UNIQUE KEY IDENTIFICATION OF AUDIO

Avantika Mussady, Parichita Ray, Lhaden Wongmu Sherpa, Sauvik Bal

Abstract— Objective of this paper is to understand the need as well as the process how the identification of a particular song is done through different approach keeping in the mind the concept of uniqueness. Our idea focuses on the uniqueness of the song as every song comprise different music along with lyrics. Recognizing a particular audio by understanding its pitch and other details by using its audio sample is known as audio fingerprint. The approach of the proposed is to identify a particular song and the method is carried out by dividing it into two parts i.e. in vocal and instrumental. There is no comparison between these two in terms of peak values and both are different entities which comprise different qualities. The values of both the parts are considered and amalgamated in such a way that a new value is obtainted that is known as prime value. The unique key is known as a prime value which is later on used to identify that particular song.

Index Terms— Audio Identification, Spectrogram, peak value, prime value, Vocal part and Instrumental part, Key-generation.

—————————— ◆ ——————————

1 INTRODUCTION

In year 1893 Sir Francis Galton was responsible to prove that no two fingerprints of human are alike and so this was tak-en further as a unique feature to identify the human, not on-ly fingerprints but also the iris, ears, hair etc. so if the same idea have been taken in different approach considering in different field then the audio identification/audio fingerprint is taken into account.

Audio identification concept deals with the process of iden-tifying a particular song by understanding its nature, it means the type of the song is encoungtered. Type of the song is calcu-lated by understanding the two major part i.e. the vocal part and the acoustic part of the song.

There's always a common question that arise in everyone's mind that is the reason behind using a technology i.e. its need. Today generation needs advance technology as it's been known that human work done can never be compared with any technology in today's era. Technology has reduced the work done of human and due to the habit of adapting new technology in life makes them easier to live. This ap-proach of audio identification gives a platform to the users to identify the song by using unique approach.

Audio identification has played very important role in un-derstanding and recognizing a particular song. This process talks about the representation of the information (mu-sic/sound) which is encapsulated depending on the type (uniqueness).

Role is to capture the signature/mark of a particular song by identifying the vocal and the background music. Examples of the audio fingerprint/identification technology are like Shazam and AcoustID which uses Spectrogram to form a time-frequency representation of audio. Spectrogram helps to ana-lyze the song by finding its property as a function typ-ically its wavelength, frequency or energy.

Using the concept of spectrogram is a good starting point as through it robust audio fingerprint canbe generated. It ana-lyzes the song and indentifies the peak but the plan is to find the primary points of the region i.e. the vocal and instrument peak value. This notion is carried out by understanding the identification process through spectrogram.

2 LITERATURE SURVEY

They have focused on the watermarking process and have discussed about the different techniques of watermarking pro-cess dealing with the Time Domain Technique, Spread Spec-trum Technique, Frequency Domain Technique, Echo Hiding Technique, and Patch Work Technique. Their concept deals with the protection of the data of the right owner with the help of watermark [1]. This paper talks about the concept of Spike-gram. Spikegram is a new method to represent audio signal used to construct the robustness of the sound focused to re-ducing the strength of the decoder. Their proposed method is robust against the variety of signal processing transform. Their idea is robust against unified speech and audio code [2]. First-ly they have optimized the issue of robust and payload which is the challenging. The embedding quantization function which they have designed to introduce minimal changes in the signal while embedding provides an opportunity to hide more watermarking data. The features have been properly utilized to achieve robustness and imperceptibility at an acceptable level. Their proposed algorithm has good performance giving good imperceptibility and robustness [3]. This paper presents a wireless multimedia simulator (WMS), which uses a com-pact graphical user interface to present the real-time packet delay with the playback of streaming media over a wireless channel based on classic radio channel models and IEEE 802.11 medium access control. This WMS helps to demonstrate the visual and auditory effects of fading errors, packet delay and loss. Statistical difference-ofmeans tests show the video with slower motion and fewer colors is likely to offer better delay tolerance, and audio is less sensitive to bit errors while video is more resistant to network congestion [4]. To achieve optimized receiver audio quality for transmissions using sin-gle or multiple paths, they have propose a self-adaptive joint Error and Rate Control framework based on packet loss pre-

————————————————

- *Avantika Mussady, Parichita Ray, Lhaden Wongmu Sherpa are currently pursuing Bachelor's degree in Computer Science & Engineering in Universit of Engineering & Management ,Jaipur, Rajasthan, India. E-mail: m.aavie.26@mail.com, parichitaray83@mail.com, lhadensherpa01@gmail.com*
- *Sauvik Bal currently working as Assistant Professor in the department of Computer Science & Engineering in Universit of Engineering & Management, Jaipur, Rajasthan, India. E-mail: sauvikbal@mail.com*

diction and on-line quality assessment. The Error Control chooses proactive FEC to preserve quality with optimal bandwidth, using a Markov Decision Process (MDP) and a stochastic inventory control, a new approach for multimedia error recovery. In their approach they use The Rate Control which uses a quality optimization model to determine the optimal dispersion over single or multiple paths. They have present results using simulation and Internet experiments to show the superiority of our mechanism over other similar techniques [5]. They have proposed a novel, reliable and preemptive dynamic congestion control mechanism DCCM based on RTP for streaming media. They have adopted the decoupling of persistence congestion from transient congestion. Their mechanism is tailored to use packet inter-arrival jitter for detecting transient congestion and estimated packet loss rate probability for detecting and persistence congestion. Their primary goal is to detect congestion before it leads to packet losses while trying to maintain fairness with other competing flows. They have been evaluated Dynamic congestion control mechanism DCCM and tested through extensive simulation and concludes that bandwidth is almost evenly shared between DCCM traffic and TCP flow, while DCCM tries to maintain smoothness by effective utilization of bandwidth [6]. Their main aim was to achieve the error robustness the signal is divided into the base and the enhancement layer when the network is approaching the congestion. The base layer is perceptually encoded using a time-varying pre- and postfilter, and this layer is transported using a high priority traffic class in Differentiated Services (DiffServ) network. The enhancement layer is the difference between the original signal and base layer, and is transmitted using a regular Best Effort traffic class. In their experiments the system delay is just 256 samples, and the layering only introduces moderate amounts of redundancy, while improving the error resilience significantly [7]. Their mechanism works only if time sensitive data stream is the minority of the network traffic and if the network is not very congested. The methodology of dropping low priority data will not scale when time sensitive data stream uses high percentage of network bandwidth. This is because of the bandwidth which is required by video/audio applications can vary in very wide range when real-time data becomes majority network traffic, that is, television (TV), telephone, visual telephone, videoconferencing, gaming, and other video/audio based applications are all deployed on Internet [8]. Their paper includes the robustness against acoustic path transmission is only evaluated. It introduces a simple audio watermarking algorithm. And then the watermark embedding and detection process will be takesplace. Then the results will gives the high fidelity robust audio watermarking algorithms. Then these algorithms will gives the high robustness and also achieves high capacity against the common signal manipulation [9]. In this paper, they have propose a digital watermarking for voice signals recorded by a digital voice recorder especially in conferences. They have discussed about some requirements for detecting falsifications in the voice signal, because the conference record has the probability that a conclusion is changed for only a falsification in a little time interval. They have shown a method of locating an altered place in the voice signal and a new idea applying a white Gaussian noise for guarantee of the valid voice signal without falsification. One characteristic of their proposal is an agitation depending on a peak spectrum for making it difficult to illegally change the embedded watermark [10]. This invention related to a movie anti-piracy system with sensing and reporting features. A movie projector projects a light image to reflect off of a screen, while an infrared energy projection source generates an infrared energy pattern which is reflected off the screen in the same field of view as the light images reflecting from the screen. Infrared energy sensors are positioned in the seating area of the theater. The infrared sensors provide sensed signals respectively to a sensor response controller. The controller is responsive to the sensor signals to provide an infrared status signal output to a control center. This device in only confined to theatre our approach is which they have planned supports the prevention of piracy even while accessing a media file on our mobile devices, laptops [11]. In potato system they would have their own digital signature as a key encryption. The customer pays for the re-distribution permission. Once he gets the license he can not only distribute the song, but can also earn money for every song he sells along with the producer. The policy rights are given indirectly into the hands of the consumers who have no interest in preventing its exploitation [12]. Fingerprinting hides user-related information in a distribution file to distinguish users. The robustness of the echo hiding method can be improved by increasing the amplitudes of the echo signals, although such would degrade the audio quality [13]. The fragile watermark is embedded in such a way that it is destroyed, if the robust watermark is tried to be removed. If the fragile watermark is destroyed, the user is unable to know where each segment is. If the segments are short enough it is very hard to put them in the correct order. Only a 50% indicates that the two audio clips (original clip and the watermarked clip) cannot be discriminated [14].

3 PROPOSED SYSTEM

The goal of audio identification is to develop such a platform where identifying any audio and storing it differently with the help of unique ID. Through that unique ID particular song can be access directly with no collision in the backend memory.

4 ALGORITHM

4.1 Statement form:
Step1.
Select a desired song.

Step2.
The desired song is split into two parts i.e. the **Vocal part** and the **Instrumental part**.

Step3.
Dividing each part of the song into segments which later on is used to derive a value, with the help of those audio segments a graph is plot for which Spectrogram is used.

Step4.
With the help of the graph, the **RMS** (root mean square) value

International Journal of Scientific & Engineering Research Volume 10, Issue 3, March-2019
ISSN 2229-5518

of the Vocal part as well as of the Instrument part is obtained.

Step5.
Extract the unique value of each part and take the last eight digits through it.

Step6.
Club the value (last eight digits) taken from the each part to obtain a unique value i.e. the unique key. This unique key is later used to identify the song.

Note:
Following the same process any song can be identified by the unique key introduced at the end of the process.

4.2 SYSTEM ARCHITECTURE

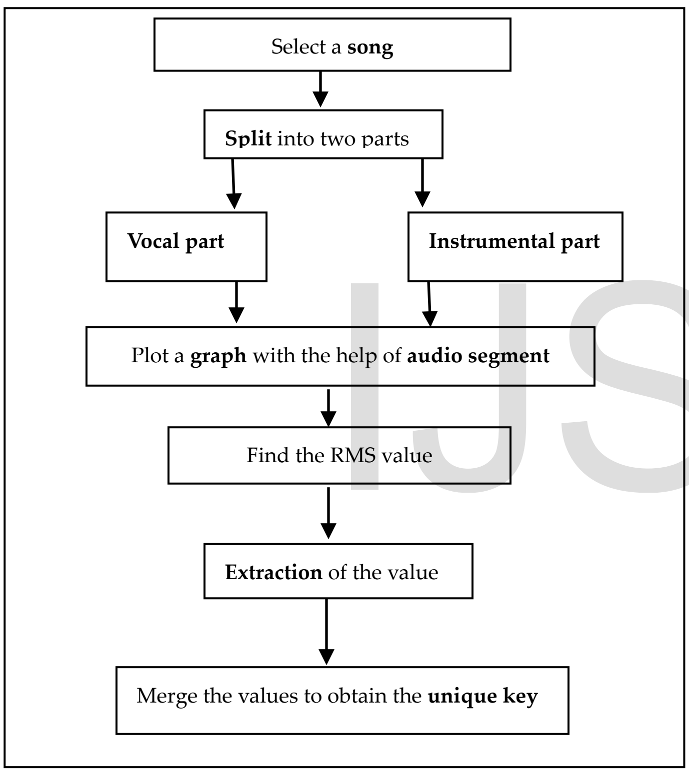

Fig.1. **Block Diagram**

5 EXPLANATION WITH GRAPH

The method of SPECTROGRAM which is used for the visual represent of the spectrums of the signals which counts frequencies which varies according to time and the intensity increases with the density.

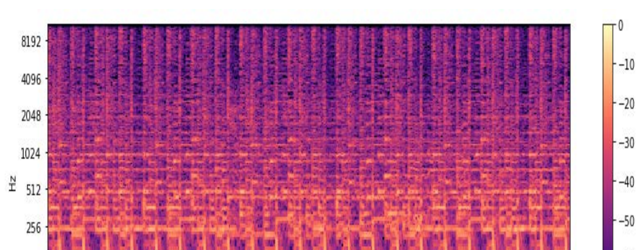

Fig.2. Spectrogram

With the help of the spectrogram the astronomical spectrum which is divided into two parts Vocal and Instrumental part is recorded.

Song: "Jaane Ja"

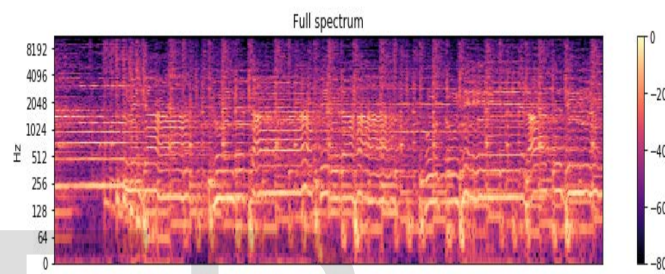

Fig.3. Spectrogram of "Jaane Ja"

Figure.3. is the representation of a song with the help of spectrogram (frequency Vs time graph). In this example, louder events like speech/lyrics are indicated by bright color (yellow/bright orange) and quieter events (instrument/noise floor) are indicated by darker colors (dark orange/blue/black).
After dividing the audio i.e. the song into vocal and instrumental parts, the peak value i.e. the RMS value is obtained.

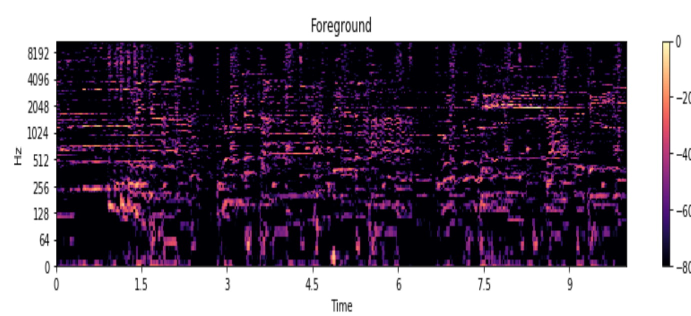

Fig.4. Vocal Part of "Jaane Ja" in the form of Spectrogram

Figure.4. is the representation of vocal part named as Foreground (to understand easily). When the term vocal is mention then it is consider as the speech/lyrics part but not the instrumental part.

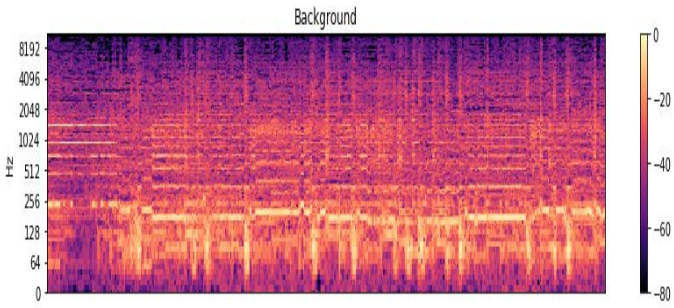

Fig.5. Instrumental Part of "Jaane Ja" in the form of Spectrogram

Figure.5. Representation of instrumental part named as Background (to understand easily). In the graph of background only considered the instrumental part not the vocal.

After the audio segmentation, find the peak value of Foreground and Background. After finding the values of both the parts, take the last 8 digits of the value and finally club those values together to get one final as a unique key.
Consider it as unique key because with the help of it this key, our approach toward audio identification is carried out successfully.

6 ANALYSIS

Song: Jaane Jaan [15-22]
Film: Jawani Diwani
Language: Hindi

Different versions of same song is taken for experiment but that are sung by different singers along with different background music. Based upon these 8 songs the proposed approach is carried out.

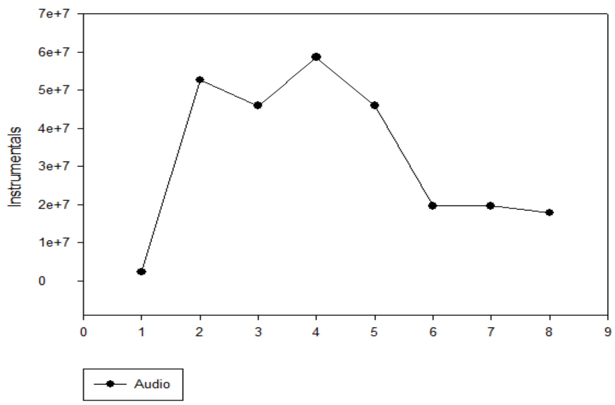

Fig.6. Instrumental part of 8 versions

Figure.6. it represents the different peak values of a song in 8 different forms, where X-axis and Y-axis are frequency and time respectively. Through this peak values the concept is understand about the song which is selected comprise same lyrics/speech but due to difference in the values of instrumental part there is no similarity.

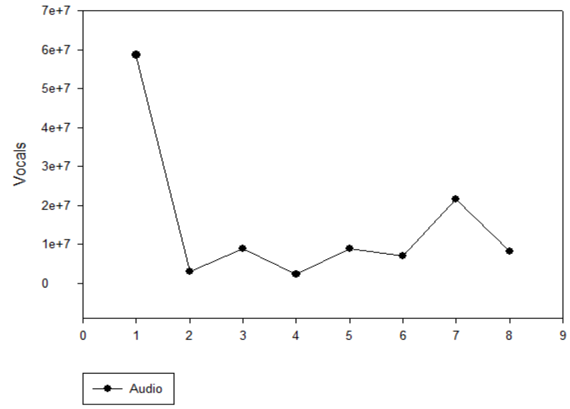

Fig.7. Vocal part of 8 versions

The rule is same like Instrumental graph. **Fig.7.** represents the different peak values of a song in 8 different forms. This 1 song in 8 different forms results in 8 different peak values. Difference lies because of different vocal peak.

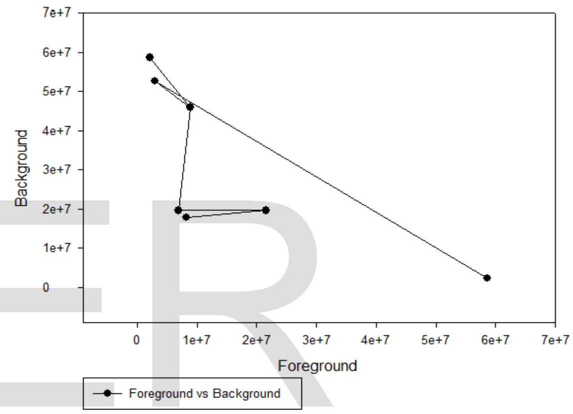

Fig.9. Background Vs Foreground of 8 versions

Fig.9. shows the graph of Background (Instrumental) Vs Foreground (Vocal). Through this graph the concept is understand that after the merge of 8 peak values of Vocal as well as Instrumental part, the final 8 peak values are the unique 8 keys of a single song but in 8 different forms. These 8 unique keys are used to define 8 different forms of a song.

Hence, different unique keys are used to identify different forms of a song; in the same way for every song the result comes as different prime values.

7 FUTURE WORKS

Based on this approach of Audio Identification, the future prospectus can be expected in various ways such as enhancing the Robustness, a property to use machine learning to take this audio identification to the next level of technology and it can also yield efficient analytical models in future.

8 ACKNOWLEDGMENT

We would like to acknowledge and extend our heartfelt gratitude with our families, faculties and friends for their immense support in successful evaluation of the results.

10 CONCLUSION

The experiment yields us the unique key value of any given audio. From the result, the value for an audio is determined based upon the values generated by splitting the instrumental segment and the vocal segment, evaluating them individually and concatenating the two values obtained. It is based upon the way in which humans identify any given audio. To test the value for uniqueness, multiple versions of songs is considered with similar vocals and instrumental values, thereby receiving unique values for each of such audio files. This unique value can be used as a prerequisite for various evaluations which is to be performed on a given audio file.

11 REFERENCES

[1] Mr. R. D. Shelke, Dr. Milind U. Nemade, “Audio Watermarking Techniques for Copyright Protection”

[2] Yousof Erfani, Ramin Pichevar, Jean Rouat, “Audio watermarking using spikegram and a two-dictionary approach”

[3] Arashdeep Kaur, Malay Kishore Dutta K. Elissa, “A Hybrid approach for Optimizing Transparency, Robustness and Capacity of an Audio Watermarking Algorithm”

[4] Xiaojing Li, Kai Dong, Wei Song, and Bradford G. Nickerson “QoE Evaluation of Multimedia Transmission over Wireless Network”

[5] Lopamudra Roychoudhuri , Ehab S. Al-Shaer, “Autonomic QoS Optimization of Real-time Internet, Audio using Loss Prediction and Stochastic Control”

[6] Shafqaat Ahmad, N .D Gohar, Aatif kamal, “A Dynamic Congestion Control Mechanism for Real Time Streams over RTP”

[7] Erik Hellerud and U. Peter Svensson, ” Robust Transmission of Lossless Audio with Low Delay over IP Networks”, Centre for Quantifiable Quality of Service in Communication Systems, Norwegian University of Science and Technology Trondheim, Norway,
2007 IEEE International Symposium on Signal Processing and Information Technology

[8] [1]Guojun Jin, [2]Haina Tang, “Control Transmission Pace at IP Layer to Avoid Packet Drop”, [1]Lawrence Berkeley National Laboratory 1 Cyclotron Road, Berkeley, CA 94720, [2] Chinese Science and Technology Network Center, Beijing, China

[9] [1]R.Subhashini, [2] Dr.K.Boopathi Bagan, “Robust audio watermarking for monitoring and information embedding”, [1] M.E Student, Department of Electronics Engineering, Madras Institute of Technology, [2] M.E, Ph.D, Professor, Department of Electronics Engineering, Madras Institute of Technology, 2017 4th International Conference on Signal Processing, Communications and Networking (ICSCN -2017), March 16 – 18, 2017, Chennai, INDIA

[10] [1]Tomoki YOSHIDA , [2]Takeru MIYAZAKI , [3]Shunsuke ARAKI , [4]Satoshi UEHARA and [5]Yasuyuki NOGAMI, “Detecting Falsification to MP3 Audio Signals for Conference Record Using Digital Watermarking”, The University of Kitakyushu, Fukuoka, Japan, Kyushu Institute of Technology, Japan, Okayama University, Japan, 2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)

[11] David H. Sitrick, James A. Fancher, “Anti-piracy protection system and methodology. https://patents.google.com/patent/US6771349

[12] GauravYadav, SowmyaKamath S , Gaurav Prasad; “Secure Music Sharing with Improved Potato System Using Cryptographic Techniques”; Signal Processing, Informatics, Communication and Energy Systems (SPICES), 2015 IEEE International Conference onSignal Processing, Informatics, Communication and Energy Systems (SPICES).

[13] Che-Jen Hsieh, Jung-Shian Li and Cheng-Fu Hung, “A Robust Audio Fingerprinting Scheme for MP3 Copyright”; Intelligent Information Hiding and Multimedia Signal Processing, 2007. IIHMSP 2007.Third International Conference onIntelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2007).

[14] Mikko L6ytynoja, NedeljkoCvejic, EevaLahetkangas, TapioSeppanen, “Audio Encryption Using Fragile Watermarking”; Information, Communications and Signal Processing, 2005 Fifth International Conference onInformation Communications & Signal Processing.

[15] https://www.youtube.com/watch?v=MKnlkHt5LC4&feature=youtu.be

[16] https://www.youtube.com/watch?v=uwPvI6pETRs&feature=youtu.be

[17] https://www.youtube.com/watch?v=r2oPV_j22LE&feature=youtu.be

[18] https://www.youtube.com/watch?v=7cXz6akcHpE&feature=youtu.be

[19] https://www.youtube.com/watch?v=EcyF7_vaPGE&feature=youtu.be

[20] https://www.youtube.com/watch?v=eLIHjjchO0M&feature=youtu.be

[21] https://www.youtube.com/watch?v=cIrx5nmtJK8&feature=youtu.be

[22] https://www.youtube.com/watch?v=eYKQvutPWpk&feature=youtu.be