Proceedings of SPCOM 2020, I.I.Sc. Bangalore, July 2020

Classifying Cultural Music using Melodic Features

Amruta Vidwans
Department of Electrical Engg
IIT Bombay
amruta28757@iitbombay.org

Prateek Verma
Department of Electrical Engg
IIT Bombay
prateekv@stanford.edu

Preeti Rao
Department of Electrical Engg
IIT Bombay
prao@ee.iitb.ac.in

***Abstract*—We present melody based classification of musical styles by exploiting pitch and energy based characteristics computed on the audio signal. Three prominent musical styles were chosen which have improvisation as an integral part with similar melodic principles, theme, and structure of concerts namely, Hindustani, Carnatic and Turkish music. Listeners of one or more of these genres can discriminate these entirely based on the melodic style. The resynthesized melody of music pieces that share the underlying raga/makam, removing any singer cues, was used to validate our hypothesis that style distinction is embedded in the melody. Our automatic method is based on finding a set of highly discriminatory features, motivated by musicological knowledge, to capture distinct characteristics of the melodic contour. The nature of transitions in the pitch contour, presence of microtonal notes and the dynamic variations in the vocal energy are exploited. The automatically classified style labels are found to correlate well with the judgments of human listeners. The melody based features when combined with timbre based features, were found to improve the classification performance on the music metadata based genre labels.**

I. Introduction

Indian classical music is categorized into two major subgenres, viz. Hindustani and Carnatic with corresponding origins in North and South India respectively . They have similar *raga* (melodic mode) and *tala*(rhythm) framework while differing in style and instrumentation. Similar to Indian classical music, in terms of the *raga* framework, is Turkish music which has the *makam* i.e. a scale and melodic movements associated with it. The structure of the concert is similar across three music styles with the concert starting with an unmetered and improvised elaboration called the *alap* (*taqsim* in Turkish music) section where the main features of the chosen raga / makam are brought out by the soloist. There has been past work on Indian classical music on motif identification [1], *raga* recognition [2][3] and sub-genre classification [4]. Classification for different styles viz. Arabic, Chinese, Japanese, Indian, African, Western Classical music using timbre, rhythm and wavelet coefficients of decomposing the audio signal via multi-level Daubechies wavelet was proposed in [5]. They emphasized that the diversity existing within Indian classical music is difficult to model. Audio pieces were divided into 9 world regions using rhythmic, tonal and timbral features with new features for non-western music in [6]. Melody based classification for western music styles was done by obtaining high-level melodic descriptors that could be easily related to the properties of the music [7]. Panteli et al. [8] observed that clusters that appear in a melodic features space obtained from across several genres of music can be distinguished by characteristic uses of vibrato, melisma, and slow versus fast syllabic singing. Acoustic features based on timbre such as MFCC, delta-MFCC, spectral features etc. borrowed from speech processing were used in [9] to distinguish different Indian music genres rather than using any specific musical attributes.

978-1-7281-8895-9/20/$31.00 ©2020 IEEE

Our goal in this work is to address the problem of the automatic classification of these culture specific music styles in the context of the melody in vocal concerts. Even though there exist obvious timbre features based on the accompanying instruments used (such as violin in the Carnatic style and harmonium in Hindustani concerts), and language cues, we wish to explore the discrimination of the three styles using pitch attributes alone. This is based on the observation that that listeners can typically discriminate the three styles based on the alap with its solo singing, often using only vowel utterances. The previous work on style classification [10] has been improved and extended to a larger dataset in addition to including Turkish music. An investigation of acoustic features for automatic classification can help us obtain insights about the significant differences between culture specific music styles. Apart from the musicological importance of the outcomes, the discovered attributes can be useful in music information retrieval (MIR) tasks such as automatic music tagging and recommendation systems.

II. Database

Concert recording CDs of widely performed *ragas/makams* by well-known artists (of various schools of music) of Hindustani, Carnatic and Turkish style were obtained. A *raga/makam* can be defined as a scale of notes having a particular arrangement and melodic movements.

In the present study, we consider *ragas*/*makam*s that use almost the same scale intervals (relative to the chosen tonic note) in the Hindustani, Carnatic and Turkish styles. There are a total of 180 distinct excerpts *alaps*/*taqsim* (30-40s each) equally distributed across styles obtained from 78 concerts by 47 different artists. We choose sets of corresponding ragas across the three styles, in terms of the musical scale intervals, to avoid biasing listeners in the perceptual tests.

III. Perception Testing

To validate our hypothesis that melody is sufficient to capture the style distinction, we carried out perceptual listening tests. To avoid any bias towards artist identity, voice quality, and pronunciation, the melodic contour, extracted using a predominant pitch detection algorithm, is re-synthesized using a uniform timbre vowel-like sound using 3 harmonics of equal strength before being presented to listeners. Similar re-synthesis methods have been used in other works in melodic content analysis [11]. The pitch tracking was carried out based on the harmonic spectral characteristics of the singing voice using [12]. A stage of temporal smoothing is applied based on expected local pitch dynamics. The detected pitch contour was manually corrected by listening to the original and re-synthesized tracks to avoid any pitch error based bias. The amplitude of the re-synthesized tone, however, follows that of the singer's voice (extracted along with the pitch) since it plays a role in melody perception. We administered a total of 140 perceptual tests across 127 listeners whose background ranged from highly trained in one of the genres (> 3 years) to general listeners with no specific exposure to a particular style. An online interface with randomized presentation of the stimuli was used to elicit the listener judgements with the assurance of prizes to the top 2 performers in terms of achieving the highest accuracy in distinguishing the three categories. Giving such motivation resulted in the listeners trying hard to give the best performance and avoid sub-par performances of crowd-based testing which can occur in interfaces such as Amazon Mechanical Turk.

The dataset of 180 *alap* clips was divided into 10 sets. Each set contained 6 clips of each style from the dataset with almost equal number of clips for each *raga*. The listeners had to listen to at least first 10s of a clip before marking their decision as Hindustani (H), Carnatic (C), Turkish (T), Not Hindustani (NH), Not Carnatic (NC), Not Turkish (NT) or Not Sure (NS). We observe that listeners are able to identify the style with an average accuracy of 69% (Correctly identifying the clip as H, C or T). The actual distribution of group-wise numbers and recognition accuracies was as follows: lay listeners (63): 60%; listeners with $<$ 3 years training (49): 67%; and listeners with $>$ 3 years training (28): 77%. The trained category participants were familiar with H or C but none of them were trained in T. We see, as we might expect, that more the music training of the listener in any style, higher is their performance in discriminating between melodic styles. The clips where the trained listeners with experience $>$ 3 years failed were truly confusing as they had mixed characteristics of both the styles for e.g. Hindustani clip having a lot of *gamak* (rapid oscillatory pitch movement) which is typically seen in the Carnatic style. The confusions between H-C and C-T are comparable, with the confusions between H-T being least. The lay listeners had more confusion in H-C than in C-T which was seen from NT option being marked more as compared to NC and NH.

IV. Melodic Feature Extraction

Melody is defined by Poliner et al. [13] as "the single (monophonic) pitch sequence that a listener might reproduce if asked to whistle or hum a piece of polyphonic music. Pitch is the perceptual correlate of the fundamental frequency of a periodic sound. All musical sounds are periodic, and pitch is the property which differentiates sounds of the same timbre and loudness. The melody is represented as a time series of pitch values corresponding to the main or predominant voice computed on the audio signal at equispaced intervals [13][14][15]. In addition to the pitch, the second dimension of the extracted melody is the associated energy dynamics. In case of polyphonic audio, this can be computed as the short-time energy values of the melodic voice obtained as the total energy in the regions of the harmonic locations of the melodic pitch.

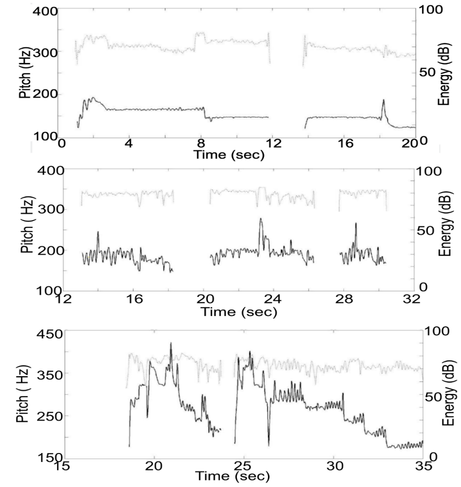

Fig. 1. Melody contours of Hindustani, Carnatic and Turkish style music. The proportion of steady notes is seen more in Hindustani with least in Carnatic. The dotted contours show the corresponding energy variations.

A. Preprocessing: Melody extraction

The database in our case consists of *alap* and *taqsim* recordings which do not contain percussive accompaniment, but only the drone (tanpura) and occasionally additional melodic accompaniment such as the violin. The challenges to pitch detection algorithms in tracking melody for Indian and Turkish music are the rapid pitch fluctuations and presence of melodic accompaniment [16][17]. The melody extraction algorithm must identify the singing voice regions and obtain a robust estimate of the pitch of the voice in the form of a continuous pitch contour sampled at 10 ms intervals [12]. Due to the difficulty of setting optimal analysis parameters across, and within, the variety of music pieces in our study, we adopt a semi-automatic approach for pitch detection [14]. This allows us to focus on our research for melodic features that can be computed on the detected pitch contour.

B. Local contour-shape features

The melodic contour in Indian classical vocal styles comprises steady and ornamented (*gamak*) regions. These localized characteristics are used to design discriminative features to classify Hindustani and Carnatic styles [10]. The proportion of steady notes to the *gamak*/ornaments is higher in Hindustani *alap* section as opposed to that in Carnatic *alap* or Turkish *taqsim* as seen in Figure 1. The regions not corresponding to the steady notes are expected to be captured using the *gamak* measure. The *gamak* measure characterizes the oscillatory behavior of the pitch contour modulations.

The pitch contour segments that are labelled as non-steady (*gamak*) region are analyzed for rate of pitch modulation. The Fourier spectrum of the temporal pitch trajectory, sampled every 10ms, shows clear peaks whenever the region is characterized by uniform oscillations. The energy ratio ER is calculated for pitch values in 1s window with 0.5s hop (i.e. 2 frames per second) by taking its Fourier transform. For calculating the ER of a segment, we take the ratio of the energy of oscillations in the regions in 3-7.5 Hz, normalized by the energy in the 1-20Hz frequency region as shown below.

$$ER = \frac{\sum_{k_{3Hz}}^{k_{7.5Hz}} |Z(k)|^2}{\sum_{k_{1Hz}}^{k_{20Hz}} |Z(k)|^2} \tag{1}$$

where $Z(k)$ is the DFT of the mean subtracted pitch trajectory $z(n)$, with samples at 10ms intervals, and k_{fHz} is the frequency bin corresponding to k Hz.

The percentage of ER computed that cross a certain threshold serves as an indicator of the vocal style. We define the *Gamak Measure* as

$$GamakMeasure = \frac{Number of ER > x}{Total number of ER computed} \tag{2}$$

The threshold x was varied from 0.1 to 0.9 to empirically set its value as 0.3 to achieve best separation between the oscillatory and relatively slowly varying segments. The *Gamak Measure* is seen to be high for Carnatic style ornaments while it is least for Hindustani style due to the presence of glides. The measure takes on intermediate values for the Turkish style where the oscillatory segments fewer than in the Carnatic style as seen in Figure 1.

C. Distance of maximum histogram peak from the tonic

A study on pitch intervals distribution in style perception was done by Vidwans et.al [10]. Their observation was that in an unfolded histogram, the Hindustani *alap*s are concentrated in the region near the tonic while the Carnatic *alap* pitch distribution is closer to the upper octave tonic. Further, we observed that in the case of Turkish music that the range of the artist is predominantly in the higher octave for the *taqsim* section. The tonic of the clips can be detected by an analysis of the drone-only region or by a multipitch method [18]. The pitch interval between the most prominent peak in the unfolded histogram computed from the pitch contour and the tonic is then used as a feature in the style classification.

D. Melodic transitions

The feature described in section IV-B characterizes the melody on a small time scale interval for the nature of pitch movements. The distance of highest peak from the tonic discounts the temporal nature of the pitch contour by taking the pitch histogram. To obtain the desired coarse representation of melody, we use wavelet representation of the pitch contour. Wavelets have been used to model time series data [19] capturing variations at different scales of resolution. The Haar wavelet function is appropriate due to its piecewise constant nature, which facilitates the quantitative analysis of melodic transitions in a stepwise fashion. The vocal regions are first concatenated and the approximation coefficients corresponding to a Level 5 approximation are obtained. Given that the pitch contour is sampled at 10 ms intervals, the piecewise output of a Level 5 approximation will have a length of 2^5 = 32 samples, corresponding to a time resolution of 320 ms which can be considered a reasonable approximation to a sequence of stable note events. In the case of Turkish Music, the transitions are moderate as compared to Carnatic music. To capture the trend, the number of upward transitions greater than 1 semitone (minimum possible jump in a *raga*) in the approximated pitch contour, normalized by the duration of the vocal region of the music piece is taken as a feature.

E. Energy based feature

Turkish music is marked by oscillations in amplitude or energy. These regular fluctuations of a note of fixed pitch are termed tremolo. To get an estimate of the periodicity of the energy contour, the energy contour is subtracted from its median filtered version and the number of zero crossings are used as a feature. The length of the median filter is decided empirically by considering the maximum accuracy achieved in the classification. The median filter length was kept to be 1s over energy contour values.

F. Microtonality based measure

Turkish music uses 53 Holdorein commas [20] i.e. it has many more musical pitch intervals than in the case of Hindustani and Carnatic music. The histogram of the notes in case of three styles shows that the Hindustani music notes are located closer to the equitempered locations while the Carnatic music notes can be slightly more removed [21]. However, in Turkish music, the note intervals are seen to displaced from the equitempered location by more than 30 cents [20]. For features based on the presence of micro-tones, we need to locate the peaks in the histogram of the audio. Unlike the context of Koduri et. al. [21] , our histogram is noisy due to the short duration of the audio excerpt, making peak picking more challenging. We overcome the noisiness by first detecting the peaks in the median filtered histogram. We then revisit the original noisy histogram to search for the highest peak in the 30 cents vicinity around the previously detected peak as shown in Figure 2. A similar approach was used by Turnbull et. al for picking boundaries for music segmentation in pop music [22].

We define four features based on the following definition:

$$l'_k = f\left(l_k\right) = mod(l_k, 100),\ if\ l'_k < 50,\ else\ l'_k = 100 - l'_k \tag{3}$$

where l_k is the location of the peak detected in the unfolded histogram. The mod operation reduces a note to its micro-tonal information only. For example, the two pitch intervals 110 cents and 290 cents will be represented by 10 cents in terms of their micro-tonal location with reference to the equi-tempered grid. Each of the features requires the peak detection step as described above. When there are micro-tones present,

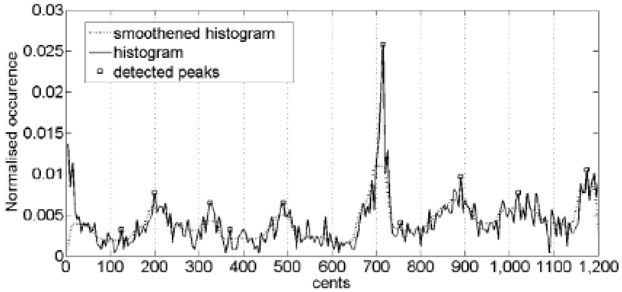

Fig. 2. Illustration of steps in detecting peaks in a folded histogram for *taqsim* section of *makam rast* by artist Hafiz Sesyilmaz (8 cent binning)

the location of these peaks are expected to lie far from the equitempered grid.

The first micro-tonality feature is the maximum deviation of the peaks present in the histogram from each other which will capture the presence of any notes which have deviated from the equitempered grid. This feature does not use explicit information about the tonic of the piece. A high value of the maximum inter peak deviation ($MIPD$) indicates the strong presence of the micro-tones in the piece. Out of the N peaks, we compute the maximum of the difference in locations l_i and l_j.

$$MIPD = max(f|l_i - l_j|), i \neq j, i < j < N \tag{4}$$

We propose three more features that use information about the tonic. They capture the deviation of the detected peaks from the pitch histogram from the equitempered grid obtained from the tonic information. The essence of these features is to capture the micro-tones rendered but in a different manner. Maximum peak deviation (MPD) calculates the maximum deviation of the peak location from the equitempered grid as

$$MPD = max(l'_k) \tag{5}$$

We do not take the mean as the number of distinct peaks in a audio for a very small excerpt may vary according to the *raga*/*taqsim* being chosen. Instead of taking the maximum value we can assign a weight of the deviation by the height of the peaks. Thus weighted peak deviation (WPD) is calculated to give importance to the peak with higher density in the spectrogram. The deviation of a peak is its distance from the equitempered scale i.e. from the nearest 100 cent interval, and f_k is the height of the peak in the corresponding histogram. Thus weighted peak deviation is defined as,

$$WPD = \frac{\sum_{k=1}^{N} f_k l'_k}{N} \tag{6}$$

We can find the micro-tonality of the rendered notes even without picking peaks in the histogram. This can be calculated by the equitempered note density feature (ED) which is the density of the notes lying in the vicinity of the equitempered location to the rest of the locations

$$ED = \frac{\sum_{k \in A} H_k}{\sum_k H_k} \tag{7}$$

where $k \in [1, 1200]$ and A = set of all $k < dev$ from equitemptered scale (at 0,100, 200 cents etc.) with dev chosen to be 30 cents, and H_k is the histogram of the notes present in the pitch contour, folded back to an octave (being represented by 1200 cents). We carried out feature selection using mutual information based criterion for considering information gain in WEKA toolbox. The final features selected were the top two features among the four microtonality based features viz. maximum peak deviation and equitempered note location density, along with other features.

V. CLASSIFICATION RESULTS AND DISCUSSION

Automatic classification is achieved based on quadratic discriminant analysis. We report two distinct sets of results. The first set is based on the automatic prediction of the genre label of each audio excerpt based on the provided track metadata, i.e. the genre as intended by the artist themselves. The second set of results involve the prediction of the listener judgements presented in III. Given that listeners achieve less than 70% accuracy when their predictions are compared to the provided music metadata labels, we can expect to see different performances in the automatic classification. In each case of the two distinct 'ground truths', a quadratic classifier (with a full covariance matrix) was trained and validated on the **6-dimensional** melodic feature vectors representing the 180 music excerpts using 5 fold cross-validation.

A. Prediction of metadata labels

In addition to the melodic features, there exist timbral based differences due to the varied nature of instrumentation, language and even recording conditions across the dataset. Widely used features in speech and music processing to model audio timbre are mel-frequency cepstral coefficients (MFCC). We used the average of 13-dimensional MFCC computed at 20 ms frame intervals across the original audio signal of each excerpt entire clip as the feature to capture the timbre. From the Table I, it is observed that addition of our novel pitch based feature is improving the performance over the baseline timbre features. Also, accuracy was compared for the various set of features namely (melodic features, timbre, and melodic+timbre based features) using the classifiers as seen in Table I. The analysis of confusion matrix of the classes showed that the confusion between the Carnatic and Turkish clips is more while the Hindustani class is well separated.

B. Prediction of subjective labels

We would like to see how well the proposed computational melodic features correlate with listener perception of the musical style from the synthesized melody. We carry this out by training and testing the classifier on labels obtained from the subjective listening tests. Considering that each synthesized audio excerpt was judged by an average of 14 listeners, and that different excerpts were judged by different sets of listeners, we need to consider ways to assign a single subjective label to each excerpt that reflects the collective listener confidence across the 3 styles for the excerpt in question. We remind the reader that each listener had to select one of the following label options for each presented excerpt: H, C, T, NH, NC, NT, and NS.

TABLE I
CLASSIFICATION ACCURACY (%) FOR MUSIC METADATA BASED LABELS AND PERCEPTION TEST BASED SUBJECTIVE LABELS

Labels	**Melody based features**	**Timbre based features**	**Timbre+Melody based features**
Metadata	92.5	85.0	94.5
Subjective	90.0	N.A.	N.A.

A reasonable approach is to assign a score to each of H, C and T categories based on the percentage of listeners choosing that label; next, to this we add the percentage of listeners who rule out the corresponding other labels (NH, NC, NT). For example the H score for an excerpt would be: %H + 0.5(%NC+%NT) - %NH - 0.33%NS. Finally, we assign a label to the excerpt corresponding to the highest of the 3 scores H, C and T computed for it. Using the assigned subjective labels as ground-truth, we obtain a classification accuracy of 90% in 5-fold cross-validation testing. It was observed that the Carnatic and Hindustani labels were never confused when predicting listener judgements. The largest confusions were between Hindustani and Turkish labels. We did not use timbre features in modeling listener judgements since the listeners had no access to the original music audio.

VI. CONCLUSION

We have successfully proposed melodic features to distinguish Hindustani, Carnatic and Turkish Music. The addition of energy and microtonality based features have helped in distinguishing Turkish musical style from Indian Classical Music. Listening tests confirmed that the style distinction is evident in melody. The high correlation between the listening tests and the classifier output indicates the high dependence of the features on melodic cues. Moreover, the clips with high confidence of classification also correspond to the clips which all the listeners have marked to be of that particular style emphasizing the fact that the model chosen is able to distinguish the styles as per the cues used by the listeners. Addition of timbre based features over the melody based features shows considerable improvement to the baseline features.

VII. ACKNOWLEDGEMENTS

The contribution of Prithvi Kantan to the processing of the perception experiment data is gratefully acknowledged.

REFERENCES

[1] J. C. Ross, T. Vinutha, and P. Rao, "Detecting melodic motifs from audio for hindustani classical music." in *ISMIR*, 2012, pp. 193–198.

[2] P. Chordia and A. Rae, "Automatic raag classification using pitch-class and pitch-class dyad distributions," in *Proceedings of the International Symposium on Music Information Retrieval, Vienna, Austria*, 2007.

[3] G. K. Koduri, S. Gulati, and P. Rao, "A survey of raaga recognition techniques and improvements to the state-of-the-art," *Sound and Music Computing*, vol. 38, pp. 39–41, 2011.

[4] S. Kini, S. Gulati, and P. Rao, "Automatic genre classification of north indian devotional music," in *2011 National Conference on Communications (NCC)*. IEEE, 2011, pp. 1–5.

[5] Y. Liu, Q. Xiang, Y. Wang, and L. Cai, "Cultural style based music classification of audio signals," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 57–60.

[6] A. Kruspe, H. Lukashevich, J. Abeßer, H. Großmann, and C. Dittmar, "Automatic classification of musical pieces into global cultural areas," in *Audio Engineering Society Conference: 42nd International Conference: Semantic Audio*. Audio Engineering Society, 2011.

[7] J. Salamon, B. Rocha, and E. Gómez, "Musical genre classification using melody features extracted from polyphonic music signals," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 81–84.

[8] B. R. B. J. Panteli, M. and S. Dixon, "Towards the characterization of singing styles in world music." in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 636-640). IEEE*, 2017.

[9] P. Agarwal, H. Karnick, and B. Raj, "A comparative study of indian and western music forms." in *ISMIR*, 2013, pp. 29–34.

[10] A. Vidwans, K. K. Ganguli, and P. Rao, "Classification of indian classical vocal styles from melodic contours," in *Serra X, Rao P, Murthy H, Bozkurt B, editors. Proceedings of the 2nd CompMusic Workshop; 2012 Jul 12-13; Istanbul, Turkey. Barcelona: Universitat Pompeu Fabra; 2012. p. 139-146*. Universitat Pompeu Fabra, 2012.

[11] K. Ganguli and P. Rao, "Discrimination of melodic patterns in indian classical music." in *Twenty First National Conference on Communications (NCC) (pp. 1-6). IEEE.*, 2015.

[12] V. Rao and P. Rao, "Vocal melody extraction in the presence of pitched accompaniment in polyphonic music," *IEEE transactions on audio, speech, and language processing*, vol. 18, no. 8, pp. 2145–2154, 2010.

[13] G. E. Poliner, D. P. Ellis, A. F. Ehmann, E. Gómez, S. Streich, and B. Ong, "Melody transcription from music audio: Approaches and evaluation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1247–1256, 2007.

[14] S. Pant, V. Rao, and P. Rao, "A melody detection user interface for polyphonic music," in *2010 National Conference On Communications (NCC)*. IEEE, 2010, pp. 1–5.

[15] J. Salamon and E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1759–1770, 2012.

[16] V. Rao, P. Gaddipati, and P. Rao, "Signal-driven window-length adaptation for sinusoid detection in polyphonic music," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 342–348, 2012.

[17] V. Rao, C. Gupta, and P. Rao, "Context-aware features for singing voice detection in polyphonic music," in *International Workshop on Adaptive Multimedia Retrieval*. Springer, 2011, pp. 43–57.

[18] J. Salamon, S. Gulati, and X. Serra, "A multipitch approach to tonic identification in indian classical music," in *Gouyon F, Herrera P, Martins LG, Müller M. ISMIR 2012: Proceedings of the 13th International Society for Music Information Retrieval Conference; 2012 Oct 8-12; Porto, Portugal. Porto: FEUP Ediçoes; 2012*. International Society for Music Information Retrieval (ISMIR), 2012.

[19] M. Vlachos, J. Lin, E. Keogh, and D. Gunopulos, "A wavelet-based anytime algorithm for k-means clustering of time series," in *In proc. workshop on clustering high dimensionality data and its applications*. Citeseer, 2003.

[20] B. Bozkurt, "Pitch histogram based analysis of makam music in turkey," *Proceedings of Les Corpus de l'oralité, Strasbourg, France*, 2011.

[21] G. K. Koduri, J. Serrà Julià, and X. Serra, "Characterization of intonation in carnatic music by parametrizing pitch histograms," in *Gouyon F, Herrera P, Martins LG, Müller M. ISMIR 2012: Proceedings of the 13th International Society for Music Information Retrieval Conference; 2012 Oct 8-12; Porto, Portugal. Porto: FEUP Ediçoes, 2012*. International Society for Music Information Retrieval (ISMIR), 2012.

[22] D. Turnbull, G. R. Lanckriet, E. Pampalk, and M. Goto, "A supervised approach for detecting boundaries in music using difference features and boosting." in *ISMIR*, 2007, pp. 51–54.