In Search of An Efficient ‘Chandas’ Identification System for Malayalam Poems from Audio

Anoop C S[1], Riyas K S[2]
[1, 2] Department of Electronics and Communication Engineering
[1] Government Engineering College Wayanad
[2] RIT Kottayam

Abstract- *The music genres in Kerala are characterized by a syllabic level rhythmic pattern called ‘Chandas’. The ‘Chandas’ basically is a syllabic level rhythmic pattern characterized by some rules for arranging light and heavy syllables in a poem or a song. This paper explains the two different approaches for identifying this rhythmic pattern of traditional Malayalam poems or songs from known music genres in Kerala. The first method obtains ‘Chandas’ by counting the number of syllables per two or four lines of a poem. The major step in this process includes detection of syllable nuclei in the audio signal. For this purpose intensity contours are used. The second method is an ANN based machine learning approach which can incorporate most statistical characteristics of these poems or songs.*

Keywords- Chandas, Poetic Meter, Malayalam, Vritham, Syllable nuclei.

I. INTRODUCTION

Even though modern people’s life is sufficiently supported by the amazing technological development, it is not possible to say that there was no scientific knowledge carried by ancient people. Because still we can see some of their marvelous works which attract us. It is even not dependent upon the social and economic status of those people. An excellent case is the music traditions existing in various parts of the world especially in culturally benefitted Indian states like Kerala. For example Vanjippattu is one of the music genres existing in Kerala associated with the Boat race culture developed in Kerala. The rhythmic pulses throughout the song are in somehow similar to the rhythm of how a boat is run. So analyzing the scientific or engineering background of these music traditions will be much interesting.

II. BACKGROUND LITERATURE

The science of chandas defines a syllable as a vowel preceded or followed by any number of consonant phonemes. The syllables are classified into Light and Heavy syllables. The time for uttering a light syllable is taken as one mathra. Chandas is a musical composition which defines a set of rules associated with the arrangement of light (laghu) and heavy (guru) syllables in a poem or a song. Chandas can be classified into two major classes -Mathra Chandas and Akshara Chandas. Mathra Chandas is determined by counting the number of mathras, taking laghu (short) as 1 and guru (long) as 2. This may be of two types:

1. Sama chandas- which has equal number of mathras in all the four quarters of the chandas.

2. Visama chandas- which does not have equal number of mathras in all the four quarters of the chandas.

Vritham is the chandas which has a specified sequence and fixed number of syllables in each quarter. These are of three types:

1. Sama vritham- in which all the four quarters of the chandas are identical, having identical pattern of short-long syllables.
2. Ardha sama vritham- in which the 1st and 3rd quarters as well as 2nd and 4th are identical with each other.
3. Visama vritham- in which none of the quarters of the chandas is identical with any other quarter of it.

Ganas are the triplets of the syllables having different combination of short and long syllables. Therefore a total of 8 ganas are possible. The gana system is widely used poetry as it makes the process of identifying the chandas much easier.

ചന്തമേ/റിയ പൂ/വിലും,ശ/ബളാഭ/മാംശല/ഭത്തിലും

സന്തതം/കരതാ/രിയന്നൊ/രുചിത്ര/ചാതുരി/കാട്ടിയും

101	001	010	010	100	101
101	001	010	010	100	101

0-laghu (light syllable)

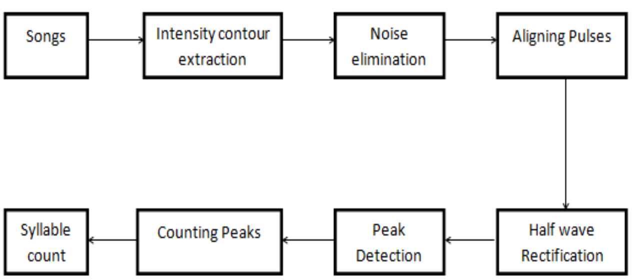
1-guru (heavy syllable)

Triplets used in poetry			
0	0	0	0
1	0	0	1
2	0	1	0
3	0	1	1
4	1	0	0
5	1	0	1
6	1	1	0
7	1	1	1

III. SYLLABLE COUNT APPROACH

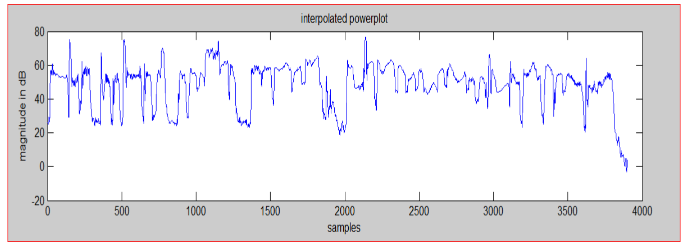
A. Block diagram

In this approach a measure of total number of syllables in 2 lines of a poem or a song is used for identifying the chandas .From this information chandas can be identified provided the poem is written in a chandas with fixed syllable count. Intensity contour are used for this purpose. This is because, the pulses present in an intensity contour actually correspond to the syllables. So by counting the number of pulses one can identify the chandas used in that song. The major work of this block is to detect syllables. But the contour contains a lot of unwanted or noisy information .Therefore syllable nuclei detection is a challenging task



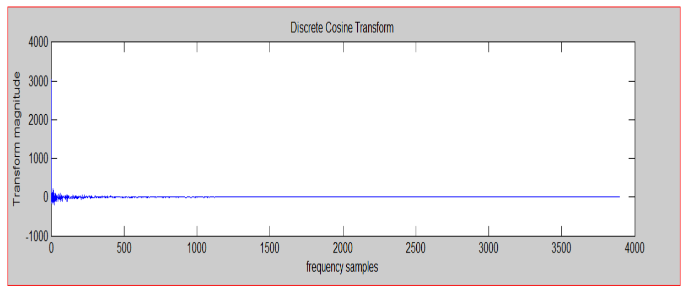
B. Power contour extraction.

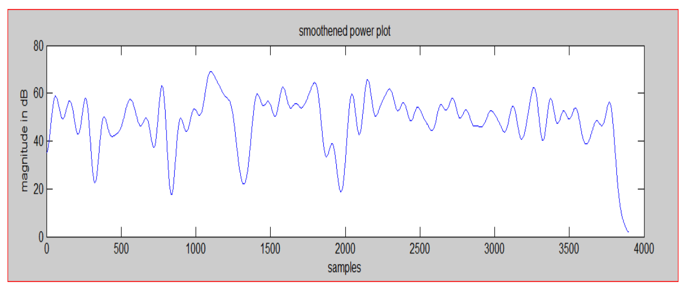
Power contour provides more information about the syllable nuclei than the original audio signal. This is also called intensity contour of the audio signal. It is usually expressed in dB magnitude. The dips in the contour actually represent the end of a syllable. So the syllable nuclei correspond to power pulses.



C. Low pass filtering to detect syllable nuclei

In this step the high frequency noise are eliminated. This makes the pulses clearly visible. Initially the DCT of the intensity contour is taken. Then the high frequency coefficients are removed. Now taking IDCT makes the syllable pulses visible clearly.





D. Pulse aligning

Since the pulses are aligned in a random fashion it is difficult to place a threshold. Therefore these pulses have to be aligned in the same level. In order to align the pulses in a straight line the slowly varying contents have to be removed. This can be done by subtracting out the very low frequency signal from the noise free power contour.

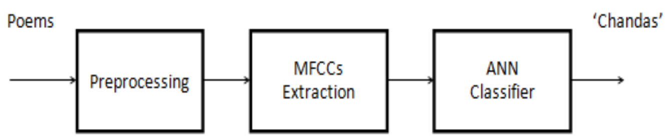
E. Half wave rectification and counting

In this step the contour is half wave rectified to get the exact sequence of pulses. Now syllable count estimation is done by counting the peaks.

IV. MACHINE LEARNING APPROACH

A. Block diagram

This method employs a machine learning approach for vritham identification. Therefore it will cover all the statistical characteristics of songs. The figure given below shows the major steps involved in this work.

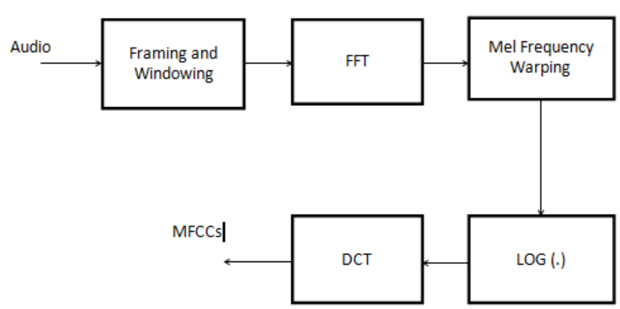


B. Database Collection

Database Creation is the initial step in any speech recognition or music information retrieval task. We made a lot of efforts to keep the quality of the data base. Because this may affect the progress and the final result of the work we have taken. The recording works are done in a studio, which ensured noise free samples. In this project the number of classes (vrithams) chosen are eight. So initially the lyrics of many songs from each class are collected. The number of candidates is 20 of which 10 are males and remaining are females. From each class 100 samples were collected and therefore a total of 800 samples are obtained. The amount of database plays an important role in the accuracy of the final system. For example the pitch value may vary largely from males to females. The samples are recorded in Lin16-PCM mono wav format using the software tool Sound forge.

C. Feature Extraction

MFCC is the most successful feature in most speech recognition and music information retrieval tasks. These are a set of coefficients obtained in a similar way how human ear works.



MFCCs are also increasingly finding uses in music information retrieval applications such as Genre classification, Audio similarity measure etc. In this work first 39 coefficients are extracted. Human vocal tract can be considered as a time varying filter. Therefore the speech or music signals are not statistically stationary. But it is observed that speech is stationary for a very short interval of time which is about 10ms. If the frame is much shorter enough samples will not be obtained, if it is longer the signal variations in the frame is too much. This is the reason why usually a 10-20ms framing is done for all recognition tasks. The next step calculates the power spectrum of each frame. This is supported by the working of human cochlea which vibrates at different spots depending on the incoming frequency. Depending on the position in the cochlea that vibrates, different nerves fire the brain understands that certain frequencies are present. The frequency resolution capacity of human cochlea becomes very poor at high frequencies. This is why clumps of periodogram bins are taken and summed them up to get an idea of how much energy exists in different frequency regions. Mel filter bank actually performs this task. The initial filter is very narrow. At higher frequencies filters get wider the ear becomes less concerned about variations in those regions. The Mel scale describes the arrangement of these filters and their widths. Taking logarithm is also due to a similar thing. It is based on the nonlinear loudness perception capability of human ear with the energy. The logarithm makes cepstral mean subtraction possible. That is why other nonlinear functions are not taken. Finally the DCT of the log filter bank energies are calculated which essentially de-correlates the filter banks.

D. Classification

After feature extraction the next step in machine learning is the classification. In this process, the classifier classifies the whole songs into different classes based on the features extracted. In this work there will be 8 classes as the numbers of chosen vrithams are 8.

Even though many classifiers are there, ANN and its derivatives show better results for Music Genre Classification. As the name indicates, Artificial Neural Network (ANN) shows resemblance with human neural networks. Similar to human brain neurons, artificial neurons are also designed. This similarity may be the reason behind the success of ANN in Speech Recognition and Music Information Retrieval tasks.
Just like how a biological neuron collects signals from sources, combines them, performs a generally nonlinear operation on the result and then outputs the final results, an artificial neuron also does similar things. Initially, inputs are multiplied by a weight. Then these results are summed, which when multiplied using a transfer function generates the result and then outputs.

An artificial network is a collection of large number of artificial neurons which are interconnected based on the statistical information obtained in the training phase. These neurons are represented as nodes in the network. A typical ANN network consists of at least 3 layers of nodes .The first layer represents the input layer and the final layer represents the output layer. The middle layer acts as a hidden layer which makes the network to perform more deeply. But there can be more than one hidden layer. The choice of number of hidden layers depends on the application. In this work a 10 layer ANN is used.

V. RESULTS

A. Syllable Count Approach

The syllable count estimation results show comparatively good results. But more erroneous counts are obtained in some cases. The accuracy of this system mainly depends on the design of filters used for syllabic nuclei detection task. In this work a fixed threshold is used for removing the high frequency noise regions. The design of a pulse aligning system also requires efficient filter deign. If these filters are made adaptive these effects can be overcome.

Chandas	No. of Syllables per 2 lines	Average Esti-mated count	Percentage of error
Druthakakali	11+11=22	23.7	.0077
Kavyanarthaki	-	-	-
Keka	14+14=28	29	.035
Mallika	18+18=36	36.05	.001
Manjari	12+10=22	22.79	.035
Nathonnatha	16+13=29	28.2	.0071
Omanakuttan	10+10=20	19	.05
Tharangini	-	-	-

B. Machine Learning Approach

Using standard set of 13 MFCC coefficients the vritham identification accuracy is somewhat poor. But if we include more coefficients the accuracy can be increased with the cost of computation time. Also by increasing the number of filters, the accuracy can be improved. In this work 39 coefficients with 40 filters are taken. The accuracy obtained is 87.685. Since this is a musical work, apart from usual MFCC set, delta and delta-delta coefficients will improve result. These additional coefficients will consider the changing behaviors of usual MFCC coefficients.

VI. CONCLUSION

This paper presents two different methods for identifying the chandas of a poem or a song. The database collected contains mainly songs from various music traditions of Kerala. The first method uses syllable count as measure for identifying the vritham. But this method may not be helpful for mathra vrithams. In addition to syllable counts if exact laghu-guru pattern can be identified along with the tempo, a perfect vritham identification system will be obtained. The final method is sufficient to incorporate most statistical properties of the song. If above two methods are combined with suggested modifications in proper manner a robust chandas identification system can be developed. Also this work can be extended to other music traditions in India as well. Since the work clarifies the relation between musical rhythm and poetic meter, the work may be a useful for rhythm analysis in major Indian traditional music traditions.

REFERENCES

[1] A. R. Rajaraja Varma, Keralapanineeyam, DC Books, Kottayam, 1968.

[2] A. R. Rajaraja Varma, Vrithamanjari, DC Books, Calicut, 1990.

[3] Ajay Srinivasamurthy, Andre Holzapfel and Xavier Serra "In Search of Automatic Rhythm Analysis Methods for Turkish and Indian Art Music", Journal of New Music Research, Volume 43, Issue 1, 31 Mar 2014.

[4] Rama N.and Meenakshi Lakshmanan, “A Computational Algorithm for Metrical Classification of Verse”, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 2, No 1, March 2010.

[5] Manu Madhavan, Mujeeb Rehman O and P C Reghu Raj, “Computing Prosodic Patterns for Malayalam”, National Conference on Indian Language Computing- NCILC 2013

[6] Rama N. and Meenakshi Lakshmanan, “An Algorithm Based on Empirical Methods, for Text-to-Tuneful-Speech Synthesis of Sanskrit Verse”, International Journal of Computer Science and Network Security, Vol.10, No. 1, January 2010.

[7] Ajay Srinivasamurthy and Xavier Serra, “A Supervised approach to hierarchical metrical cycle tracking from audio music recordings”, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florence, Italy, 2014.

[8] Athanasios Lykartsis and Alexander Lerch, “Beat histogram features for rhythm based musical genre classification using multiple novelty functions”, Proc. of the 18th Int. Conference on Digital Audio Effects (DAFx-15) ,Trondheim, Norway, Nov 30 - Dec 3, 2015 .

[9] Andre Holzapfel and Florian Krebs, “Tracking the "odd":metre interference in culturally diverse music corpus”, 15th International Society for Music Information Retrieval Conference (ISMIR 2014), Banglore, October, 2014.

[10] Anssi P. Klapuri, Antti J. Eronen, and Jaakko T. Astola, “Analysis of the Meter of Acoustic Musical Signals”, IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no.1, pp. 342-355, 2006.

[11] Dan P. W. Ellis, “Beat Tracking by Dynamic Programming”, Journal of New Music Research, vol. 36, no. 1, pp. 51-60, 2007.

[12] Aggelos Pikrakis, Iasonas Antonopoulos, and Sergios Theodoridis, “Music meter and tempo tracking from raw polyphonic audio”, Proc. of 5th International Conference on Music Information Retrieval (ISMIR 2004), Barcelona, Spain, October 2004.