

New Approach for Identification of North Indian Musical Instrument For Monophonic Tone

Vrushali Gunjal, Priyanka Desai, Jayshree Odhekar, Ghare Prachi

Computer Engineering, DYPCOE, Ambi, Pune, Maharashtra, India

ABSTRACT

A system has been already developed which can automatically identify the source of monophonic musical instrument sounds. Pre-processing of sound recordings includes calculation of the short term RMS energy envelope, Principal Component Analysis and Ratio of product transformations of the resulting Principal Components. An Artificial Neural Network [ANN] and a K-Nearest Neighbour Classifier [KNN] were compared to determine which provided best classification ability. The overall system performance was tested on the basis of sounds recorded from North Indian musical instruments chosen to represent the family of each major musical instrument and playing notes over the range of one octave under varying sound conditions. Classification precisions in the range 93.8 - 100 % were achieved. This paper provides the results of some primary work carried out to discover the potential of artificial neural network (ANN) and k-nearest neighbor classifiers (KNN) to monophonic musical instrument sounds. Some instruments were studied, chosen to represent each of the major North Indian musical instrument families. Recordings made were pre-processed by calculating the short term RMS energy envelope and performing Principal Component Analysis and calculating Ratio Product transformations of the resultant Principal Components. Two classifiers were presented with the information to determine which provided the best classification results for instrument identification.

Keywords: Music Information Retrieval, North Indian classical Music, MIR Toolbox, Timbre

I. INTRODUCTION

In the scientific and practical causes for building computer systems that can recognize and identify the instruments in music. More than a century after Helmholtz's revolutionary research, arguments still abound over the definition of musical "timbre," and on the relative importance of various acoustic features of musical instrument sounds. There are no developed scientific concepts about how humans, as listeners, identify sound sources, yet there are multiple software applications in the market which sound source identification by computer would be useful. For example, building a computer system that can explain musical multimedia data (Foote, in press; Wold, Blum, Keislar, & Wheaton, 1996) or transliterate musical performances for purposes of teaching, imaginary study, or structured coding with better concepts and models. By making such systems, we need to stand to learn about the human system we seek to imitate. This work is concurrently a scientific attempt to understand timbre by enumerating the relevance of various audile cues for

instrument identification and a practical stab to build a piece of an annotation/transcription system. The vast literature on the production and observation of musical instrument sound suggests many theoretically salient audio features of musical sound. In this system, we consider quite a few of these features and validate their extraction from musical instrument tones. The system can apply pattern-recognition techniques of both to evaluate the utility of these features in an identification situation and to build a useful classifier.

In this paper, Section II reviews the related work. The existing system is given in Section III. Section IV gives the proposed approach and algorithms. Section V discusses conclusions and finally acknowledgment & references.

II. RELATED WORK

An examination into timbre and instrument grouping has become more popular in last few years. Methods used in speech analysis were applied to musical echoes in order to construct a timbre space. The MelCepstrum procedure was applied to obtain parameters for the

description of sounds and then Self-Organizing Maps (SOM) and Principal Element Analysis were applied to this data then it produces a low-dimensional timbre space. This provides good supernatural analysis, but no temporal measures were combined in the analysis. Features were extracted from a wide range of musical instruments. These were analyzed using a variety of different grouping techniques. It was found that by using Quadratic Discriminant Analysis performed best in distinguishing between instrument families for classification.

Experiments to clearly identify specific musical instruments have also been reported in last few years. Brown distinguished between oboe and saxophone by calculating kestrel coefficients and applying a k-means algorithm to form clusters. Eronen and Klapuri examined in a wide range of temporal and spectral features from a large variation of north Indian instruments. Martin and Kim used features calculated from the log-lag correlogram rather than features based on the Short-Time Fourier Transform (STFT) to classify musical instruments hierarchically. Kaminsky and Materka examined the RMS of a group of instruments and reduced this data using PCA. This data was then classified using an Artificial Neural Network (ANN) and a K-Nearest Neighbour (KNN).

A. Proposal

This study plans to create and spontaneous musical instrument classifier by extracting and scrutinizing applicable features. These features are used as demonstrations of the timbre of the musical instrument. The efficiency of each feature is examined on a number of north Indian instruments as explained in this section.

B. Concept of Sound Timbre

The popular definition of Timbre is "the psychoacoustician's multidimensional waste basket category for everything that cannot be labeled pitch or loudness". The timbre typically refers to be a feature of a sound that allows us to distinguish between two sounds that are of the same pitch, loudness, and duration. Timbre is a lightly defined term representing many components together in a sound. In literature, no complete definition of timbre is available; neither any unit of timbre has been yet

defined. A lot of researchers have given names identical to timbre. Timbre has always remained a non-tangible entity.

C. Timbre Used For Instrument Identification

Musical instruments are identified by timbre identification. There has been a nonstop disagreement and discussion on whether human throat can be considered as a kind of musical instrument or not? We do not want to spoil in this debate of right or wrong. A lot of researchers have confidence in that timbre is to be related only to the sound generated by any instruments or objects. While on the other side if the human throat is expected to be a kind of wind instrument in which sound is produced because of the trembling occur in the vocal tract by movement of some quantity of air passing through it then, we may consider timbre features of the sound generated through the human throat.

Also, there are many characteristics of the timbre of a musical instrument if, we consider the timbre used. These are bridges, cotton threads etc. that are used to improve the sound generated through the instruments as in the case of a Tanpura, Sitar or Guitar instrument, for example. Ideally, in such situation, no two or more musical instruments can have same or similar timbre of the sound. Thus, even if we identify two musical instruments as, for example, two instruments like a ghatam and a dholak then still we are not able to recognize and identify which ghatam and which dholak if there are more than two such instruments.

III. EXISTING SYSTEM

Any sound is classified by an object called as Timbre. In music, timbre also knew as tone color or tone quality from psychoacoustics is the quality of a musical note, sound, or tone that differentiates the types of sound production, such as voices and string instruments, musical instruments, percussion instruments and wind instruments. The physical features of sound that define the discernment of timbre include spectrum and envelope. Timbre is multi-dimensional, fuzzy, undefined, unit less attribute of sound that exclusively identifies it. Some typical examples of timbre are "Soft, Loud, Low, High, Rough, Smooth" etc. Let it be a musical instrument sound, timbre plays very important role in classifying the sounds.

In this research paper, we have considered regular used North Indian Classical Music instruments like flute solo performances. For a monophonic instrument identification problem, the voice classification is very easy. While for North Indian Classical Music which is Homophonic (single sound streamline go together with by background instruments such as Tanpura, Harmonium, Guitar that is played continuously as the tune progresses) the classification of voice from the background voice is challenging task. The output sound is a combination of all sounds generated together giving a feeling of harmonious sound which is difficult to separate into parts or layers. Similar to singing voice there exist timbre for different musical instruments. The North Indian Classical music considers, for example, Flute, Violin, Harmonium, Sarod, Sitar, Shehnai, as leading Classical Instruments that are separately played. These instruments are used in the similar way a singer sings classical ragas. From instrument identification viewpoint a lot of research has been made towards a classification of an instrument in an orchestra recording or simple instrument dividation by identification of its timbre. The usual categorization or classification of the instruments is done into Flute, Violin, Clarinet, Trumpet etc. or the classification is into woodwind instruments, string instruments, bow instruments etc. Very insignificant research has been made into the identification of which flute? Which trumpet? Which violin? If, for example, we have five Flutes available, do we have the capability in our system that individually identifies the instrument units into Flute A, Flute B, and Flute C etc.? Researches, often declare usage of singing voice as a kind of musical instrument by saying that human voice: the most natural of all musical devices that are most widely used. But on the other side, there is minimum research and applications to detect separate 'Unit' of an instrument rather than a just type of north Indian musical instrument.

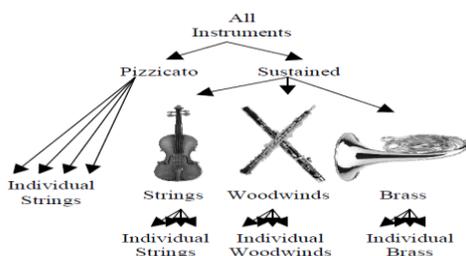


Figure 1: Classification of Musical Instruments

The north Indian musical instruments are also characterized by their timbers. The flute sounding husky and used for ragas containing minor notes of music are

undoubtedly completely different than the flue sounding immortal for raga's covering major notes. In this paper, we have used flute's performances from North Indian Classical Musical tones. It is to be stated here that the classical performances of any musical instruments playing North Indian classical Music are very similar to the performances of classical singers of this kind of music.

A classical singer uses Tanpura harmonium as a supplementary instrument and sings primarily long notes contained in the ragas. Then, after completion of nearly all the notes described by the raga's, that lengthy and sung adequately, the rhythm starts slowly and then regularly the speed of the rhythm and singing and the rate of change of musical note sung per unit of time increases accordingly. Similar performance is observed with classical performances of north Indian classical musical instruments such as Violin or Flute and much more.

IV. PROPOSED APPROACH AND ALGORITHM

In this approach, we propose to have specifically two databases each containing 50 files. Which files have wav format recordings each of 5 sec duration, 16 bit PCM and which is having 11,025Hz sampling frequency respectively. The Database-1 contains 5 different flute musical instrument (which are different units) recordings with 10 samples of each. The Database-2 contains 5 different flutes (Which are different units) recordings with 10 samples of each recording. The arrangement of both datasets done this way has many purposes to be served. First, the system will be tested on audio that contains a combination of musical instruments. Second, the performance, importance, and role of Timbral sound descriptors can be verified and tested. Then, the comparison of the performance of statistical classifiers can be done on the basis of instrument sound inputs. Also, we wish to confuse the recognition system with respect to, whether it is a Flute or Violin. By this way in Database-1 or in Database-2, if we have total 10 numbers of Flutes and if Flute 7 is identified as Flute 7 for example, then that means a musical instrument Flute unit of number 7 has been properly identified as Flute number 7. Thus by this procedure of audio input, this makes the same system to work as a musical instrument identifier.

In order to study the audio imaginative phases of this above-stated system, a hybrid selection algorithm is applied only on the audio descriptors, defined under the taxonomy of timber. As per MPEG standard, there are more than 50 sound descriptors and can be categorized according to a variety of taxonomy depending upon their point of view. In north Indian musical instrument identification research taxonomy of audio descriptors is surveyed that includes following audio descriptors under the title of Timbre in MIRtoolbox. The audio descriptors are: attack time, zero-crossing rate, and attack slope, roll off, roughness, brightness, MFCC. The attack time and slope are not useful for harmonic sound samples there for there are neglected giving total 6 audio descriptors including MFCC.

Algorithms:

A. RMS-

For a brief audio signal (frame) consisting of N samples, the amplitude of the signal measured by the root Mean square is delineated by equation mentioned below. RMS may be a measure of the loudness of an audio signal and since changes in loudness are vital cues for brand new sound events it is employed in audio segmentation. During this project RMS options, are accustomed observe boundaries among completely different musical instruments. The tactic for detecting boundaries relies on the difference measure of those amplitude distributions.

$$P = \sqrt{\sum_{t=1}^N x^2(t)}$$

B. PCA-

The aim of principal component analysis (PCA) is to scale back the spatiality of an information set that consists of an oversized variety of reticulated variables, whereas retentive the maximum amount as attainable the variation gift within the information set. This is often achieved by transformation with a brand new set of variables, the principal components (PCs), that are unrelated and ordered so the primary few retain most of the variation gift all told of the initial variables.

If PCA weren't used here, the ANN/KNN would wish to investigate RMS Energy values for every musical signal. With input values, the KNN would become comparatively massive and resulting training and test effortful and slow.

The PCs for the RMS Energy values were calculated using the matrix for the dataset, C, with the subsequent:

$$z = A^t x^*$$

Where: z could be a column matrix containing the PCs themselves, A could be a matrix that has columns consisting of the eigenvectors of the matrix, C, and x* could be a column matrix containing a consistent version of the dataset vector x.

Using above equation to see the PCs and so applying Kaiser's rule, solely three PCs are maintained for the RMS Energy information. For these three PCs, the additive share of total variation diagrammatic was 88.9%. Therefore the spatiality of the input data vector to the ANN/KNN has been reduced a substantial saving. To obtain some plan of the problem of discrimination of the four instruments supported these three PCs, XGOBI, an interactive dynamic graphics program for information visualization developed at Bellcore, was want to plot a 2D scatter plot of the three PCs for the ANN/KNN training dataset. Except for some overlap between the guitar and piano clusters, fairly sensible separation exists between instruments. This suggests the approach chosen for the system has potential.

C. ANN-

A multilayer perceptron ANN with sigmoid nonlinearity exploitation the back Propagation training rule was used because of the classification paradigm. This was used primarily as a result of its widespread use, quality and successful application by different researchers in playing similar classification tasks.

Its architecture was designed using the rules recommended by bird genus. With four instruments being classified, four output layer units (M) were needed. With 3 RMS Energy PCs, 3 input layer units (N) were necessary. One hidden layer (HL) was utilized, with the initial variety of units (h) being determined exploitation Maren's suggestion, i.e., $h = \sqrt{(N \times M)} \approx \text{four}$. The optimum design, but was finally determined exploitation empirical results.

During training of the ANN, pattern learning was employed in preference to batch learning. A learning rate, $\alpha = 0.25$ and a momentum term, $\eta = 0.15$ were used with but 30,000 displays to the ANN and with $\alpha = 0.15$ and $\eta = 0.08$ on the far side this value. All input file were scaled to the vary 0.1 - 0.9. Throughout testing, the ANN outputs (x) were the edge at a value of 0.5. For 0

$x < 0.5$, a zero result was assumed. Conversely, for $0.5 \leq x < 1.0$, a results of 1.0 was assumed.

D. KNN-

The K-nearest neighbor classifier is an example of a non-parametric classifier. The fundamental algorithmic program in such classifiers is easy. For every input feature vector to be classified, a search is created to search out the placement of the K nearest training examples, and then assign the input to the category having the biggest members during this location. Euclidian distance is usually used because the metric to measure neighborhood. For the special case of $K=1$, we'll get the nearest neighbor classifier, that merely assigns the input feature vector to the identical category as that of the nearest training vector.

The Euclidean distance between feature vectors:

$X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_n\}$ is given by:

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

The KNN algorithm, as mentioned earlier, is extremely straightforward yet rather powerful, and employed in several applications. However, there are things that require being thought of when KNN classifiers are used. The Euclidian distance measure is usually employed in the KNN algorithm. In some cases, use of this metric may lead to an undesirable outcome. For example, in cases where many feature sets (where one feature set has comparatively massive values) are used as a combined input to a KNN classifier, the KNN are biased by the larger values. This results in a really poor performance. A potential methodology for avoiding this problem would be to normalize the feature sets.

The aim is to use the KNN classifier for locating the category of an unknown feature X because it may be seen in the figure, of the nearest neighbor ($K=5$) four belong to a category and just one belongs to category b and therefore X is allotted to category a.

Some of the disadvantages of the K-nearest neighbour classifiers are:

Need the complete feature vectors of all training data once a brand new vector feature is to be classified and therefore massive storage needs. The classification time

is longer compared to other classifiers. The K-nearest neighbor classifier has some qualities that are vital like it needs no training and this can be useful particularly once a new training data is accessorial. Uses native info and therefore will learn complicated functions without having to represent them expressly. During this project, k-NN is adopted to permit the new data accessorial.

V. CONCLUSION

This paper has described some preliminary work using both ANN and KNN classifiers for the automatic source identification of monophonic musical instrument sounds. The results achieved are encouraging although somewhat surprising given that only temporal, but not frequency, data was utilized. This may be because of the limited range of instruments used, the laboratory controlled conditions of the instrument recordings created and also the smart discrimination obtained with the four instruments chosen.

VI. ACKNOWLEDGMENT

We are thankful to faculty of Computer Engineering Department, DYPCOE, Savitribai Phule Pune University for their support. This research paper would not be possible without all of them.

VII. REFERENCES

- [1] Dalibor Mitrovic, Matthias Zeppelzauer, And Christian Breiteneder, "Features for Content -Based Audio Retrieval," *Advances in Computers*, vol. 78, pp. 71-150, 2010.
- [2] Geoffroy Peeters, "A large set of audio features for sound description (Similarity and Classification) in the CUIDADO Project," Ircam, Analysis/Synthesis Team, 1PI Igor, Stravinsky, 75001, Paris, France, Analysis report V 1.0, 23rd April, 2004.
- [3] S., J. W. Beauchamp, S. Meneguzzi McAdams, "Discrimination of Musical Instruments Sounds Resynthesized with Simplified Spectrotemporal Parameters," *JASA*, vol. 2, p. 104, 1999.
- [4] S. H. Deshmukh, Dr. S. G. bhirud, "A Novel Method to Identify Audio Descriptors, Useful in Gender Identification from North Indian Classical Music Vocal," paper ISSN: 0975-9646.
- [5] T. H. Park, "Towards Automatic Musical Instrument Timbre Recognition," New Jersey, USA, 2004.

- [6] G. Peeters, "A large set of audio features for sound description (Similarity and Classification) in the CUIDADO Project," 75001, Paris, France, 23rd April, 2004.
- [7] Tae Hong Park, "'Towards Automatic Musical Instrument Timbre Recognition'," PRINCETON UNIVERSITY, PRINCETON, Thesis Report Nov 2004.
- [8] S. Deshmukh and S.G. Bhirud, "A Hybrid Selection Method of Audio Descriptors for Singer Identification in North Indian Classical Music", in Emerging Trends in Engineering and Technology (ICETET), Himji, Japan, 2012.