A Novel Approach of Classifying and Recognizing the Audio Scenario's Profile

Ajay Kadam
Student: Department of Computer Engineering
Dr. D. Y. Patil School of Engineering and Technology
Lohegaon, Pune, India

Ramesh M. Kagalkar
Professor: Department of Computer Engineering
Dr. D. Y. Patil School of Engineering and Technology
Lohegaon, Pune, India

ABSTRACT
The small piece of sound clip can provide lots of information regarding the background of sound when it is captured, different types of sounds in that clip etc. Only human can recognize the sounds from clip if he knows or have heard them before. But then also there is a limitation of humans, human can recognize up to certain amounts of sound clip if more clips are there human will get confused or can't distinguish them. Hence developed system which will store different sound sample in the database and will recognize the same if comes again in system as an input.

The aim of system is to identify the sounds in the given input sound clip, compare the extracted features with database samples and generate the proper text description for relevant sound with the image. Here developed system will convert the musical piece into byte array. Then time domain values will be converted into frequency domain with the help of modified FFT formulae. The chunks of 4096 bytes will be created for processing. After which system will store or match the top four values of each chunks in database for insertion or comparison of sound sample. The database is the basic need of the system. I have created database of sound over 800 samples with two distinct indoor and outdoor classes. The system has provision to add more samples in it from manual selection as well as by the recording real time samples. The system is showing the results in descriptive manner for different the different types of samples. If want to discuss the accuracy of system then it is 94% accurate for offline mode while in online mode its accuracy degrades to 50% because of noise issue.

Keywords
fingerprinting; Pure tone; White noise

1. INTRODUCTION
The proposed research objective is to add to a framework for programmed recognition of sound. In this framework the real challenge is to distinguish any information sound stream investigates it and anticipates the likelihood of diverse sounds show up in it. To create and industrially conveyed an adaptable sound web crawler a flexible sound search engine. The calculation is clamor and contortion safe, computationally productive, and hugely adaptable, equipped for rapidly recognizing a short portion of sound stream caught through a phone microphone in the presence of frontal area voices and other predominant commotion, and through voice codec pressure, out of a database of over accessible tracks. The algorithm utilizes a combinatorial hashed time-recurrence group of stars examination of the sound, yielding ordinary properties, for example, transparency, in which numerous tracks combined may each be distinguished.

The human movement is reflected in a rich mixture of acoustic occasions, delivered either by the human or by articles took care of by people, so the determination of both the character of sounds and their position in time can help identify and depict that human action. The improvement of frameworks for programmed grouping of a sound sign into classes of occasions, or the change of the exhibitions of programmed arrangement of a sound sign into occasions[1].

In proposed framework is to give the info as a sound stream then after preprocessing and highlight extraction sound stream is perceived and showed on the showcase that is brought from database which is now put away & right stable class is distinguished and showed on the screen. The example coordinating and acknowledgment could be possible through predefined database.

Generally to activate verbalization signal capture for verbalization apperception, the push-to-verbalize has been widely utilized in handheld mobile contrivances in which one can push a special button in the contrivance to activate or deactivate verbalization capture[6]. It is immune to the environmental noise, but may not the case virtually required in other consumer contrivances; for example, an interactive digital TV is supposed to heedfully auricularly discern all the time and automatically detect only human voice by denotes of the hands-free verbalization acquisition. In integration, it requires filtering out potential acoustic event sounds (e.g., hand-claps, phone ringing, door-slam, and so on) for the robustness of speech recognition system [5].

Not at all like other other audio or verbalization signals, sound events have a generally brief time compass. They are generally recognized by their exceptional spectra-fleeting mark. This paper proposes a novel order technique taking into account probabilistic separation bolster vector machines (SVMs). Here concentrate on a parametric way to deal with describing sound signs utilizing the dissemination of the sub band fleeting envelope (STE), and part methods for the sub band probabilistic separation (SPD) under the system of SVM. Here demonstrate that summed up gamma displaying is very much contrived for sound portrayal and that the probabilistic separation bit gives a shut structure answer for the figuring of difference separation, which colossally lessens computational expense. The directed investigations on a database of ten sorts of sound occasions. The outcomes demonstrate that the proposed characterization technique altogether beats traditional SVM classifiers with Mel-recurrence campestral coefficients (MFCCs). The quick processing of probabilistic separation likewise makes the proposed technique a conspicuous decision for online sound occasion acknowledgment [2].

Here think about the spectral and transient periodicity representations that can be utilized to portray the qualities of

the mood of a music sound sign. A nonstop esteemed vitality capacity speaking to the onset positions after some time is initially separated from the sound sign. From this capacity system register at every time a vector which speaks to the attributes of the nearby mood. Four capabilities are concentrated on for this vector. They are gotten from the plentifulness of the discrete Fourier change (DFT), the auto-relationship capacity (ACF), the result of the DFT and the ACF interjected on a half and half slack/recurrence hub and the connected DFT and ACF coefficients. At that point the vectors are inspected at some particular frequencies, which speak to different proportions of the neighborhood beat. The capacity of these periodicity representations to depict the musicality attributes of a sound thing is assessed through an arrangement errand. In this, system test the utilization of the periodicity representations alone, joined with rhythm data and consolidated with a proposed arrangement of cadence highlights. The assessment is performed utilizing commented and assessed rhythm. System demonstrate that utilizing such straightforward periodicity representations permits accomplishing high acknowledgment rates at any rate equivalent to beforehand distributed results [3].

A music piece can be considered as an arrangement of sound occasions which speak to both transient and long haul fleeting data. Nonetheless, in the assignment of programmed music class characterization, the majority of content order based methodologies could just catch fleeting neighbourhood conditions (e.g., unigram and bigram-based event insights) to speak to music substance. In this paper, Here propose the utilization of time-obliged consecutive examples (TSPs) as compelling highlights for music classification arrangement[12]. Above all else, a programmed dialect distinguishing proof method is performed to tokenize every music piece into an arrangement of shrouded Markov model files. At that point TSP mining is connected to find classification particular TSPs, trailed by the reckoning of event. Frequencies of TSPs in every music piece[15][21]. At last, bolster vector machine classifiers are utilized in view of these event frequencies to perform the arrangement assignment. Examinations led on two generally utilized datasets for music kind characterization, GTZAN and ISMIR2004Genre, demonstrate that the proposed strategy can find more discriminative fleeting structures and attain to a superior acknowledgment precision than the unigram and bigram-based measurable methodology [4].

2. FUNDAMENTAL APPROACH

- The proposed system gives hands on technical tool to elaborate and to find the classes of different from its sound pattern of a sample space.
- In this approach data base is created, which consists of .wave form of different activity, event, scenes sounds and its information of each sample model of various types.
- A Novel Approach For Classifying and Recognizing the Audio Scenario's Profile 18.
- Once the sound sample query is given to the system first it converts into .wave format and extract features. Each sound file is fingerprinted a process in which reproducible hash tokens are extracted.
- Both database and sample sound files are subjected to the same analysis. The fingerprints from the unknown sample are matched against a large set of fingerprints derived from the sample database. The system matches are subsequently evaluated for correctness of match.

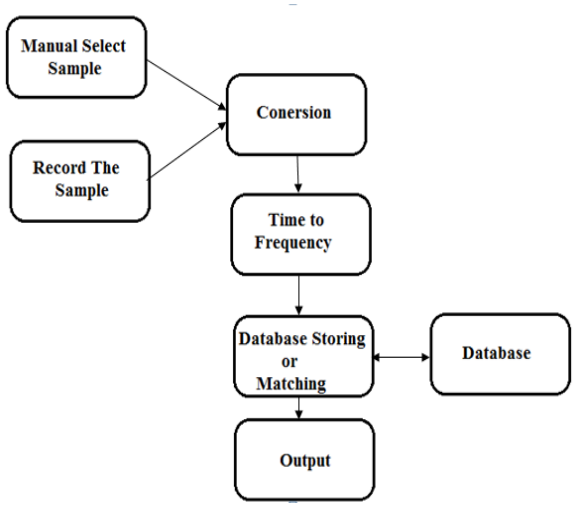


Figure 1: Block diagram of system.

3. ACTUAL PROCEDURE

The above figure 1 clearly shows the actual working with stepwise procedures of proposed system. Starting with the system needs the input sound. System has provision to give input manually as well as recording sound in real time. The input sound will be converted into byte format. These are the time domain values which will be later converted into frequency domain with the help of modified formulae of fft. lastly the chunks of 4096 bytes will be created and the upper 4 values (hash points) will be used for further processing. The extracted features will be matched with the saved hash points for finding out the result. The output of system contains the classification of relevant sound with its exact name along with its properties as well as relevant image.

4. IN PROPOSED SYSTEM, SYSTEM PERFORM TWO MAIN OPERATIONS ON SOUNDS THAT ARE TRAINING THE SOUND AND TESTING.

1. Algorithm for Training the Sound in Database

Input: Sound sample

Output: Text message

Steps:

1. Select sound sample
2. Convert to byte array
3. Generate song id
4. Determining matching points with respect to time
5. Save to database.

2. Algorithm for Testing the Sound with Database Samples

A. Offline

Input: Sound Sample

Output: Byte array, Hash values, matching sound of sample with details

Steps

1. Select the sound sample.
2. Convert to byte array.
3. Convert byte data from time domain to frequency domain.
4. Determine the key values (hash points).
5. Find exact match from database.

B. Online

Input: Recorded Sound Sample

Output: Relevant Matching Sound

Steps

1. Record sound sample from microphone.
2. Convert to byte array.
3. Convert byte data from time domain to frequency domain.
4. Determine the key values (hash points).
5. Find exact match from database.

5. MATHEMATICAL MODEL

A mathematical model is a description of a system using mathematical concepts and language. The process of developing a mathematical model is termed mathematical modeling.

Set Theory

1. Let S={ASI , T , DB , TR , TS , φ }

Where,

ASI is Audio Stream Input

T is Text

DB is Database

TR is Training Data Set

TS is Test Data Set.

Φis null or empty set

2. ASI={ASI1 ,ASI2 , ……, ASIn,φ }

Where,

ASI is Audio Stream Input

3. T={T1 , T2 , ……Tn , φ }

Where,

T is the Text that matches with the sample.

4. DB={ DS1 , DS2 , ……, DSn , φ}

Where,

DS is the Data Set containing n number of data set of Audio Streams.

5. DS= {ASIDS , ASDS , φ }

Where,

ASIDS is the Audio Stream Input Data Set.

ASDS is the Audio Stream Data Set.

The table 1.1 represents activities which are done in the set theory during its recognition. It shows the activities, relation and its description during sound recognition. Initially, input sound is given and is trained in database. Then it gives query to match input sound with database based on its extracted features and its text will be displayed.

Table 1.1: Activity table of sound recognization.

Activities	Relation	Description
Activity 1	$A1(ASI \rightarrow S)$	Input sound given to system.
Activity 2	$A2(TR \rightarrow DB_i)$	Sounds are trained accordingly and given to database.
Activity 3	$A3(Qi \rightarrow DB_i)$	Query is given to database to recognize sound
Activity 4	$A4(ASI_i \rightarrow T_i)$	Each sound will be classified and shown in text.

Relevant Mathematics:

The systems hash values are generated by the use of FFT formulae with some modifications.

FFT Formulae:

The FFT of a series values can be calculated as,

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i 2\pi k \frac{n}{N}} \qquad k = 0, \ldots, N-1.$$

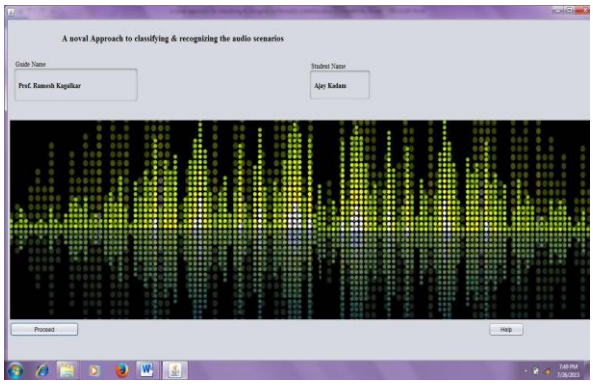
Now, the highlighted area values have replaced with 2.Hence the modified FFT formulae used to calculate FFT is as

$$X_k = \sum_{n=0}^{N-1} x_n e^{\,2} \qquad k = 0, \ldots, N-1.$$

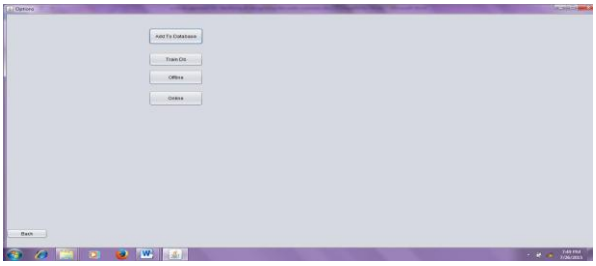
6. SCREENSHOTS OF SYSTEM

Here is an implemented system for a novel approach for classifying and recognizing the audio scenario's profile. Here onwards you will have snapshots of system so that it will be easy to user for knowing about system as well as to interact with system features.

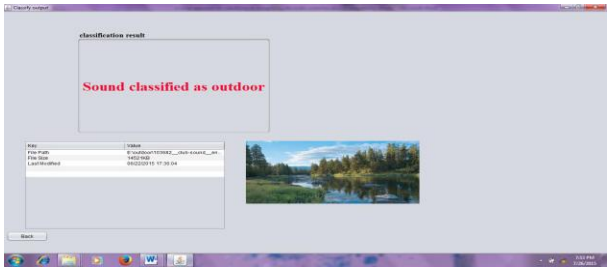
1. The figure shown below is of the initial home screen of the proposed system



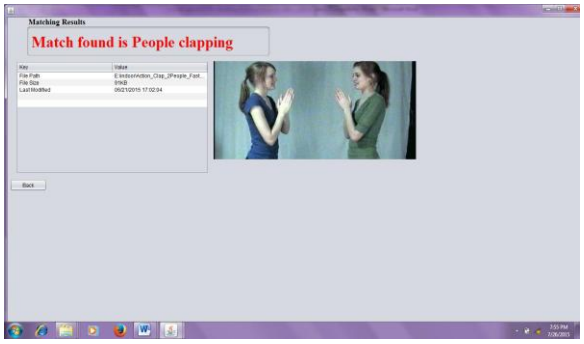
2. Now as the user proceeds further, either of these options can be selected



3. After selecting classify option.



4. After selecting exact match option.



7. CONCLUSIONS

The system will take a sound and it will do the pattern matching with saved database sound patterns if pattern gets matched then it will give the output as typical sound name and its further details with relevant image. The system gives hands on specialized apparatus to expand and to discover the classes of not the same as its sound example of a specimen space. This is conceivable by proposed methodology. In this methodology information base is made, which comprises of .wave type of distinctive movement, occasion, scenes sounds what's more, its data of every example model of different sorts. When the sound example question is given to the framework first it changes over into .wave design and concentrate highlights. Every stable document is fingerprinted a procedure in which reproducible hash tokens are separated. Both database and test sound documents are subjected to the same examination. The fingerprints from the obscure example are coordinated against a substantial arrangement of fingerprints got from the test database. The framework matches are consequently assessed for rightness of match.

The system can be extended for the classification and recognization of mixed sounds. Also database can be extended for more accuracy and variety of sounds matching. In future if sound sample is not present in the systems database then system should search it in cloud based framework where standard database of sound sample are stored for public use.

8. ACKNOWLEDGEMENT

It gives me an immense pleasure and satisfaction to present this Dissertation report on "A Novel Approach For Classifying and Recognizing The Audio Scenario's Profile", which is the result of unwavering support, expert guidance and focused direction of my guide Prof. Ramesh M. Kagalkar to whom I express my deep sense of gratitude and humble thanks, for his valuable guidance throughout the presentation work. The success of this dissertation has throughout depended upon an exact blend of hard work and unending co-operation and guidance, extended to me by the superiors at our college. Furthermore, I am indebted to our ME Co-odinator, Prof. Mrs. Roshani Raut (Ade), HOD, Prof. Ms. Arti Mohanpurkar, Principal, Dr. Uttam Kalwane whose constant encouragement and motivation inspired me to do mybest. Last but not the least, I sincerely thank to my colleagues, the staff and all others who directly or indirectly helped me and made numerous suggestions which have surely improved the quality of my work.

9. REFERENCES

[1] Namgook Cho & Eun-Kyoung Kim "Enhanced-voice activity detection using acoustic event detection & classification" in IEEE Transactions on Consumer Electronics, Vol. 57, No. 1, February 2011.

[2] Geoffroy Peeters"spectral and temporal periodicity representations of rhythm," ieee transactions on audio, speech, and language processing, vol. 19, no. 5, july 2011

[3] Jia-Min Ren, Student Member, IEEE, and Jyh-Shing Roger Jang, Member, IEEE "discovering time constrained sequential patterns for music genre classification" IEEE transactions on audio, speech, and language processing, vol. 20, no. 4, may 2012.

[4] Namgook Choo & Taeyoon Kim "Voice activation system using acoustic event detection and keyword/speaker recognition" 01/2011; DOI: 10.1109/ICCE.2011.5722550

[5] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Scream and gunshot detection and localization for audio-surveillance systems," in Proc. IEEE Conf. Adv. Video Signal Based Surveill., 2007, pp. 21–26.

[6] Ajay R. Kadam and Ramesh M. Kagalkar, "Enhanced sound detection technique", c-PGCON-2015, Fourth Post Graduate Conference for Computer Engineering students at MET's Bhujbal Knowledge City, Nashik held on 13th and 14th March 2015.

[7] Ajay R. Kadam and Ramesh Kagalkar," Audio Scenarios Detection Technique", International Journal of Computer Applications (IJCA),Volume 120 June 2015 Edition ISBN : 973-93-80887-55-4.

[8] Ajay R. Kadam and Ramesh Kagalkar," A Review Paper on Predictive Sound Recognition System" CiiT International Journal of Software Engineering and Technology, June Issue 2015 on 30/6/2015.

[9] Ajay R. Kadam and Ramesh Kagalkar," Predictive Sound Recognization System ", International Journal of Advance Research in Computer Science and Management Studies, Volume 2, Issue11 ISSN(Online):2321-7782.

[10] Kyuwoong Hwang and Soo-Young Lee, Member, IEEE "Environmental Audio Scene and Activity Recognitionthrough Mobile-based Crowdsourcing" IEEE Transactions on Consumer Electronics, Vol. 58, No. 2, May 2012.

[11] R. Radhakrishnan, A. Divakaran, and P. Smaragdis, "Audio analysis for surveillance applications," in Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust., 2005, pp. 158–161.

[12] Proc. (RT-07) Rich Transcription Meeting Recognition Evaluation Plan, [Online]. Available: http://www.nist.gov/speech/tests/rt/rt2007

[13] J. Tchorz and B. Kollmeier, "A model of auditory perception as front end for automatic speech recognition," J. Acoust. Soc. Amer., vol. 106, no. 4, pp. 2040–2050, 1999.

[14] T. Jaakkola and D. Haussler, "Exploiting generative models in discriminative classifiers," in Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 1998, vol. 11, pp. 487–493.

[15] V. Wan and S. Renals, "Speaker verification using sequence discriminant support vector machines," IEEE Trans. Speech Audio Process., vol. 13, no. 2, pp. 203–210, Mar. 2005.

[16] T. Jebara and R. Kondor, "Bhattacharyya and expected likelihood kernels," Lecture Notes in Computer Science, vol. 2777, pp. 57–71, 2003.

[17] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMM pervectors for speaker verification," IEEE Signal Process. Lett., vol. 13, no. 5,pp. 308–311, May 2006.

[18] Abeer Alwan, Steven Lulich, Harish Ariskere "The role of subglottal resonances in speech processing algorithms" The Journal of the Acoustical Society of America (Impact Factor:1.56). 04/2015;137(4):2327-2327.

[19] Fred Richardson, Douglas Reynolds, Najim Dehak "Deep Neural Network Approaches to Speaker and Language Recognition" IEEE Signal Processing Letters (Impact Factor: 1.64). 10/2015; 22(10):1-1

[20] Jens Kreitewolf, Angela D Friederici, Katharina von Kriegstein "Hemispheric Lateralization of Linguistic Prosody Recognition in Comparison to Speech and Speaker Recognition."NeuroImage (Impact Factor: 6.13). 07/2014;102DOI: 10.1016/j.neuroimage.2014.07.0

[21] Kun Han, Yuxuan Wang, DeLiang Wang, William S. Woods, Ivo Merks,Tao Zhang "Learning Spectral Mapping for Speech Dereverberation and Denoising" Audio, Speech, and Language Processing, IEEE/ACM Transactions on 06/2015; 23(6):982-992. DOI: 10.1109/TASLP.2015.2416653

[22] Mikolaj Kundegorski, Philip J.B. Jackson, Bartosz Ziółko" Two-Microphone Dereverberation for Automatic Speech Recognition of Polish" Archives of Acoustics 01/2015; 39(3). DOI: 10.2478/aoa-2014-0045

[23] Abeer Alwan, Steven Lulich, Harish Ariskere" The role of subglottal resonances in speech processing algorithms" The Journal of the Acoustical Society of America (Impact Factor: 1.56). 04/2015; 137(4):2327-2327. DOI: 10.1121/1.4920497

10. AUTHORS PROFILE

Mr. Ajay Kadam received Bachelor of Engineering degree in Computer Science & Engineering in …….and now pursuing Post Graduation (M.E.) in department of Computer Engineering from Dr. D.Y.Patil School of Engineering and Technology in current academic year 2014-15. He I currently working on audio sound reorganization.

Ramesh. M. Kagalkar was born on Jun 1st, 1979 in Karnataka, India and presently working as Assistant. Professor, Department of Computer Engineering, Dr.D.Y.Patil School Of Engineering and Technology, Charoli, B.K.Via –Lohegaon, Pune, Maharashtra, India. He has 13.8 years of teaching experience at various institutions. He is a Research Scholar in Visveswaraiah Technological University, Belgaum, He had obtained M.Tech (CSE) Degree in 2005 from VTU Belgaum and He received BE (CSE) Degree in 2001 from Gulbarga University, Gulbarga. He is the author of text book Advance Computer Architecture, which cover the syllabus of final year computer science and engg department, Visveswaraiah Technological University, Belgaum. He is waiting for submission of two research articles for patent right. He has published more than 18 research papers in International Journals and presented few of there in international conferences. His main research interest include Image processing, Gesture reorganization, speech processing, voice to sign language and CBIR. Under his guidance two ME students awarded degree in SPPU, Pune, two more students at the edge of completion their ME final dissertation reports and 5 final year M.E students are started a research work on different area and they have publish their research papers on International Journals and International conference. 2 ME Ist year students are started a research work on different area.