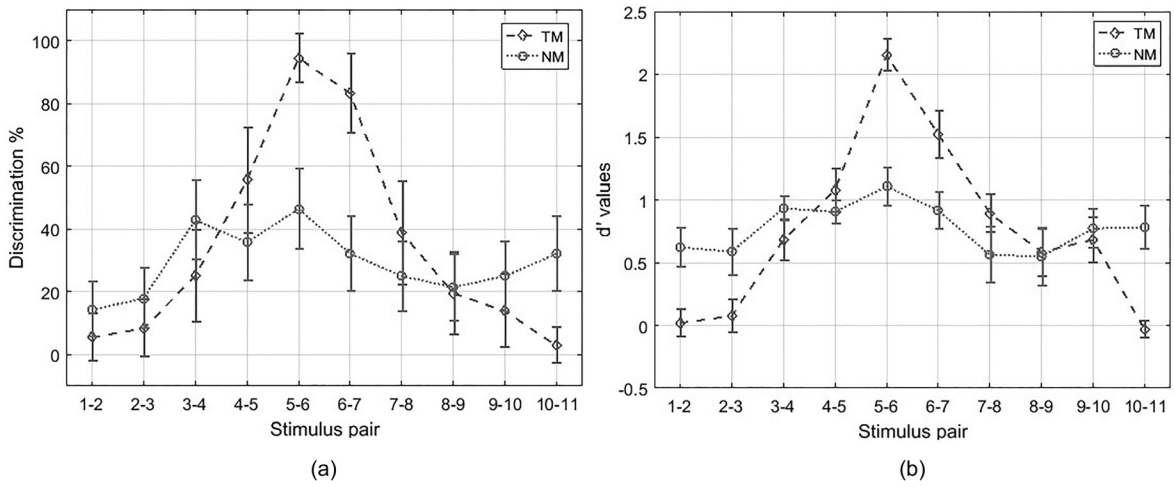

FIG. 11. Mean and standard deviation of TMs and NMs (a) discrimination scores and (b) d' values, averaged over participants and trials for the CP discrimination task for the characteristic DPGRS phrase.

analysis (see Fig. 12) revealed a correlation of 0.74 ($p = 7 \times 10^{-5}$).

2. Non-characteristic phrase context

The non-characteristic phrase context of R-duration manipulation was presented to a subset of 16 TMs from the group who did the characteristic phrase context. Figure 13 compares the discrimination performance for the two cases, characteristic DPGRS and non-characteristic DPMGRS phrases. We note a distinct difference in behavior in the out-of-context phrase with the TMs' perception of acoustic differences being almost uniform across the stimulus pairs, ruling out the possible presence of categories.

VII. EXPERIMENT 3: LISTEN AND IMITATE

Experiment 3 involves testing for the existence of categories via production. The participant's task is to imitate a presented prompt as closely as possible. The goal is to test whether participants memorize the stimulus and tend to recall it from their working memory, or whether they perceive it as the closest P template already stored in their long-term memory and thus reproduce the P. In the present context of two categories, a TM participant might be expected then to produce instances of one of two distinct patterns when confronted with any stimulus prompt drawn from the continuum between two category Ps. This method has been successfully applied in studies on speech intonational categories (Pierrehumbert and Steele, 1989; Redi, 2003).

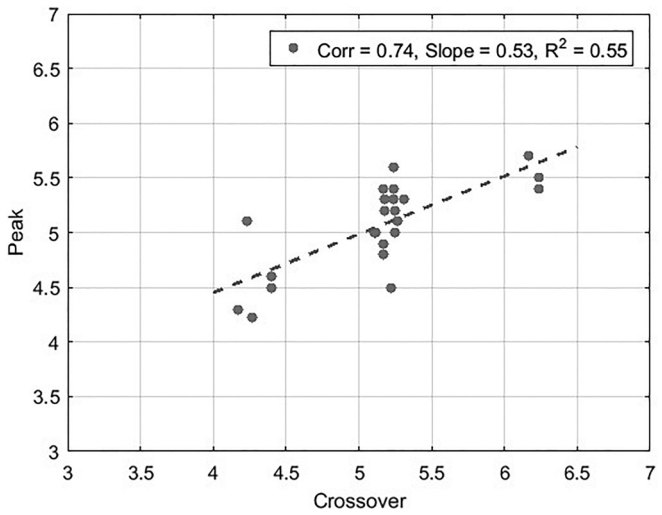

FIG. 12. Correlation between individual crossover and discrimination peaks for 23 TM participants.

A. Method

The experiment is conducted for both the characteristic DPGRS and the non-characteristic DPMGRS phrases. To recapitulate, the corresponding GRS segments for both phrase categories have the same melodic shape, differing only in the pre-context of the GRS segment with coinciding G- and S-onsets, as seen in the example of Fig. 9.

As the name listen and imitate suggests, the experiment involves recording of participants' vocal rendition of the stimuli. Participants were instructed to hum the stimulus prompt as closely as possible with no other information about the stimulus provided. A total of 22 stimuli (11 each from the 2 phrase categories, with R-duration values chosen as in Table IV) were presented each followed by a recording of the participant's corresponding hummed response. The stimuli within each of the two trial blocks were repeated twice and presented in a randomized order. After each stimulus was played, a sufficiently long pause was provided with only the tanpura drone as a backing track. While the prompt was accompanied by metronome clicks on the onsets of G and S notes, this was omitted in the singing pause to avoid constraining the participant's imitation too much. Each stimulus, followed by the hummed response, takes about 15 s in a single play. This accumulates to $44 \times 15 = 660$ s ~ 11 min. Even if the number of plays requested by the participant is higher (which occurred in rare instances, particularly with the more complex non-phrase context), the total time taken for a trial block was no more than 20 min. The two trial blocks are separated by at least 1h.

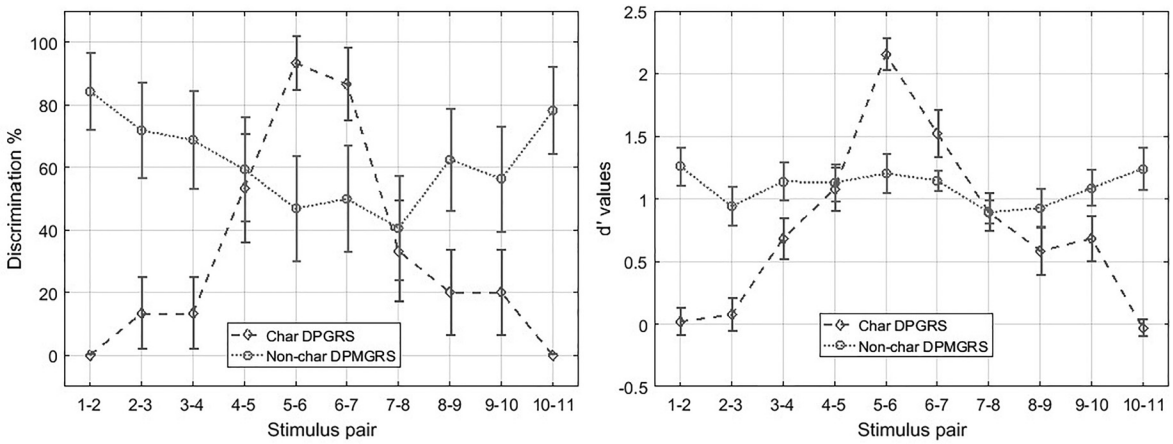

FIG. 13. Mean and standard deviation of TMs' (a) discrimination score and (b) d' values averaged over participants and trials for the CP discrimination task for the non-characteristic DPMGRS phrase.

The recordings were carried out in a quiet environment on a high fidelity digital recorder (Edirol R-09H, Roland) with an audio encoding of 44.1 kHz sampled 16-bit mono PCM (pulse-code modualtion, .wav) format. Participants listened to the stimulus through an over-ear headphone (HD 180, Sennheiser) with a moderate volume level. In the singing pause, the background tanpura played only in the headphone, ensuring a clean audio for the sung melody extraction and further analyses. One interesting finding was that all participants attempted to maintain the tempo of the prompt by tapping even though not instructed to.

We segmented the constituent notes from the pitch contours of the recorded phrases by the method presented in Sec. III B and measured the duration of each of the notes.

B. Statistical model fitting

We observe the dependence of the imitated phrase, specifically the duration of the R note, on the corresponding stimulus prompt. For an accurate imitation, we would expect to see a linearly correlation. The presence of learned categories would be expected to influence this relationship, however, making it more of a sigmoid shape.

We also analyze the individual participants' response data using logistic regression to obtain the corresponding sigmoidal curve fits. The sung duration averaged across trials for a given participant can be viewed as representing the proportion of one of the categories elicited by a specific prompt. The bias intercept and slope coefficients of a sigmoid curve fit can be revealing about the extent to which the specific participant uses the duration cue in categorization (Morrison and Kondaurova, 2009). One of the inputs to the logistic regression model is the sung (raw) duration of the segmented R note. First, the participant's array of 44 R-duration values for the stimulus continuum (indices 1–11, repeated 2 times for each phrase category in 2 trial blocks) is min-max normalized between [0,1] The second input to the model is a categorical array of the stimulus continuum. We provide a categorization of 5 Deshkar + 6 Bhupali candidates as obtained from the identification results of Fig. 10. The model is fitted in a log odds space. The log odds transformation converts proportions in the range 0–1 into logits in the range $-\infty$ to $+\infty$. Logit values from the fitted model can be converted to probabilities so that fitted curves in the log odds space become sigmoidal curves in the probability space. Thus a logistic regression model was fitted to each participant's curve of sung R duration versus prompt R duration for each of the characteristic (DPGRS) and non-characteristic (DPMGRS) phrase contexts. The model included a bias coefficient and a duration-tuned coefficient, as given by Eq. (2),

$$p(\text{Bhupali}|\text{R}) = \frac{1}{e^{-(\alpha+\beta \text{R}_{\text{duration}})}+1}, \tag{2}$$

where the left-hand side is the proportion of Bhupali responses for each R-duration value, α is the intercept, i.e., displacement along the x axis, and β is the regression coefficient (Morrison and Kondaurova, 2009), i.e., the sigmoid growth rate.

C. Results and discussion

The sung phrase shape is indicative of the perceived phrase shape. Thus, the variation in sung duration with reference to the corresponding prompt duration can capture perceptual categorization, if any. We separately process the data corresponding to each participant category and each phrase context. In order to detect any systematic differences in behaviors between participant groups and phrase contexts, we compare the corresponding logistic fit parameter distributions for the different participant categories and phrase contexts.

1. Characteristic DPGRS

Figure 14(a) shows the mean and standard deviation (across participants and trials) of the observed R duration for each reference prompt, labeled by its R duration, in characteristic phrase context. We see that the TMs are relatively insensitive to the changing prompt, especially in the region of low R duration where they seem to be producing the

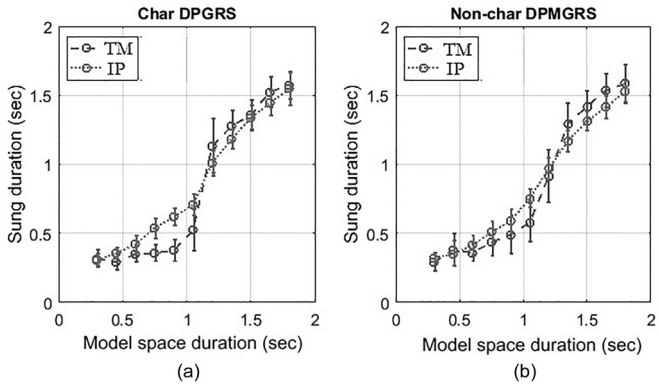

FIG. 14. Sung duration mean and standard deviation across participants (TM; IP) and trials versus prompt duration of R note in the context of the (a) characteristic phrase and (b) non-characteristic phrase.

P phrase of raga Deshkar irrespective of the presented prompt. As the presented R duration increases, there is a point beyond which a rapid jump in the sung R duration occurs for the subsequent increase in prompt R duration. Post this, there is a slow increase in sung R duration with prompt R duration. The distribution of the sung duration of R note (not shown in the plot) is bimodal with a sharp peak at 0.4 s and a more extended shallow peak around 1.3 s. This indicates that TMs tended to reproduce phrase Ps from long-term memory in the imitation task, with some sensitivity to prompt R duration in the Bhupali category. Indi-pop musicians, on the other hand, followed the melodic shape almost exactly as evident from the near diagonal nature of the fitted curve. That is, the Indi-pop musicians show a more proportionate increase in sung duration with prompt duration throughout the continuum. All participants were observed to trade off the G- and R-note durations in their rendition to keep the overall duration between the G- and S-note onsets constant.

2. Non-characteristic DPMGRS

The case for the non-characteristic DPMGRS is interesting because there is no expected category for this common descending phrase that appears in many ragas. In Fig. 14(b), we note that for the Indi-pop participants' group, the curve is not different from that corresponding to the characteristic phrase for the same group, which is expected. In the case of the TMs, we see a less distinct S-shape, indicating a better correlation between the prompted and sung R durations in the non-characteristic context.

An explanation for the more sigmoidal shape for the TM compared to the Indi-pop singer (IP) participants in Fig. 14(b) became evident from the informal discussions with some of the TM participants after the experiment. They mentioned that the phrase evoked the memory of a certain raga performance, although they differed in the specifics and named different ragas (Yaman, Sudh Kalyan, and Vachaspati, among others). Thus it appears that TMs have a tendency to ascribe an identity or form from their long-term memory when they hear something that sounds like a musical phrase.

3. Comparing statistical model parameter distributions

Figure 15 shows the distributions of the logistic regression fit parameters (β and α) computed for each participant in a given participant group and phrase context. We make the following observations.

- The distributions of β, the sigmoid growth-rate parameter, for TMs are separated from the corresponding distributions for Indi-pop musicians with a significant difference of the means ($p = 2 \times 10^{-16}$). The higher β in the case of the TMs indicates a steeper crossover between categories. This is consistent with the shapes of the curves in Fig. 14 where we noted that the relationship between the sung and prompted R durations for Indi-pop musicians is close to linear rather than sigmoidal in both phrase contexts. This implies that CP can be ascribed to the TM but not to the IP participants.
- The means of the distributions of α between TM and IP participants differ significantly, but this stems from the interaction of the model parameters. A change in the

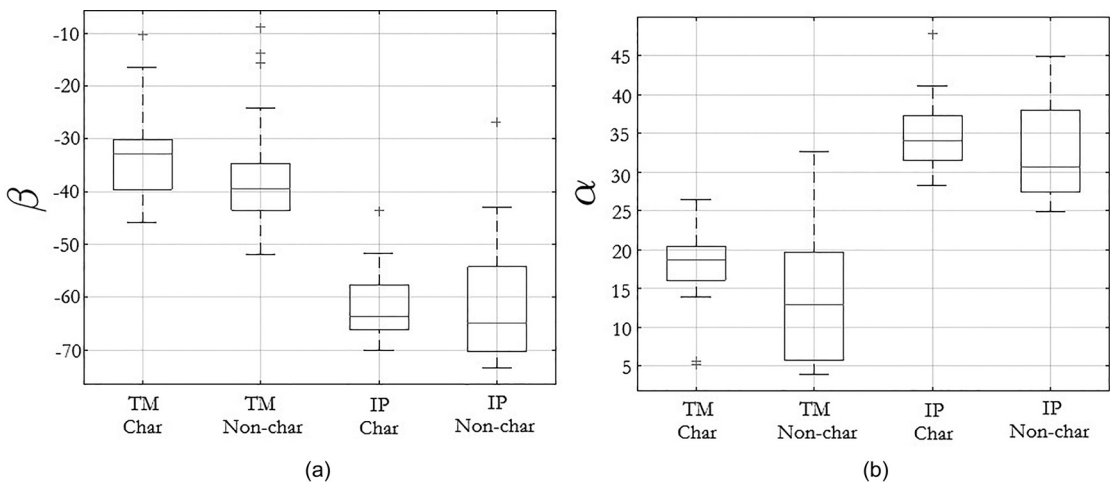

FIG. 15. Box-plots of individual participant's (a) β and (b) α values for characteristic DPGRS and non-characteristic DPMGRS phrases for TM and IP participant groups.

growth rate without a change in the crossover point of the sigmoid leads to a corresponding change in the intercept α.

- The TMs show a higher mean in the β distribution for the characteristic phrase compared to that for the non-characteristic phrase, but the difference is not significant ($p = 0.18$). This indicates that the curves of an individual TM participant have similar steepness across characteristic and non-characteristic contexts. This counterintuitive result is explained by the previous observation that while the DPGRS phrase does not have a characteristic shape in any of the ragas it can occur in, TM participants tend to interpret the presented prompt semantically and reproduce their own specific interpretation. We thus obtain a narrow distribution of the growth-rate coefficient β across the TM participants for the non-characteristic context. However, due to the differing individual-dependent interpretations, the crossover locations of the sigmoid curve fits tend to vary across a range leading to the greater dispersion in α.
- IP participants show greater dispersion in the model fit parameters for the non-characteristic phrase relative to that of the characteristic phrase. This may be attributed to the higher complexity of the latter from the greater number of notes, making it more challenging to accurately reproduce for these untrained singers.

VIII. SUMMARY AND GENERAL DISCUSSION

With the broader goal of computational modeling of melodic similarity, we considered the similarity and categorization of melodic motifs. Across genres as diverse as Western folk and Indian art music, melodic motifs play an important role in global similarity judgements by humans, and musical pieces are classified by the frequency of appearance of the recognized motifs (Cambouropoulos *et al.*, 2001; Volk and Kranenburg, 2012; Boot *et al.*, 2016). We therefore considered, in this work, melodic similarity at the phrase level in the context of human judgements, which could be influenced by implicit genre-specific musicological knowledge or learned schema involving phrase categories. Although we considered learned schema and long-term memory effects in this work, it is likely to be more generally applicable to music listening where the repeated use of a motif in a piece leads to an "imprint" that starts to develop from the first hearing of the piece and becomes associated with an abstracted cue (Deliège, 2001).

The perceptual discrimination of acoustic variants of Ps in music has received relatively limited attention compared to that accorded to Ps in speech. As reviewed earlier, chords and melodic musical intervals have been investigated for CP by musicians, NMs, and amateurs (Burns and Ward, 1978; Siegel and Siegel, 1977; Acker *et al.*, 1995; Barrett, 1999). There is no similar previous work, however, with continuous melodic shapes where the variations can be temporal in nature. On the other hand is the work on CP of rhythms based on integer-ratio intervals (Jacoby and McDermott, 2017; Desain and Honing, 2003). Characteristic phrases in raga music, represented by continuous pitch curves, serve well to study the effects of learned categories on perception. In this work, we chose the dimension of note duration, an acoustic cue linked to musical emphasis, to investigate the perceptual discrimination of variations in the vicinity of a prototypical phrase of a well-known raga. Both, the prototypical melodic shape and the distinguishing cue were abstracted from measurements of the manually annotated phrase in a corpus-based study. The perception experiments then used a stimulus set comprising variations of the phrase along the chosen dimension but realized with a uniform loudness and timbre across stimuli.

Our experiments with TMs confirmed the existence of a melodic shape P in a raga-belongingness-based goodness rating experiment. All participants showed improved discrimination scores in stimuli pairs with increasing physical separation between the stimuli. However, in the case of the TMs, the P acted like a perceptual magnet where physical distances in its neighborhood were not discriminated as well as the similar separation in a region away from the P (but still in the space of the raga phrase). NMs did not show any difference in their perception between the vicinity of the P and away from it. It was found that to the TM listeners, stimuli at the higher extreme of R-note duration range evoked an impression of the allied raga Bhupali phrase with the same sequence of notes. Based on the assumption of categories, CP experiments for identification and discrimination were carried out. Identification scores across TM participants indicated unambiguous categorization to a raga label at the extreme ends of the unidimensional R-duration range accompanied by a steep crossover of category label roughly midway. Discrimination of stimuli separated by a single step in R duration showed a peak at the category crossover and very low scores near the extremes, confirming the presence of CP. NM participants showed relatively constant discrimination scores across the R-duration range. A production-based experiment (listen and imitate), borrowed from speech intonation studies, also captured the differences in phrase perception by TMs and untrained singers very effectively.

All in all, our perception experiments confirm that the learned schema of raga phrases warps the perceptual distance with respect to physical distance along the cue dimension. The distinct behavior of TMs in both listening and production experiments points to the role of long-term memory in raga phrase perception. The perceptual attractor property of the P phrase can be explained by the observed spread in the acoustic cue dimension in the corpus study. It is thus possible that several exemplars of the phrase are stored in the musician's long-term memory. The strong context dependence of the categorization cue is revealed in experiments with the same melodic motif variation in a non-characteristic phrase setting. While NMs continue to show no evidence of a discrimination peak, as expected, musicians trained in the genre show evidence of a peak even in the non-characteristic context, but diverge among themselves in the location of the peak in the cue continuum. This latter phenomenon indicates that the TMs are predisposed to classifying the musical phrases they hear into categories. In melodic contexts where the relation between the acoustic cue dimension and categories is ambiguous, we see individual differences in the detected category boundary location.

As for the practical implications of the presented work, retrieval of recurring motifs has been an important focus in melody-based MIR research. This has been typically attempted using time series pattern matching methods with distance computation between continuous pitch segments (Typke *et al.*, 2007; Rao *et al.*, 2014; Gulati *et al.*, 2016; Dutta *et al.*, 2015). Dynamic time warping has been applied, possibly with constraints that serve to weight certain melodic events more than others. It has been recognized, however, that distance computation based on high level musical descriptors matches human similarity judgements better compared to that obtained by simple pitch contour distance (Kroher *et al.*, 2014). Given the importance of the correct recognition of recurrent motifs for global similarity characterization (e.g., folk tune classification or raga identification), the present work confirms the importance of salient cues in the categorical judgements. High level features can be derived from the cues, which themselves can be established for a given context via musicological knowledge or corpus studies, as illustrated in this work. Further, given the distinctly different discrimination behaviors observed for TMs and untrained listeners, measurement of CP can potentially be used to assess musical ability. A similar proposition was made by Vuust *et al.* (2011) based on auditory event-related potentials (ERP) measures, which are known to be strongly correlated with behavioral discrimination measures.

Our experimental material was drawn from the pentatonic ragas Bhupali and Deshkar with their overlapping characteristic phrases in terms of note sequences. In general, it is not uncommon to find that the same named phrase corresponds to more than one raga. What is unique to a raga is the distinctive melodic shape in terms of one or more of the relative note durations, specific non-standard note intonations, and the nature of transition segments connecting the notes (van der Meer, 1980; Rao and Rao, 2014). The selected GRS motif in this work differs across the two ragas in the temporal extent of R and also in the intonation of G, which is pitched slightly higher in Deshkar (Kulkarni, 2011), thus providing multiple cues to the phrase category in practice. In the reported experiments of this paper, the G intonation corresponded to that of raga Deshkar and was not varied. However this does not seem to have affected the CP of the phrase along the R-note duration dimension implying that the intonation cue may be lower in the cue hierarchy. Future work should consider both cues and their interplay in phrase categorization. Finally, we note that we considered synthetic stimuli that captured only the essential melodic shape of the motif and neglected all other dimensions, including non-essential pitch embellishments, timbre, and loudness dynamics. Complex, natural stimuli are known to elicit more categorical effects (Van Hessen and Schouten, 1999; Kroher *et al.*, 2014), i.e., perception becomes more categorical as naturalness increases. Stimuli in which relevant cues are enhanced/caricaturized might be more psychoacoustic than categorically perceived. Future work could also consider the perceptual influence of embellishments and ornamentation such as brief excursions of pitch within a note that performers often use for aesthetic effect.

ACKNOWLEDGMENTS

This work received partial funding from the European Research Council under the European Union's Seventh Framework Programme (Grant No. FP7/2007-2013)/ERC Grant agreement 267583 (CompMusic).

[1]https://dunya.compmusic.upf.edu/hindustani (Last viewed April 6, 2019).

Aaltonen, O., Niemi, P., Nyrke, T., and Tuhkanen, M. (**1987**). "Event-related brain potentials and the perception of a phonetic continuum," Biol. Psychol. **24**, 197–207.

Acker, B. E., Pastore, R. E., and Hall, M. D. (**1995**). "Within-category discrimination of musical chords: Perceptual magnet or anchor?," Percept. Psychophys. **57**, 863–874.

Allan, H., Mullensiefen, D., and Wiggins, G. A. (**2007**). "Methodological considerations in studies of musical similarity," in *Proc. of Int. Soc. for Music Information Retrieval (ISMIR)*, pp. 473–478.

Asano, R., and Boeckx, C. (**2015**). "Syntax in language and music: What is the right level of comparison?," Front. Psychol. **6**, 942–957.

Autrim-NCPA (**2017**). "Music in motion: The automated transcription for Indian music (AUTRIM) project by NCPA and UvA," available at https://autrimncpa.wordpress.com/ (Last viewed April 26, 2017).

Bagchee, S. (**2006**). *Shruti: A Listener's Guide to Hindustani Music* (Rupa Co., New Delhi, India).

Barrett, S. (**1999**). "The perceptual magnet effect is not specific to speech prototypes: New evidence from music categories," Speech Hear. Lang.: Work Prog. **11**, 1–16.

Batliner, A., and Schiefer, L. (**1987**). "Stimulus category, reaction time, and order effect-an experiment on pitch discrimination," Proc. ICPhS **5**(3), 46–49.

Boot, P., Volk, A., and de Haas, W. B. (**2016**). "Evaluating the role of repeated patterns in folk song classification and compression," J. New Music Res. **45**, 223–238.

Burns, E. M., and Ward, W. D. (**1978**). "Categorical perception—Phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals," J. Acoust. Soc. Am. **63**, 456–468.

Cambouropoulos, E., Crawford, T., and Iliopoulos, C. S. (**2001**). "Pattern processing in melodic sequences: Challenges, caveats and prospects," Comput. Humanit. **35**, 9–21.

Cowdery, J. R. (**1984**). "A fresh look at the concept of tune family," Ethnomusicology **28**, 495–504.

Cutting, J. E., and Rosner, B. S. (**1974**). "Categories and boundaries in speech and music," Percept. Psychophys. **16**, 564–570.

Deliège, I. (**2001**). "Prototype effects in music listening: An empirical approach to the notion of imprint," Music Percept.: Interdiscip. J. **18**, 371–407.

Desain, P., and Honing, H. (**2003**). "The formation of rhythmic categories and metric priming," Perception **32**, 341–365.

Downie, J. S. (**2003**). "Music information retrieval," Annu. Rev. Inform. Sci. Technol. **37**(1), 295–340.

Dutta, S., Krishnaraj, S. P., and Murthy, H. A. (**2015**). "Raga verification in Carnatic music using longest common segment set," in *Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pp. 605–611.

Egan, J. P. (**1975**). *Signal Detection Theory and ROC Analysis Academic Press Series in Cognition and Perception* (Academic, London, UK).

Fiske, H. E. (**1997**). "Categorical perception of musical patterns: How different is 'different?,'" in *Bulletin of the Council for Research in Music Education*, pp. 20–24.

Ganguli, K. K., Gulati, S., Serra, X., and Rao, P. (**2016**). "Data-driven exploration of melodic structures in Hindustani music," in *Proc. of the International Society for Music Information Retrieval (ISMIR)*, New York, pp. 605–611.

Ganguli, K. K., Lele, A., Pinjani, S., Rao, P., Srinivasamurthy, A., and Gulati, S. (**2017**). "Melodic shape stylization for robust and efficient motif detection in Hindustani vocal music," in *Proc. of National Conference on Communications (NCC)*, IEEE, pp. 1–6.

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

Ganguli, K. K., and Rao, P. (**2017**). "Towards computational modeling of the ungrammatical in a raga performance," in *Proc. of the International Society for Music Information Retrieval (ISMIR)*, Suzhou, China.

Ganguli, K. K., and Rao, P. (**2018**). "On the distributional representation of ragas: Experiments with allied raga pairs," Trans. Int. Soc. Music Inform. Retr. **1**, 79–95.
Goldstone, R. L., and Hendrickson, A. T. (**2010**). "Categorical perception," Cognit. Sc.: Wiley Interdiscip. Rev. **1**, 69–78.
Gulati, S., Bellur, A., Salamon, J., Ranjani, H. G., Ishwar, V., Murthy, H. A., and Serra, X. (**2014**). "Automatic tonic identification in Indian art music: Approaches and evaluation," J. New Music Res. **43**, 53–71.
Gulati, S., Serrà, J., Ishwar, V., Şentürk, S., and Serra, X. (**2016**). "Phrase-based rāga recognition using vector space modeling," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 66–70.
Harnad, S. (**2003**). "Categorical perception," in *Encyclopedia of Cognitive Science* (Nature Publishing Group/Macmillan, London).
Hirst, D., and Di Cristo, A. (**1998**). *Intonation Systems: A Survey of Twenty Languages* (Cambridge University Press, Cambridge, UK), pp. 1–44.
Iverson, P., and Kuhl, P. K. (**1995**). "Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling," J. Acoust. Soc. Am. **97**, 553–562.
Jacoby, N., and McDermott, J. H. (**2017**). "Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction," Curr. Biol. **27**, 359–370.
Kroher, N., Gómez, E., Guastavino, C., Gómez, F., and Bonada, J. (**2014**). "Computational models for perceived melodic similarity in a cappella flamenco singing," in *Proc. of Int. Soc. for Music Information Retrieval (ISMIR)*, pp. 65–70.
Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (**1992**). "Linguistic experience alters phonetic perception in infants by 6 months of age," Science **255**, 606–608.
Kulkarni, S. (**2011**). *Shyamrao Gharana* (Prism Books Pvt. Ltd., Bangalore, India), Vol. 1.
Ladd, D. R., and Morton, R. (**1997**). "The perception of intonational emphasis: Continuous or categorical?," J. Phonetics **25**, 313–342.
Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). "The discrimination of speech sounds within and across phoneme boundaries," J. Exp. Psychol. **54**, 358–368.
MacKain, K. S., Best, C. T., and Strange, W. (**1981**). "Categorical perception of English /r/ and /l/ by Japanese bilinguals," Appl. Psycholinguist. **2**, 369–390.
Marsden, A. (**2012**). "Interrogating melodic similarity: A definitive phenomenon or the product of interpretation?," J. New Music Res. **41**, 323–335.
McMurray, B., Dennhardt, J. L., and Struck-Marcell, A. (**2008**). "Context effects on musical chord categorization: Different forms of top-down feedback in speech and music?," Cognit. Sci. **32**, 893–920.
Morrison, G. S., and Kondaurova, M. V. (**2009**). "Analysis of categorical response data: Use logistic regression rather than endpoint-difference scores or discriminant analysis," J. Acoust. Soc. Am. **126**, 2159–2162.
Mullensiefen, D., and Frieler, K. (**2004**). "Measuring melodic similariry: Human vs. algorithmic judgments," in *Proc. of Interdisciplinary Musicology*.
Novello, A., McKinney, M. F., and Kohlrausch, A. (**2006**). "Perceptual evaluation of music similarity," in *Proc. of Int. Soc. for Music Information Retrieval (ISMIR)*.
Pearce, M., Mullensiefen, D., and Wiggins, G. (**2010**). "Melodic grouping in music information retrieval: New methods and applications," in *Advances in Music Information Retrieval*, Studies in Computational Intelligence (Springer, Berlin Heidelberg), Vol. 274, pp. 364–388.
Pierrehumbert, J. B., and Steele, S. A. (**1989**). "Categories of tonal alignment in English," Phonetica **46**, 181–196.
Pisoni, D. B., and Lazarus, J. H. (**1974**). "Categorical and noncategorical modes of speech perception along the voicing continuum," J. Acoust. Soc. Am. **55**, 328–333.
Powers, H. S., and Widdess, R. (**2001**). *India, Subcontinent of*, 2nd ed. (Macmillan, London), Chap. III.
Raja, D. (**2016**). *The Raga-ness of Ragas: Ragas Beyond the Grammar* (D. K. Print World Ltd., New Delhi, India).
Rao, V., and Rao, P. (**2010**). "Vocal melody extraction in the presence of pitched accompaniment in polyphonic music," IEEE Trans. Audio Speech Lang. Process. **18**(8), 2145–2154.
Rao, S., and Rao, P. (**2014**). "An overview of Hindustani music in the context of computational musicology," J. New Music Res. **43**(1), 24–33.
Rao, P., Ross, J. C., Ganguli, K. K., Pandit, V., Ishwar, V., Bellur, A., and Murthy, H. A. (**2014**). "Classification of melodic motifs in raga music with time-series matching," J. New Music Res. **43**, 115–131.
Redi, L. (**2003**). "Categorical effects in production of pitch contours in English," in *Proceedings of the 15th International Congress of the Phonetic Sciences*, pp. 2921–2924.
Repp, B. H. (**1984**). "Categorical perception: Issues, methods, findings," Speech Lang. **10**, 243–335.
Rodd, J., and Chen, A. (**2016**). "Pitch accents show a perceptual magnet effect: Evidence of internal structure in intonation categories," in *Speech Prosody 2016*, pp. 697–701.
Rohrmeier, M., and Widdess, R. (**2012**). "Incidental learning of modal features of north Indian music," in *12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*, Thessaloniki, Greece, pp. 23–28.
Scavone, G. P., Lakatos, S., and Harbke, C. R. (**2002**). "The sonic mapper: An interactive program for obtaining similarity ratings with auditory stimuli," in *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, Japan, July 2–5, 2002.
Schneider, K. (**2012**). "The German boundary tones: Categorical perception, perceptual magnets, and the perceptual reference space," Dr. Phil. thesis, Institut fur Maschinelle Sprachverarbeitung der Universitat Stuttgart, Germany.
Schneider, K., Dogil, G., and Möbius, B. (**2009**). "German boundary tones show categorical perception and a perceptual magnet effect when presented in different contexts," in *Tenth Annual Conference of the International Speech Communication Association*.
Siegel, J. A., and Siegel, W. (**1977**). "Categorical perception of tonal intervals: Musicians can't tell sharp from flat," Percept. Psychophys. **21**, 399–407.
Stanislaw, H., and Todorov, N. (**1999**). "Calculation of signal detection theory measures," Behav. Res. Methods, Instruments Comput. **31**, 137–149.
Typke, R., Wiering, F., and Veltkamp, R. C. (**2007**). "Transportation distances and human perception of melodic similarity," Musicae Scientiae **11**, 153–181.
van der Meer, W. (**1980**). *Hindustani Music in the 20th Century* (Martinus Nijhoff Publishers, Dordrecht, Netherlands).
van der Meer, W. (**2008**). "Improvisation versus reproduction, India and the world," New Sound: Int. Mag. Music **32**.
Van Hessen, A. J., and Schouten, M. (**1999**). "Categorical perception as a function of stimulus quality," Phonetica **56**, 56–72.
Vempala, N. N., and Russo, F. A. (**2012**). "A melodic similarity measure based on human similarity judgments," in *Proc. of the Int. Conf. on Music Perception and Cognition*.
Vempala, N. N., and Russo, F. A. (**2015**). "An empirically derived measure of melodic similarity," J. New Music Res. **44**, 391–404.
Vijaykrishnan, K. G. (**2007**). *The Grammar of Carnatic Music* (De Gruyter Mouton, Berlin).
Volk, A., and Kranenburg, P. V. (**2012**). "Melodic similarity among folk songs: An annotation study on similarity-based categorization in music," Musicae Scientiae **16**, 317–339.
Vuust, P., Brattico, E., Glerean, E., Seppänen, M., Pakarinen, S., Tervaniemi, M., and Näätänen, R. (**2011**). "New fast mismatch negativity paradigm for determining the neural prerequisites for musical ability," Cortex **47**, 1091–1098.
Welch, B. L. (**1947**). "The generalisation of student's problems when several different population variances are involved," Biometrika **34**, 28–35.
Wickens, T. D. (**2002**). *Elementary Signal Detection Theory* (Oxford University Press, New York).
Widdess, R. (**2013**). "Schemas and improvisation in Indian music," in *Language, Music and Interaction*, edited by R. Kempson, C. Howes, and M. Orwin (College Publications, London).