ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 5 May 2024

Detecting Forgeries and Protecting Authenticity on Digital Audio Authentication

[1] M. Sangeetha, [2] R. Lakshmi Narasimha, [3] P. Sriram Chowdary, [4] R. Pranav, [5] L. Kailash

[1] Assistant Professor, Department of Computer Science and engineering, Kalasalingam academy of research and education Virudhunagar, Tamil Nadu, India

[2] [3] [4] [5] Department of Computer Science and engineering Kalasalingam academy of research and education Virudhunagar, Tamil Nadu, India

Corresponding Author Email: [1] sangeetha.m@klu.ac.in, [2] r.lakshminarasimha2003@gmail.com, [3] srirampendyala123@gmail.com, [4] 9920004111@klu.ac.in, [5] Kailashcs1004@gmail.com

Abstract*— The spread of more complex fake methods has increased the need for strong digital audio identification systems, which can be used in a wide range of real-life situations to both stop fraud and look into it. These kinds of programs are necessary to stop problems like fake proof, copyright violations, and people getting into your data without permission. This study talks about a new kind of automatic identification system that can tell the difference between fake and real sounds. It is based on three basic psychoacoustic principles of how humans perceive sound. The main idea behind our suggested system is to make it work like the human hearing system, which can tell the difference between real and fake audio records. Additionally, our system has the amazing ability to tell the difference between sound caught in different places using the same microphone, which makes it much more useful. To do this, we use psychoacoustic principles to pull out traits from the noise data that are useful. Gaussian mixture models are used to analyze these traits. This makes it easier for computers to make decisions automatically, which authenticates the audio material and sorts the recording surroundings into groups. Our automatic identification system is the best at stopping audio fraud and manipulation because it uses advanced statistical models and principles of how people hear sounds. It can be used in many areas, from forensics to property protection and data security, making it a flexible answer to many of the important problems that the digital age has brought up.

Index Terms*— Digital audio authentication, Psychoacoustic principles, detection of Forgery, Audio classification, Gaussian mixture model.

I. INTRODUCTION

Today, the huge popularity of smart devices like smartphones and the amazing progress made in many other technologies, especially mobile and wireless networks, have completely changed the way we live. Digital video has become an important part of our daily lives and an essential thread in the structure of our society because of this change. Multimedia formats like pictures, sounds, and videos are used for sharing information, keeping records, and making art. However, as the digital world changes, so do the problems that come with making sure that video information is correct and reliable. The effects on court cases are one of the most important effects of this digital multimedia boom. Video and audio files are now commonly accepted as proof in courts, making them a key factor in shaping choices that have far-reaching effects. There are, however, two sides to this use of video in the court scene. Multimedia material can be easily made and shared using technology that can also be used for bad things. This is why there is a worrying amount of fake and unauthentic multimedia. These fakes can be used to change people's minds, trick a lot of people, or even stop the legal system from working. A lot of progress has been made in verifying pictures and videos, but digital voice verification is still very new and needs to be worked on right away. Checking and analyzing digital audio records to make sure they are real and find any kind of fraud is a very important task. The justice system, the entertainment industry, and many other fields are all affected by this problem in big ways. Digital audio analysis and identification have become very important tools for dealing with this problem. They make it possible to listen to audio records carefully, spot changes or fakes, and finally make sure the material is real. Such technologies and methods are useful for many things, from making sure that audio evidence is real in court to making sure that audio recordings are reliable in journalism and from protecting the artistic integrity of music and sound production to making sure that voice-based communication is safe. Nowadays, it's easy for the lines between reality and fiction to become blurry. Because of this, verifying digital audio and video is a constantly changing field that needs new ideas and close attention.

Now more than ever, telling the difference between the real and the fake, between the real and the fake, is essential for keeping our society and our multimedia-rich world honest. This introduction sets the stage for a more in-depth look at the problems, progress, and possible uses of digital audio identification and investigations.

II. LITERATURE REVIEW

A Managerial Approach to Industrial Safety and Accident Prevention a Managerial Approach to

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 5 May 2024

Industrial Safety and Accident Prevention February 2013, International Journal of Science, Engineering, and Technology Research.

In order to stay ahead of the competition, different industries are focused on different parts of their production methods. To make changes, it is important to figure out what is wrong with the output system. This essay looks at a small part of a company's production system to find problems with the safety system, make the needed changes, and give companies tips on how to reach their goals and avoid accidents. We are more aware of the need for workplace safety measures because of the rise in accidents involving workers.

H Takata, H. Nakamura, T Hachino "On prediction of electric power damage by typhoons in each district in Kagoshima Prefecture via LRM and NN", SICE Annual Conference, 2004.

Kagoshima Province has been hit by typhoons many times, causing natural disasters. They sometimes cut off the power, which does a lot of damage to power equipment. To make sure that the power is quickly restored, it is important to figure out exactly how much damage was done by typhoons in each place. A two-stage classifier is used in this study to look at damage forecast in each area of Kagoshima Prefecture. A linear regression model (LRM) is used in the first step, and neural networks (NN) are used in the second. Predictions for typhoon weather can be used to get an idea of how many distribution poles and lines will be harmed. By using real data, the method is guaranteed to work.

Ramli Adnan. Abd Manan Samad, Zainazlan Md Zain, Fazlina Ahmat Ruslan "5 hours flood prediction modeling using improved NNARX structure: case study Kuala Lumpur", IEEE 4th International Conference on System Engineering and Technology, 2014.

Flooding is a type of natural disaster that is getting more dangerous around the world. Flooding can cause disasters that hurt people and damage property. An exact estimate of the flood water level is very important for flood modeling because it gives people living near the flood area time to leave. However, artificial neural networks (ANNs) are often used to deal with nonlinear problems, so they are a good choice for modeling because flood water levels change in a very nonlinear way.

Long Wang, Xiaoqing Wang, Aixia Dou, Dongliang Wang "Study on construction seismic damage loss assessment using RS and GIS" International Symposium on Electromagnetic compatibility, 2014.

This study shows a quick way to evaluate an earthquake situation. To get information about damage from remote sensing photos, the approach uses two different ways: one is based on a damage index, and the other is on picture classification. As usual, visual analysis is used in the damage index setting. You can get damage scores from experts, then ground strength data, and loss estimate factors from the experienced risk matrix.

III. METHODOLOGY

Visual waveform analysis and spectrogram study have been used by humans for a long time to find problems or changes in audio files as part of audio identification. These methods aren't perfect because people can make mistakes and might miss small signs of cheating. Recent improvements to tools used to change sounds make it even harder to find problems. Visual and audio inspections often don't show any problems, even when tampering has happened, making it hard to be sure of its authenticity. In the end, these human methods aren't very good at what they do and might not be able to consistently tell the difference between original and changed music.

Drawbacks:

- Still, one of the main problems with these approaches is that they depend on human opinion, which means that mistakes can't be ignored.
- For that reason, it is very hard to spot those oddities. To give you an example, looking at the spectrum and spectrogram of the changed audio doesn't show any problems, and hearing is also fine.

The suggested method uses psychoacoustic ideas to sort sound from different places that a single microphone picks up. It can tell the difference between settings on its own by using Gaussian mixture models on computed data. It's also important that it can identify unknown people regardless of the audio content, which means it doesn't depend on the content. In order to rate performance, human-indistinguishable synthetic noise is made in various settings and evaluated subjectively by three people. The results show that the system is very strong and works well, with a classification accuracy of 99.2% ± 2.6. It also achieved a perfect 100% accuracy in situations like audio authentication that are text-dependent and text-independent.

Benefits:

- The suggested method correctly sorts things into groups 99.2% of the time, giving an accuracy of 2.6.
- In addition, the suggested method works 100% of the time in other situations, like text-dependent and text-independent voice verification.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 5 May 2024

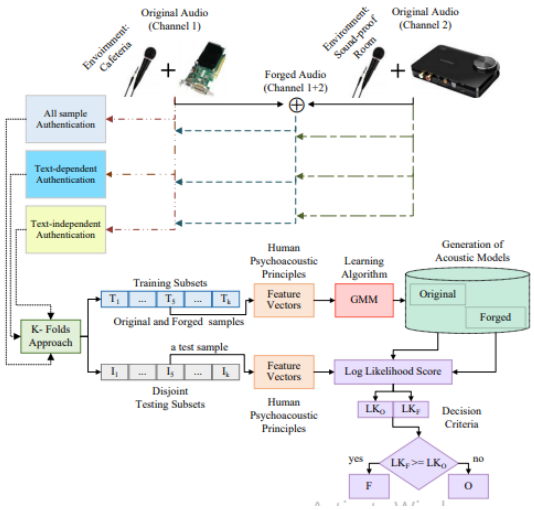

Fig. 1. Proposed Architecture

Modules:

1. **Dataset Upload:**

You can use a library like pandas to upload your audio dataset in a format like CSV or any other suitable format.

2. **Dataset Preprocessing:**

To load and prepare the audio files, use tools such as librosa or pydub. This can include things like resampling, getting rid of noise, and extracting features.

3. **Feature Extraction:**

Use the right sound qualities. MFCC (Mel-Frequency Cepstral Coefficients), Chroma traits, and other things make up typical audio qualities. In this case, you can use librosa.

4. **Load & Build Gaussian Mixture Model (GMM):**

A scikit-learn library can be used to load and build a Gaussian Mixture Model. You can use GMM to describe how the traits you take out of your audio data are spread out.

5. **Audio Authentication:**

You can use your GMM model for voice confirmation once it has been taught. To make sure the sample is real, you can compare its retrieved properties to the model during the validation process. It might not be verified if the sample is very different from the model.

6. **Authentication Logic:**

Based on the GMM model's output and the authentication logic you've designed, verify the authenticity of the audio. This can involve setting a threshold on the likelihood score or using a machine learning Classifier.

IV. IMPLEMENTATION

Algorithms

Gaussian Mixture Model:

For simplicity's sake, let's say there are K groups. This means that values of \mu and \Sigma are given for every k. If there was only one distribution, the maximum-likelihood method would have been used to figure them out. That being said, the probability density is a straight line function of the densities of all K of these distributions, and there are K of them, to be exact.

$$p(X) = \sum_{k=1}^{K} \pi_k G(X|\mu_k, \Sigma_k)$$

Expectation-Maximization (EM) Algorithm:

If you don't have enough data, data that is missing points, or secret factors, the Expectation-Maximization (EM) approach lets you find the most likely values for model parameters in a step-by-step way. To predict a new set of data, EM picks random numbers for the data points that are missing. The new values are used to predict a better start date until the values are fixed. This is done by adding up the missing points.

In the Expectation-Maximization (EM) method, the estimate step (E-step) and the maximization step (M-step) are the two most important repeated steps. They change the model parameters until the model converges.

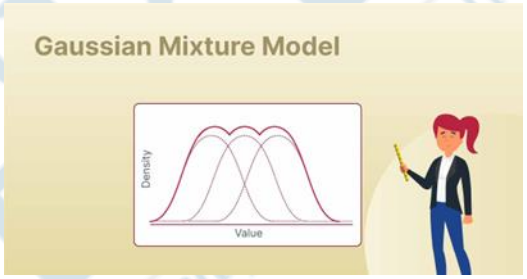

Fig. 2. Block Diagram

In a Gaussian mixture model, each cluster is connected to a multimodal Gaussian distribution. The mixture model is what these distributions add up to when they are added together. The Gaussian distributions show how the data in each cluster is spread out, and the weights show how likely it is that a certain data point is in that cluster.

You can use the expectation-maximization (EM) method to guess the Gaussian mixture model's factors and parameters. To reach equilibrium, this means moving between guessing the parameters of the Gaussian distributions and the weights of the mixture model.

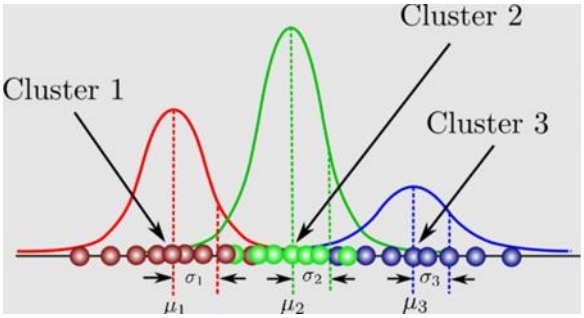

Fig. 3. Clusters of GMM

GMM is used in many different contexts, including image segmentation, grouping, and density estimation. The probability density function of a set of data points may be estimated using GMM for density estimation. When clustering, data points from the same Gaussian distribution can be grouped together using GMM. Additionally, GMM

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 5 May 2024

may be used to divide a picture into various parts for image segmentation.

Numerous application cases, such as identifying client groups, spotting fraudulent behavior, and group photos, may be solved with Gaussian mixture models. The Gaussian mixture model can detect clusters in the data in all of these situations, even if they might not be immediately apparent. Gaussian mixture models are therefore an effective tool for data analysis and must be taken into account for every clustering job.

V. EXPERMENTAL RESULTS

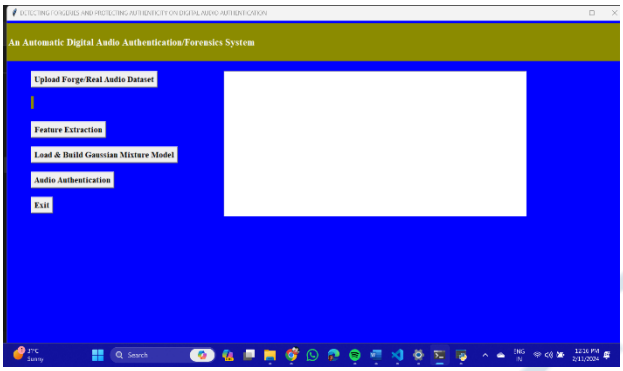

Fig 4 Home Page

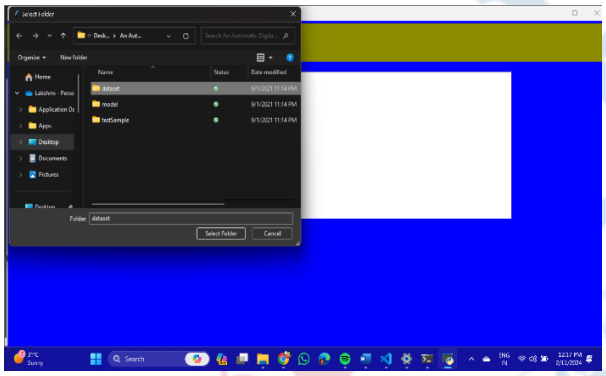

Fig 5 Upload Audio Dataset

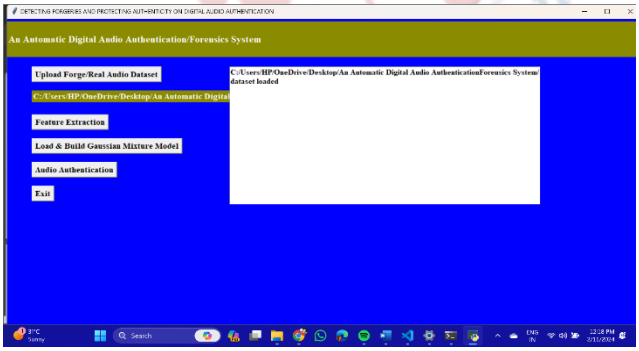

Fig 6 Audio Dataset Loaded

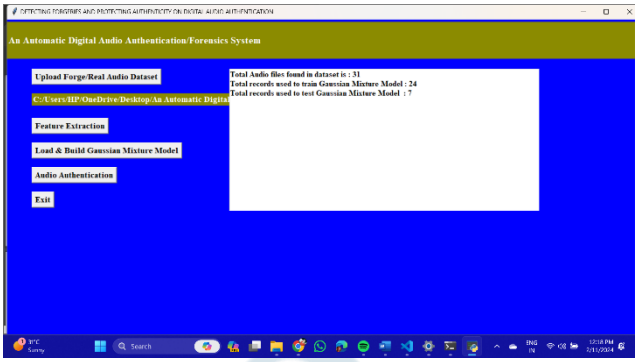

Fig 7 Split dataset into Train & Test

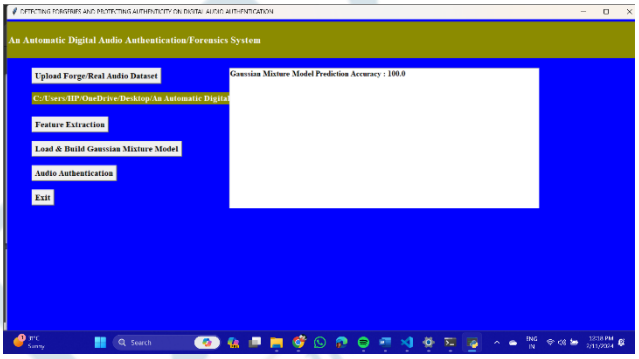

Fig 8 Accuracy of GMM

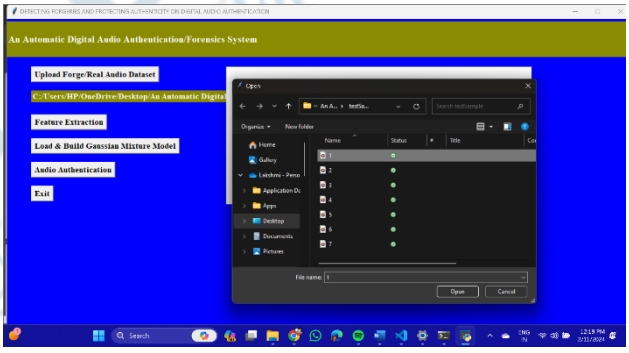

Fig 9 Upload Audio Dataset

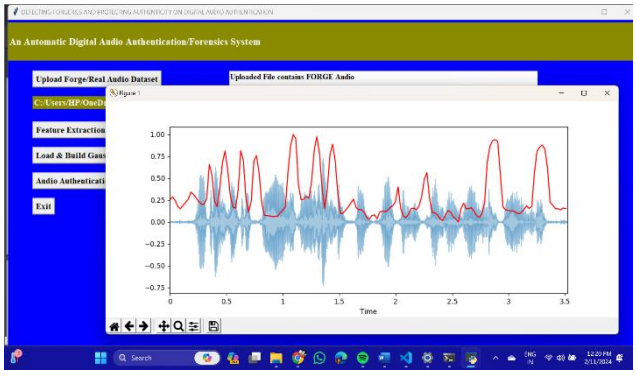

Fig 10 Output Graph of Model

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 5 May 2024

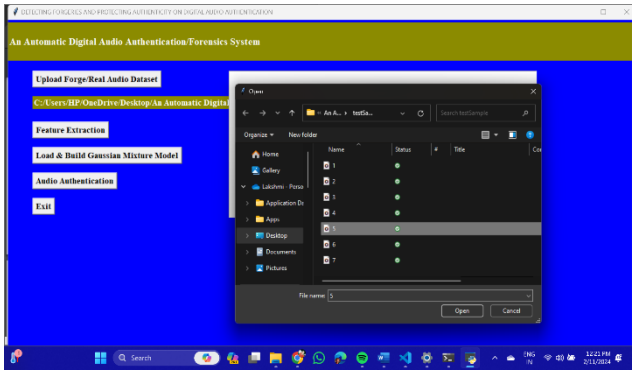

Fig 11 Upload New Audio Dataset

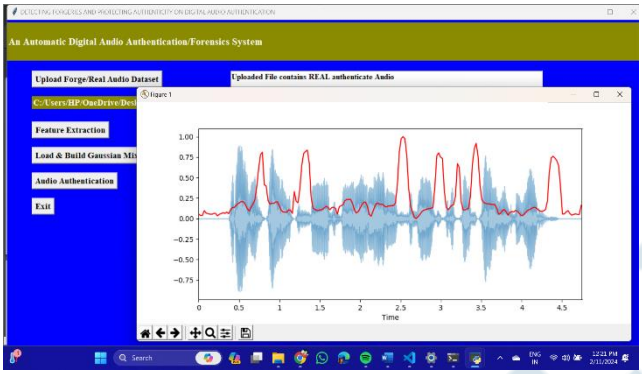

Fig 12 Output of New Dataset

VI. CONCLUSION

In the end, this study presents a cutting edge automatic audio identification system that uses three basic human psychoacoustic principles to tell the difference between real and fake audio. A very good system gets great results by taking feature vectors from both real and fake sounds and then using Gaussian Mixture Models (GMM) for automatic identification. Notably, the suggested system claims to be able to spot fake sounds with an amazing 100% accuracy, even when recording in two different rooms with the same mics on each channel. It also does a great job of classifying these different recording settings, getting a 99% success rate, which shows how strong and flexible it is. One of the best things about this method is that it can be used for both text-dependent and text-independent grading. You can't say enough good things about these two types of tests because they are necessary to see how flexible the system is in different voice identification situations. These tests did show some good signs; the system was able to reach a top accuracy of 100%. That shows how well the system works at reliably checking the accuracy of audio content, no matter if the content is linked to specific written prompts or doesn't have any text-based restrictions. This study has important effects because audio authentication is becoming more important in many areas, such as security, forensics, and digital media. Being able to automatically verify recordings with such high accuracy is a big step forward in the field, and it gives security pros and investigative experts a powerful tool. The system's ability to tell the difference between different recording settings makes it even more useful and flexible in real life. To sum up, the suggested automatic audio identification system, which is based on psychoacoustic principles and driven by GMM, has shown that it is very good at finding fake audio and sorting recording settings into different groups. The fact that it does well in both text-dependent and text-independent tests shows how flexible and reliable it is. This study is a big step forward in the area of audio identification, and it could lead to better security and investigative tools in many situations.

VII. FUTURE SCOPE

This study could lead to improvements in voice identification methods in the future, which is a good sign. More research could be done to make the system more resistant to different recording situations and settings so that it can be used in real life. It would also be helpful to work on making the model better at telling the difference between different kinds of recording devices and microphones. More study may be done to see if this method can be expanded to work with a wider range of audio files. If it can, it might be able to be used for more than just binary authentication, like complex classification jobs. Also, as supervised learning techniques get better and more advanced natural language processing techniques are added, the system could produce even more accurate and useful results, making it an important tool for audio evidence and security.

REFERENCES

[1] B. B. Zhu, M. D. Swanson, and A. H. Tew_k, ``When seeing isn't believing [multimedia authentication technologies],'' IEEE Signal Process. Mag., vol. 21, no. 2, pp. 40_49, Mar. 2004.

[2] A. Piva, ``An overview on image forensics,'' ISRN Signal Process.,vol. 2013, p. 22, Jan. 2013.

[3] A. Haouzia and R. Noumeir, ``Methods for image authentication: A survey,'' Multimedia Tools Appl., vol. 39, pp. 1_46, Aug. 2008.

[4] K. Mokhtarian and M. Hefeeda, ``Authentication of scalable video streams with low communication overhead,'' IEEE Trans. Multimedia, vol. 12, no. 7, pp. 730_742, Nov. 2010.

[5] S. Gupta, S. Cho, and C. C. J. Kuo, ``Current developments and future trends in audio authentication,'' IEEE Multimedia, vol. 19, no. 1, pp. 50_59, Jan. 2012.

[6] R. Yang, Y.-Q. Shi, and J. Huang, "Defeating fake- quality MP3," presented at the Proceedings of the 11th ACM workshop on Multimedia and security, Princeton, New Jersey, USA, 2009.

[7] Q. Yan, R. Yang, and J. Huang, "Copy-move detection of audio recording with pitch similarity," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015, pp. 1782-1786.

[8] X. Pan, X. Zhang, and S. Lyu, "Detecting splicing in digital audios using local noise level estimation," in 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pp. 1841-1844.

[9] A. J. Cooper, "Detecting Butt-Spliced Edits in Forensic Digital Audio Recordings," in 39th International Conference: Audio Forensics: Practices and Challenges, 2010.

[10] D. Campbell, E. Jones, and M. Glavin, "Audio quality assessment techniques—A review, and recent developments," Signal Processing, vol. 89, pp. 1489-1500, 8// 2009.

[11] R. C. Maher, "Overview of Audio Forensics," in Intelligent Multimedia Analysis for Security Applications, H. T. Sencar, S. Velastin, N. Nikolaidis, and S. Lian, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 127-144.

[12] B. E. Koenig and D. S. Lacey, "Forensic Authentication of Digital Audio Recordings," Journal of Audio Engineering Society, vol. 57, pp. 662-695, 2009.

[13] Audacity Team, "Audacity(R): Free Audio Editor and Recorder. Version 2.1.2 retrieved on November 25, 2016 from http://www.audacityteam.org/.", ed, 2016.

[14] GoldWave Inc., "GoldWave: Digital Audio Editing Software. Version 6.24 Retrived on November 25, 2016 from https://www.goldwave.com/goldwave.php," ed, 2016.

[15] C. Kraetzer, A. Oermann, J. Dittmann, and A. Lang, "Digital audio forensics: a first practical evaluation on microphone and environment classification," presented at the Proceedings of the 9th workshop on Multimedia & security, Dallas, Texas, USA, 2007.

[16] G. Muhammad, Y. A. Alotaibi, M. Alsulaiman, and M. N. Huda, "Environment Recognition Using Selected MPEG-7 Audio Features and Mel-Frequency Cepstral Coefficients," in 2010 Fifth International Conference on Digital Telecommunications, 2010, pp. 11-16.

[17] M. Huijbregtse and Z. Geradts, "Using the ENF Criterion for Determining the Time of Recording of Short Digital Audio Recordings," in Computational Forensics: Third International Workshop, IWCF 2009, The Hague, The Netherlands, August 13-14, 2009. Proceedings, Z. J. M. H. Geradts, K. Y. Franke, and C. J. Veenman, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 116-124.

[18] D. P. Nicolalde and J. A. Apolinario, "Evaluating digital audio authenticity with spectral distances and ENF phase change," in 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 1417-1420.

[19] D. P. N. Rodriguez, J. A. Apolinario, and L. W. P. Biscainho, "Audio Authenticity: Detecting ENF Discontinuity with High Precision Phase Analysis," IEEE Transactions on Information Forensics and Security, vol. 5, pp. 534-543, 2010.

[20] H. Zhao, Y. Chen, R. Wang, and H. Malik, "Audio splicing detection and localization using environmental signature," Multimedia Tools and Applications, pp. 1-31, 2016.