HIERARCHICAL CLASSIFICATION OF CARNATIC MUSIC FORMS

Ranjani. H. G.
Dept. of Electrical Communication Engineering,
Indian Institute of Science, Bangalore - 12, India
ranjanihg@ece.iisc.ernet.in

T. V. Sreenivas
Dept. of Electrical Communication Engineering,
Indian Institute of Science, Bangalore - 12, India
tvsree@ece.iisc.ernet.in

ABSTRACT

We address the problem of classifying a given piece of Carnatic art music into one of its several forms recognized pedagogically. We propose a hierarchical approach for classification of these forms as different combinations of rhythm, percussion and repetitive syllabic structures. The proposed 3-level hierarchy is based on various signal processing measures and classifiers. Features derived from short term energy contours, along with formant information are used to obtain discriminative features. The statistics of the features are used to design simple classifiers at each level of the hierarchy. The method is validated on a subset of IIT-M Carnatic concert music database, comprising of more than 20 hours of music. Using 10 s audio clips, we get an average f-ratio performance of 0.62 for the classification of the following six typesof Carnatic art music: */AlApana/, /viruttam/, /thillAna/, /krithi/, /thani-Avarthanam/* and */thAnam/*.

1. INTRODUCTION

Carnatic classical music is an aesthetic art form of South India. A typical Carnatic concert is rich in music content, and has structures at various levels. For example, use of various */rAga/* [1] (melodic framework) and */tALa/* (rhythmic framework) in a concert is well known [1]. Additionally, the concert can be lead-vocal or lead-instrumental in nature [2]. Apart from this, a typical concert comprises of a mix of various forms of Carnatic music, such as, */AlApana/, /thAnam/, /krithi/, /viruttam/, /thani-Avarthanam/, /thillAna/* and */swarakalpana/* [3]. All of these forms fall under any of two broad categories: */manOdharma/* (extempore) or */kalpita/* (those already composed) rendering [1,4].

In this work, we explore different features from the music signal to classify the given piece of Carnatic vocal concert music in one of the several forms, i.e., */AlApana/*, */thAnam/*, */krithi/*, */viruttam/*, */thani-Avarthanam/* or */thillAna/*. This kind of classification will find application in segmenting a concert into different parts, as many students of music and connoisseurs (*/rasikA/*) want to concentrate on one of the parts for either learning or enjoyment. Similarly, people would like to access these different parts of a concert from a database of Carnatic music, again for specific listening or improvising.

To the best of our knowledge, there is little or no work to automatically classify a given piece of Carnatic music into its various forms or classes. There exists no single feature set that can classify all the classes, owing to the varying nature of the signals. In addition, as can be expected of music, a sub-set of the classes exhibit common overlapping structures, while the rest differ. A hierarchical approach can organize the classes to groups and attribute each such group with a discriminative feature set based on the domain-knowledge. We propose acoustic features to represent and distinguish each class at each hierarchy level.

The paper is organized as below: Details of the forms of music is elaborated in Section 2. In section 3, we present the proposed hierarchical structure. The feature set extracted and the design of classifiers are detailed in Section 4. Section 5 summarizes the experiments along with the results followed by discussion and conclusions in section 6.

[1] All the Carnatic music terms are written in emphasized text within /./, to indicate phonetic pronounciation

2. CARNATIC MUSIC FORMS

Carnatic music has a long, rich and illustrious tradition and hence many forms of the music exist through the */guru-shishya parampara/* [1, 4]. Broadly, in a concert which lasts about 2-3 hours, some of the main forms are presented using construed musical ideas of the performer. These are sung (or played) in the context of different */rAga/* and different */sAhitya/* (lyrics) to create an enjoyable concert for the audience.

We expand on the details of some of the forms in a Carnatic music concert. These are, however, by no means exhaustive. One minute long example excerpts, for each of these forms, can be downloaded from:
http://www.ece.iisc.ernet.in/~ranjanihg/audio/index.html

2.1 */AlApana/*

/AlApana/ is a purely extempore melodic form of music. The exposition of a */rAga/* (melodic framework of Indian classical music) with no rhythm is called the */AlApana/*. It comprises of a sequence of phrases sung intended to create a mood for the subsequent composition, i.e., */krithi/*. The performer chooses the phrases based on the grammar of

the */rAga/* and elaborates systematically. The duration of an */AlApana/* can vary from 1 - 45 minutes and primarily depends on the taste and mood of the performer and also the subsequent composition. Typical vocal syllables used in */rAga AlApana/* are */ta/, /da/, /na/ and /ri/.*

In vocal concerts, the */AlApana/* comprises of extempore presentation of phrases by the lead artist, which would be pursued by the melodic-instrumental accompanist(s); occassionally this is followed by a solo */AlApana/* performance by the melodic-instrument accompanist(s) too.

2.2 */thAnam/*

A */thAnam/* is also a melodic improvisation form in Carnatic music. In its present day form, */thAnam/* is rendered as a part of */rAgam-thAnam-pallavi/* (or popularly RTP) of a concert. A */thAnam/* blends melody with medium-paced rhythm (*/laya/*) and brings about the intricacies of the */rAga/*. The syllables used are */aa/, /nam/, /tham/, /na/, /thom/, /tha/, /nom/*, such that the audience perceive the word '*/anantham/*' (endless) and '*/Anandam/*' (happiness). Most often, */thAnam/* is rendered without percussion instruments.

2.3 */viruttam/*

A */viruttam/* is a devotional and metrical verse in one of the (Indian) languages, sung as improvised music. It generally is elaborated in a */rAga/* or a */rAgamAlika/* (where sequence of */rAgas/* used for each verse of a given composition). Though the verse has a metric structure, */viruttam/* is devoid of rhythm in the rendering and hence may not be identifiable in the song.

2.4 */thani-Avarthanam/*

In Carnatic concerts, */thani-Avarthanam/* are percussion solo performances which exhibit creative and technical skill. It, generally, follows the main performance in a concert. In presence of multiple percussionists, alternate performances followed by group is the convention. Through out the performance the tempo of the performance is maintained.

2.5 */thillAna/*

A */thillAna/* is a popular and energetic performance, originated for dance performance, and realized during the end of a music concert. A major part of this rhythmic performance, comprises of a limited set of beat synchronous melodic utterances of meaningless rhythmic syllables. Few example syllables are */dheem/ /ta/ /na/ /dhir/ , /tha/ /ki/ /ta/ /jham/*, and the likes. Another feature of this form is that the final verse of a '*/thillAna/*' consists of lyrics similar to a */krithi/*.

2.6 */krithi/*

A */krithi/* in Carnatic music is a pre-composed piece of music or */kalpita sangeeta/*. It comprises of sub-structures: */pallavi/*, */anupallavi/* and */charana/* which roughly correspond to refrain and verses in Western music. Some compositions also include */chittai swaras/*, which are made of only the Indian solfege syllables i.e.,(*/sa/ /ri/ /ga/ /ma/ /pa/ /da/ /ni/*). [2]

[2] Some more finer aspects to a */krithi/* such as */niraval/*, */varnam/*, */mangalam/*, */gIthe/* exist, which we have currently grouped into one class */krithi/*.

2.7 */swara kalpana/*

A */swara kalpana/* is also an improvised melodic performance with rhythm and the rendition uses only Indian solfege syllables. The performer must conform to the grammar of the chosen raga along with the rules of swarakalpana. This is an extempore performance or */manOdharma sangeeta/*, though for a naive listener, this form is very similar to the */chittai swara/* section of the */krithi/*, which is */kalpita sangeeta/*.

3. PROPOSED APPROACH

From the above discussed forms, it is clear that we have to examine both melodic properties as well as rhythmic aspects, in grouping the different Carnatic music forms. Further, we can exploit certain specific syllable vocalizations which have become well accepted conventions among singers. Because of such high level features of the music form, we cannot find one feature set which will permit us to do a single level multi-class classification. Additionally, features chosen for discriminating between a chosen set of classes, may have no impact on other classes, owing to the presence of simultaneous structures among the different classes. Instead, we resort to different feature set at different levels and propose a hierarchical approach to classify a given piece of Carnatic vocal music into one of the above forms (Section 2).

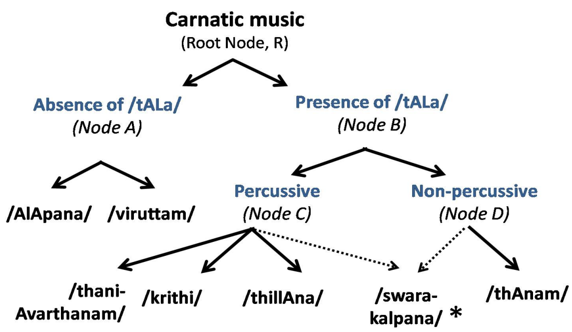


Figure 1. [Color online] Proposed hierarchical classification of Carnatic music forms. Text in blue indicates proposed hierarchy classification.

The classes are hiearchically divided as shown in Figure 1. At the root level (Node R), a grouping based on whether the form has a */tALa/* framework or not, leads */thAnam/*, */krithi/*, */thillAna/*, */swarakalpana/* and */thani-Avarthanam/* in the former (Node B) and */AlApana/* and */viruttam/* in the latter (Node A). The */tALa/*-based music pieces can be further classified based on whether they contain percussion instruments (in a concert) along with melody leading to a divide between */thillAna/*, */krithi/* & */thani-Avarthanam/* and */thAnam/*, seen as Nodes C and D in Figure 1. However, the presence of percussion for */swara-kalpana/* [3] is a choice of the percussion artist and hence can belong to either of the two nodes. At Node C, we can further introduce another branch based on the presence

[3] ,*It is not yet clear as to how to distinguish this from a possible */chittai swara/* and hence, we consider this as a part of */krithi/*, in this work.

of the instrument, and thus separate */thani-Avarthanam/* from the rest, if presence of percussion-only instrument is detected. This can be approached by incorporating instrument or instrument class identification [12]. However, the current approach intends to minimize the depth of the hierarchy so as to reduce propagation of errors [13].

4. FEATURE SETS AND CLASSIFIERS

We consider features relevant to each node in the hierarchy so as to emphasize the discriminative capabilities between the select classes.

4.1 Presence or absence of */tALa/*: (Root Node R)

4.1.1 Rhythmogram

Consider a t^{th} segment of a music signal $x(n)$ as, $s(t, m) = [x(n), n = (t-1)H_E + m], m = (1 : N_E)$, where N_E is the segment length and H_E is the segment hop. Let E_t be the short-time energy of the t^{th} segment of the music-signal, calculated as $E_t = \sum_{m=1}^{N_E} s^2(t, m)$. The window length N_E, corresponding to 32 ms, and a window hop H_E to 10 ms are considered, similar to speech analysis, motivated by the suitable frequency and time resolutions, respectively.

Consider the sequence, $[E_t,\ t = (1 : N)]$, which we further segment as: $E(r, \tau) = [E_t,\ t = (r-1)H_{ac} + \tau],\ \tau = (1 : N_{ac})$, where N_{ac} is the energy segment length and H_{ac} is the energy segment hop. We refer to the short-time energy signal as the 'novelty signal'. The parameters H_{ac} and N_{ac} must be chosen to suit the slowest tempo in the music. We have chosen, N_{ac} to correspond to 10 s and H_{ac} to 50 ms. We, then, compute the time-varying autocorrelation function of $E(r, \tau)$ given as: $R_{EE}(r, k) = \sum_{\tau=1}^{N_{ac}-k} E(r, \tau)E(r, \tau+k)$, and, $k = 0 : (N_{ac}-1)$. This is the autocorrelation of the r^{th} novelty signal segment for the k^{th} lag. Figure 2 shows this novelty-autocorrelation function as it evolves over time.

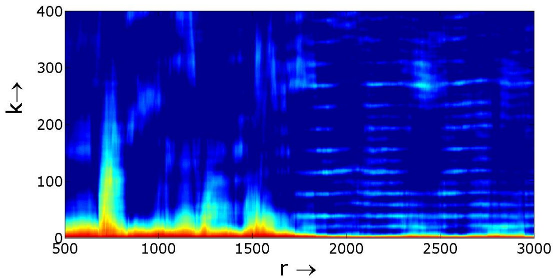


Figure 2. [Color online] Plot of $|R_{EE}(r, k)|$, the novelty-autocorrelation function for an audio clip consisting of */alapana/* followed by */krithi/*, as a 3D plot. Color=$|R_{EE}(,)|$.

The peaks in autocorrelation lags capture the long term periodicity of the energy signal, and are indicative of the metric structure of the */tALa/* [4]. This has been refered to as the rhythmogram [5]. (The method described above differs from [5], in taking autocorrelation of short-time energy in former, while use of autocorrelation of onsets is considered in [5].) The rhythmogram can be made discrete by picking P peaks of the rhythmogram for every r^{th} segment of the energy signal. We refer to this as thresholded binary-rhythmogram, denoted as $\mathcal{R}_b(r, \{l_i\})$ where $\{l_i\}$ indicates the i^{th} peak location for $i = [1 : P]$. */tALa/* structure, if present, can be well observed using this as a feature as shown in Figure 3.

In the Indian music context, tempo is a choice of the singer/ performer apart from being related to the composition. [5] Thus, for Node-R classification, we only need a measure to check the presence of */tALa/*; hence, we can use simple features (unlike in the literature which is mainly for Western music [8, 9] and uses the tempo to aid in classification or segmentation).

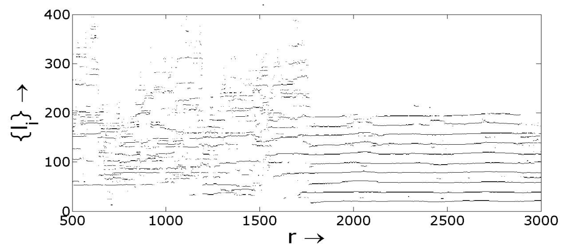


Figure 3. Plot of $\mathcal{R}_b(r, \{l_i\})$ for the audio clip corresponding to Figure 2, for $P = 10$

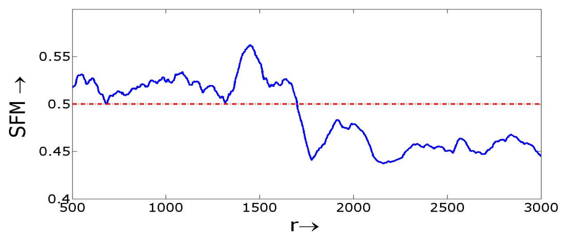


Figure 4. Plot of variation of Spectral Flatness Measure for the data corresponding to Figure 3, for $P = 10$

4.1.2 Classification at Root Node R

We can see from Figure 3 that the peak locations, $\{l_i\}$, show consistency in regions of music corresponding to presence of */tALa/* as against those which do not have */tALa/*. By constructing a histogram of the peak locations, we can expect peaky distributions, $P_r(k)$ for regions containing */tALa/*, while the non-*/tALa/* sections would have wider spread in the histogram. Motivated by this, we use spectral flatness measure (SFM or the GM:AM ratio) of the histogram of $\mathcal{R}_b(r, \{l_i\})$ which can be represented as:

$$P_r(k) = \frac{\text{\# of peaks with } l_i = k, \text{over } r = 1 : R}{\text{Total \# of peaks at all } k}$$

$$SFM(r) = \frac{[\prod_{k=1}^{K} P_r(k)]^{\frac{1}{K}}}{\frac{1}{K}\sum_{k=1}^{K} P_r(k)}, \text{ and, } K \triangleq N_{ac} - 1$$

The evolution of SFM across time for the music clip considered in Figure 3 is shown in Figure 4. We propose a low SFM as a score to indicate the presence of */tALa/*, with 0.5 value denoting boundary between the two classes. The effectiveness of the estimated distribution of peaks can

[4] It can be seen that slower the tempo, the peaks in $|R_{EE}(r, k)|$ will be farther apart.

[5] It can be argued that picking P peaks makes the feature invariant to tempo w.r.t further processing

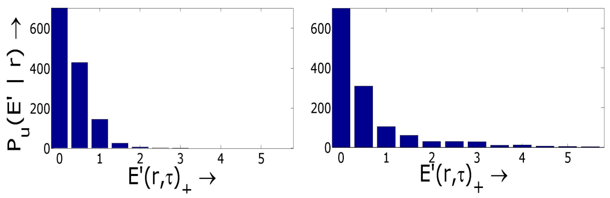
be increased by considering more points from $\mathcal{R}_b(r, \{l_i\})$. Hence, a trade-off between performance and duration, R can be expected. A value corresponding to $10\ s$ is chosen for R.

4.2 Percussive Vs Non-percussive: (Node B)

4.2.1 Onset strength

Percussion instruments give rise to stronger onsets, than melody based instruments, owing to their sharp attack pattern [6, 7]. The method to calculate onsets is briefed as follows: The energy signal, E_t, used for rhythmogram earlier is again considered. The positive first order difference function, $E'(r,\tau)_+ = (E(r,\tau) - E(r,\tau-1))_+$ [6] is computed; depending on the instrument and its corresponding attack pattern, this difference will be large (for percussive) or small (for non-percussive). The positive first order differences can be an estimate of onset strength at the r^{th} frame, or, $E'(r,\ \tau)_+ \in \mathbb{R}^+$.

Let $P_u(E'_+|r)$ denote histogram of $E'(r,\tau)_+$, (onset strength) for a 10 s audio clip containing solo voice (singing with */tALa/*) and solo voice with percussion is shown in Figure 5. It can be seen that these distributions can be modeled using an exponential distribution with a significantly different parameter, λ.



(a) (b)

Figure 5. Histograms $P_u(E' \mid r)$ for (a) soft onsets and (b) hard onsets of */tALa/* class of data from randomly chosen $10\ s$ segments of non-percussive and percussive kind.

Estimating the parameter of the exponential distribution, using the maximum likelihood framework, we get: $\lambda_+(r) = \frac{N_{os}}{\sum_\tau E'(r,\tau)_+}$, where N_{os} is the number of samples in the segment considered. Figure 6 shows a sample plot of variation of $\lambda_+(r)$ over successive r segments, for an audio concert clip reflecting the change from */thAnam/* to */krithi/*. The parameter is calculated for every 0.1 s, for a window size N_{os} corresponding to N_{ac}, ie., 10 s.

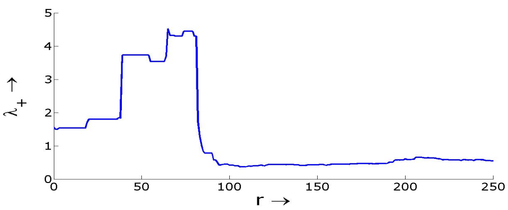


Figure 6. $\lambda_+(r)$ variation as a function of data segment, estimated for an audio clip transiting from */thanAm/* to */krithi/*.

A similar approach is seen in [10], where a Gaussian distribution is used to check the presence of onsets. Our approach differs not only in using exponential distribution, but also in using it for discriminating between soft and hard onsets.

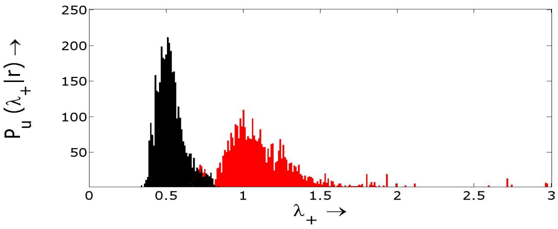


Figure 7. [Color Online] Histograms of, $P_u(\lambda_+ \mid r)$, depicting parameter distributions for 2 audio files representing (red) non-percussive and (black) percussive */tALa/* class.

4.2.2 Classification at Node B

From Section 4.2.1, estimated λ_+ parameter is seen to be effective in discriminating between forms of music (containing */tALa/*) that contain percussive and non-percussive instruments. Characterizing the same, across varied performances, shows that λ_+ distribution for percussive instruments is as shown in black and in red for non-percussive, shown in Figure 7. The deviation from the prototype parameters characterized by mean and variance of the distributions, provides correctness scores of classification.

4.3 */AlApana/* Vs */viruttam/*: (leaf nodes of A)

4.3.1 Formant peaks as features

The differentiator between */AlApana/* and */viruttam/* music pieces is that, the former uses limited syllables compared to the latter. */viruttam/*, which gives more importance to articulation, so that the lyrics are heard clearly. $(F1, F2)$ are estimated using Praat software [11] (Burg's method) [7], using a segment length $N_f = 32\ ms$, and segment hop $H_f = 10\ ms$. Figure 8 depicts a histogram of $(\frac{F2}{F1})$ of both the classes. It can be observed that there is lesser $(\frac{F2}{F1})$ variation in the alapana signal. We can thus use the rate of movement of $(\frac{F2}{F1})$ as a discriminative feature set in the rhythm-less form of Carnatic music.

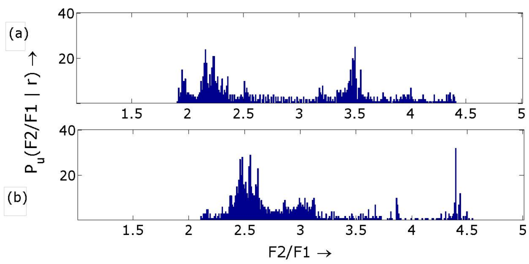


Figure 8. Histogram of $(\frac{F2}{F1})$ for a 10 s of audio containing (a) */viruttam/* (b) */alapana/* for a $10\ s$ second clips from randomly chosen example clips of the classes.

4.3.2 Classification for /AlApana/ *and* /viruttam/

Section 4.3.1 illustrated the effectiveness of $(F1, F2)$ to discriminate between */AlApana/* and */viruttam/*. It is observed that *AlApana* is more likely to have realizations of

[6] We define a function, $(x)_+ \triangleq \max(0, x)$

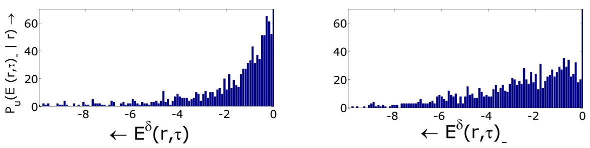
[7] Complexity in estimating reliable formants for high pitched sounds is well known and is beyond the scope of current paper.

/a/, yielding a choice $(\frac{F2}{F1})_a \sim \mathcal{N}(\mu_a, \sigma_a)$, with parameters $(\mu_a, \sigma_a) = (2.5, 0.3)$, with suitable Bayes' decision boundary. Thus, classification is based on maximum likelihood, where the likelihood is calculated for the ratio of mean of the formants extracted from 10 s data.

4.4 */thani-Avarthanam/ Vs /thillAna/ Vs /krithi/* (leaf nodes of C)

4.4.1 Cessation strength as feature

Between */krithi/* and */thillAna/*, from Sections 2.5, 2.6 and audio clips, one can prominently observe that */thillAna/*, inspite of presence of vocals, give a stronger perception of beat cycle than in the case of a */krithi/*. On closer observation, one can attribute it to the regular cessation of energy in */thillAna/*, owing to the syllables used (which mainly comprise of prominent stops as onsets of the syllables). This cessation of energy can be gauged as 'offset'. We observe that the strength of this offset is higher for */thillAna/* than for */krithi/*. Using terminology similar to Section 4.2.1, we can express $E^{\delta}(r,\tau) = (E(r,\tau) - E(r,\tau-\delta))_-$ [8], where $E^{\delta}(r,\tau) \in \mathbb{R}^-$. The parameter δ, is required as the cessation cannot be instantaneous, requiring δ duration to effect a noticable change; or in other words, it is a gradual offset. A sample histogram depicting this is shown in Figure 9. Hence, we can use the exponential parameter, λ_-, as a parameter to differentiate between */krithi/* and */thillAna/*, which consist of sustained energy owing to vocal or instrumental content.



(a) (b)

Figure 9. Histogram $P_u(E^{\delta} \mid r)$ of $E^{\delta}(r,\tau)$ for a 10 s of audio containing (a) */krithi/* (b) */thillAna/*

4.4.2 Dynamic range of the energy contours

The spectral content of */thani-Avarthanam/*, the percussion-only performance, comprises of mainly short lived bursts of energy, or "vertical lines" in the spectrogram. A periodic and large variation in the intensity as a function of time can be observed due to strong onsets and offsets, and no sustaining period, resulting in lower average intensity of the audio signal. Figure 10, depicts the distribution of average intensity in 10 s clips of */thani-Avarthnam/* and */thillAna/*, which can be used as a feature to distinguish these classes.

4.4.3 Classification combining the offset and dynamic range

Similar to Section 4.2.1, estimated λ_- parameter can discriminate between */thani-Avarthanam/* and */thillAna/* against */krithi/*. Discrimination between */thani-Avarthanam/* and */thillAna/* uses mean energy, $\mu(E)$, as a feature.

[8] We define, $(x)_- \triangleq \min(0, x)$

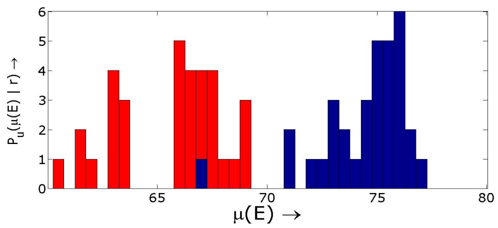


Figure 10. Histogram of average intensity, $\mu(E)$, in a 10 s of audio containing (blue) */thillAna/* and (red) */thani-Avarthanam/*

4.5 */thanam/* (leaf node of D)

For the present, */thanam/* is the only class considered at node D and hence on classifying a particular piece to contain */tALa/* without percussion at Node B, leads to this class in the proposed schema.

5. EXPERIMENTS AND RESULTS

All experiments are carried out on excerpts of vocal Carnatic music concerts. The database used comprises of randomly chosen 100 files from IITM concert database [14], amounting to 12 vocal concerts, by a total of 9 artistes (6 male and 3 female artists), and a total of ~20 hours duration. The tonic frequencies ranges from 129 to 205 Hz. The ground truth is generated after manually listening to the music pieces. The number of occurences and its duration, of each of the Carnatic form in the database, is detailed in Table 1. Also indicated are the number of non-overlapping 10 s clips of each of the music form, which is used for evaluating the algorithm. The inherent skewness in class distributions is to be noted. Change from one form to another in a Carnatic music concert is customary, and hence, any given 10 s clip can comprise of at most 2 classes. In such cases, the considered ground truth gives weightage to the temporally dominant class.

The results of the classification at each node is shown in Table 2. The performance of the proposed hierarchical model is evaluated at each node of the hierarchy, taking into account the performance of the parent nodes. However, errors at all nodes are given equal weightage, similar to "flat classification" evaluation [13]. The performance of the algorithm for each 10 s audio data is measured using the f-ratio [9]. The f-ratio is only indicative of the predictability of the algorithm to a particular class, and is invariant to tn, which is required in the unbalanced distribution of multi-class. Hence, we consider accuracy, $A = \frac{tp+tn}{tp+tn+fp+fn}$ also as a performance metric, which is not invariant to tn [15].

Though the performance of the proposed approach depends on the size of audio data analysed, we report results for 10 s clip only. An average f-ratio of 0.62 is seen for the proposed hierarchical classification, which is quite promising, given the complexity of the high level structures in the

[9] f-ratio is $F \triangleq \frac{tp}{tp+(\frac{fp+fn}{2})}$ while precision is $P \triangleq \frac{tp}{tp+fp}$. Recall, $R \triangleq \frac{tp}{tp+fn}$, & tp, fp, fn indicating the number of true positive, false positive & false negatives respectively.

music signal being examined. Also, the performance was observed to be not affected in terms of gender of the singer or the tonic variations, (at nodes R, B, C and D) as feature set considered is mostly independent of pitch.

Detailed analysis shows that */tAnam/*, */AlApana/* and */viruttam/* are the most confused classes. This is because the fp (of */tALa/* class) at Node R, trickle down to leaf node D always. This can be expected, as leaf nodes of A always have 'soft' onsets, thus reducing the predictability score of */thAnam/*. Similarly, fn errors (of */tALa/* class, mostly comprising of */tAnam/* class), and have higher syllabic repetitive features similar to that of */AlApana/*. Also, it is seen that */thani-Avarthanam/*, */thillAna/* and */krithi/* display better performance inspite of being a node below in the hierarchy. */krithi/* shows least recall performance which has been observed to be due to assumption of */chittai-swaras/* and */swara-kalpana/* in the */krithi/* class, resulting in misclassification to */thillAna/* class. Also, the performance at each node is sensitive to choice of decision boundary. Hence, it is crucial to have additional discriminative features along with high performance classifiers at each node of the hierarchy. The hierarchical framework is proposed for vocal concerts, and a generalizable framework across vocal and instrumental concerts is to be contemplated.

Table 1. Details of Carnatic music forms in the chosen dataset

Carnatic Form	Duration(min)	# of segments
/AlApana/	373	2236
/viruttam/	50	302
/thillAna/	23	136
/krithi/	666	3996
/thani-Avarthanam/	66	396
/thAnam/	22	132

Table 2. Performance of the algorithm on the Carnatic music forms

Carnatic Form	Precision	Recall	f-ratio	Accuracy
Presence of */tALa/*	0.89	0.97	0.92	0.88
Percussive	0.94	0.86	0.88	0.85
/viruttam/	0.80	0.60	0.63	0.80
/AlApana/	0.45	0.79	0.47	0.67
/thillAna/	0.87	0.65	0.7	0.74
/krithi/	0.93	0.5	0.72	0.71
/thani-Avarthanam/	0.95	0.7	0.75	0.94
/thAnam/	0.4	0.68	0.50	0.81

6. CONCLUSIONS

We have proposed a hierarchical approach to classification of Carnatic music forms, using signal processing features derived from energy contours, formant movements and parametric statistic distribution of such features. Based on the parameters of the statistical distributions, simple classifiers are proposed. An initial set of feature-vectors are designed, considering the "production model" of each particular class. An average f-ratio of 0.62 and accuracy of 0.77 is obtained, thus reflecting the promising nature of the feature vectors. The hierarchical approach to group the classes, based on the knowledge of the data is intuitive, and helps deriving better feature vectors for within the group classification. Hence, the trade-off between reduced accuracy with increasing depth of the hierarchy, against complex features for a flat single level classification can be observed as expected. More sophisticated classifiers in an automatic supervised framework can be developed based on these features to get better performance.

7. REFERENCES

[1] B. Benary: "Composers and Tradition in Karnatic Music," *Asian Music,* Vol. 3(2), pp 42-51, 1972.

[2] K. L. Armand and A. L. Armand: "One Hundred Years of Music in Madras: A Case Study in Secondary Urbanization," *Ethnomusicology,* Vol. 27(3), pp. 411-438, 1983.

[3] R. Morris: "Variation and Process in South Indian Music: Some Kritis and their Sangati," *Music Theory Spectrum* Vol. 23(1), pp. 74–89, 2001.

[4] Indira V Peterson: "The 'KRTI' as an integrative cultural form: Esthetic experience in the religious songs of two south Indian classical composers," *Journal of South Asian Literature*, Vol. 19(2),pp 165–179, 1984.

[5] K. Jensen: "A Causal Rhythm Grouping," *Computer Music Modeling and Retrieval*, Vol. 3310, pp 83–95, 2005.

[6] Moore, B. C. J. and Glasberg, B. R. and Baer, T.: "A model for the prediction of thresholds, loudness, and partial loudness," *Journal of the Audio Engineering Society*,Vol. 45.4, pp 224–240, 1997.

[7] J. P. Bello, L. Daudet et al.,: "A Tutorial on Onset Detection in Music Signals," *IEEE Trans. Speech and Audio Processing*, Vol 13.5, pp 1035–1047, 2005

[8] D. W. Ellis: "Beat Tracking by Dynamic Programming", *Journal of New Music Research*, pp 51–60 ,2007.

[9] G. Tzanetakis and P. Cook: "Musical Genre Classification of Audio Signals," *IEEE Trans., Speech and Audio Processing*, Vol. 10(5), pp 293–302, 2002.

[10] C. Duxbury, M. Sandler and M. Davies: "A Hybrid Approach To Musical Note Onset Detection", *Proc. of 5th Int. Conference on Digital Audio Effects (DAFx)*, pp 33–38, 2002.

[11] Boersma, P. and Weenink, D: "Praat: doing phonetics by computer [Computer program]," Version 5.3, 2013.

[12] G. Grindlay and D.P.W. Ellis: "Transcribing Multi-instrument Polyphonic Music with Hierarchical Eigeninstruments" *IEEE Journal Of Selected Topics In Signal Processing*, Vol. 5 (6), pp 1159–1169, 2011.

[13] E. P. Costa, A. C. Lorena et.al,: "A Review of Performance Evaluation Measures for Hierarchical Classifiers," *Proc. of the Association for the Advancement of Artificial Intelligence*, pp 1-6, 2007.

[14] A. Bellur, V. Ishwar and H. Murthy: "A knowledge based Signal Processing Approach to Tonic Identification in Indian Classical Music," *Proc. 2nd CompMusic Workshop*, pp 113–118, 2012.

[15] M. Sokolova, G. Lapalme: "A systematic analysis of performance measures for classification tasks," *Proc Information Processing and Management*, Vol 45, pp 427–437, 2009.