



The Changing Dynamics of Authentication of Audio Files: a Timescape Review

Jashanjot Kaur^{1,2}, Shivani Sharma³, Mohit Joshi⁴, Abhinav Sood¹

¹ Department of Forensic Science, Chandigarh University, Gharuan, Mohali, Punjab, India-140413

² Punjab Bureau of Investigation, Chandigarh, sector-9D Chandigarh, India-160009

³ Central Forensic Science Laboratory, Chandigarh, DFSS, MHA, GoI-160036,

⁴ Klynveld Peat Marwick Goerdeler, Bangalore

Corresponding author mail id- jashanjotkaur82@gmail.com

Abstract

The rapid and easy proliferation of multimedia files in the digital world has contributed to a significant number of crimes pertaining to the illicit possession, modification, and distribution of multimedia files. The determination of authenticity of digital audio recordings is crucial to prevent misinformation spreading and is also the pre-requisite for the audio recordings to be considered as evidence in the court room. This review paper is designed to give the overview of the both active and passive techniques that exist for determination of authenticity of digital audio recordings in the literature. Besides highlighting the various anti forensic measures and the forensic techniques to counter and detect these, the paper defines the pros and cons along with the limitations of each of these techniques in real life scenarios. Thus, this paper aims to provide a comprehensive vision about the past, present and the future of authentication and thus the integrity verification of the digital audio signals.

Keywords: *Multimedia forensics, Authentication, Digital Audio signals, Forensic evidence, Forgery detection, Tamper detection, Integrity verification, Blind techniques, Non-Blind techniques, AI generation.*

Introduction

When we talk about the twenty-first century, the digital era of communication has certainly altered the ways through which we connect and communicate to the world. The modern age idea of communication and sharing of information has facilitated us with limitless options and wide range of channels to transfer information from one end to another. The means of exchange of information is immensely different today than it was a mere two decades ago when our access to data was limited and we basically relied upon the analog media. With the proliferation of online services and mobile technologies, the analog world has been completely transformed into whole new avenue of digital big data era wherein digital multimedia in form of audio, images, documents, videos, graphics, electronic signals and interactive files features have taken over the crude carriers of information. Most of the today's communication and exchange of information is exclusively in digital format. The data growth statistics provide a sneak peek that more than 300 million multimedia data is circulated on social media per day.

Living in a digitized world wherein every human being is entangled in the web of multimedia in form of audio recordings, images, videos or be it digitized documents, we can very aptly say that we breathe through multimedia. All thanks to the wide range of digital devices entering into the world market every other day, the affordable services of internet, the data storage facilities and connectivity, that have led to the massive flood of digital multimedia. On one hand the digitization and advancement of technology had benefitted us as a faster, cost effective and an easy way of communication, but this has a dark side too.

The widespread distribution of multimedia files has increased exponentially over the past few years with the fast proliferation

of audio-video acquisition and distribution devices and tools. A significant number of cybercrimes pertain to illicit possession, modification, and distribution of multimedia files. Multimedia files (like audios, images, videos, DVR footages, documents, etc) which form an important part of evidence in the day-to-day trials of cases in legal forum are modified or altered by criminal minds with the malicious intentions to hide an important piece of evidence or to misinterpret the information and falsify the true facts and ultimately mislead the justice administration system. With the advancement in tools and techniques every now and then, it has now become alarmingly much easier to forge or manipulate a multimedia file and far more difficult to detect these forgeries and manipulations than ever. New software tools make it way easier to access, create, alter and communicate the content. In such a scenario the verification of integrity (i.e. unauthorized modification of data) and the authenticity (i.e. detection of origin and data alterations) of the content of the digital multimedia files needs to be the core focus area both in social and legal concerns. With the large amount of digital data being circulated at the speed even less than a clock tick and the easy availability of certain freeware and advanced professional editing softwares like Adobe Audition, GoldWave, Audacity, Wondershare Filmora, Photoshop, EaseUs video editor, the multimedia editing has become an easy task and even non-professionals can tamper with the multimedia files without leaving any traces¹. The detection of multimedia editing has been a task in itself which has only intensified with the recent advent of deepfake and Artificial Intelligence. In present day scenarios, the spread of Artificial Intelligence has entered in every realm of the internet, making such manipulations with just a click. This has only increased the questions of the authenticity of the digital multimedia files, when immense number of creation and editing tools are freely available on the internet giving even a maniac a whole set of opportunities to not only edit the files for own benefits but also to harm the others.

Hence, the multimedia forensics is becoming an integral part of digital and cyber forensics² and this combination poses a potential danger. The courts are witnessing the ever increasing incoming of multimedia files (especially digital audio and video files) as evidence in every sector of litigation and criminal justice. As is the case with other court evidences; to acknowledge the digital recordings as a bonafide evidence in the court of law it is necessary to prove the integrity and authenticity of the digital recordings. Thus, the determination of authenticity and verification of integrity of digital audio recordings is the pre-requisite for the recordings to be considered as evidence in the court room. Besides, the audio content forgery detection in broadcasting, especially in this AI era is crucial to prevent spreading misinformation³.

The concept of digital data authentication was first proposed by Diffie and Hellman. Though he did not propose the solutions for digital data authentication, but he brought forward the revolutionary idea of need for one-way authentication and a system of public key cryptography to be based on public and private keys^{4,5}. Following this later three researchers Rivest- Shamir- Adleman of Massachusetts Institute of Technology came up with the solution of *asymmetric* encryption known as RSA public/private key algorithm which could be used to create a shared secret for message encryption⁶.

The very famous Audible Magic's CopySense system that used the "the perceptual characteristics of audio and video" for authentication has been a smart tool that could see past transformative changes to audio and video files such that the transcoding between file formats, equalizing audio, cropping an image or even blurring of a picture could not fool the system. But, the digital security copyright protection technique (copy sense fingerprinting technology) used by the Audible Magic had testing times in 2007 when Gigaom, a tech site was able to upload a copyrighted video from Comedy Central's "The Daily Show" and in another case the Myspace site, which also used Audible Magica's Fingerprinting, also had similar success. This led to a big question mark on the Audible Magica's Fingerprinting capabilities^{7,8}.

For determining the authenticity and verification of audio files, the courts ask the forensic laboratory to testify that whether the audio recording or file is original or not or whether there is any editing or tempering in the questioned audio file. The challenge here is to ascertain that the evidence collected has not been altered using any of the anti-forensic practices and to ascertain that the same is authentic and trusted. The first requisite for this is to determine and verify the recording source of the file i.e. the identification of the recording device or the source of the audio file^{9,10,11} which in itself is one of the biggest challenge with the widespread usage of Artificial Intelligence in the creation of such media files. Only after the detection of the source and verification of the fact that it is generated from the actual source and is not all-together artificially fabricated, the further scientific analysis of the concerned audio file to check for any editing, morphing or tampering or their traces thereof and henceforth identify the manipulation zones in the file. Whilst there are a lot many well established and efficient procedures for assessing the authenticity of analogue recordings^{12,13,14}, it's notable that digital recordings may be edited or falsified in ways that defy the attentions of the forensic examiner.



Figure1. The meaning of the speech recording of the spoken statement "I am not guilty!" can easily be changed to the statement "I am guilty?" or even to "I am guilty!" by cropping or changing base frequency or pitch.

If we dive into the avenue of forgeries in audio files, we can see a whole range of alterations from the very basic and simple alterations like compression, volume enhancement, noise reduction, to content altering forgeries like splicing an evidential part of an audio file, or copy and move or cut or cut and move certain portion of an audio file thus changing the semantic information contained therein the audio file, to deliberate addition of background noise to the audio content and still counting. As demonstrated by Figure 1, a small change in the audio signal can alter the semantic meaning of the file and change the information interpreted from it¹⁵. This is where multimedia forensics comes into play to verify the integrity and to authenticate content contained therein the audio files, to detect forgery in their content and to localize the forged and tempered regions of the content in the audio files.

Categorically speaking, the audio tampering methods can be classified into two broad categories viz a viz (i) Container based tampering and (ii) Content based tampering. The modification in the audio file structure, metadata and its description is counted in as the container based tampering such as the tampering of the time stamp, file format, and hex data of an audio file. While on the other hand the modifications in the bits and bytes of audio wave forms are counted in as the content based tampering of audio files such as time frequency, cut, enhancement, copy-move, and splicing etc.¹⁶ Among these content based modifications of the audio files like splicing the region, copy-pasting occur in the spatial and spatio-temporal domains whereas modifications like deletion, duplication, frame insertion, and shuffling occurs in the temporal domain¹⁷. To fulfill the aim the of deliberate modifications in audio files in the spatial, temporal and spatio-temporal domains and to make the whole process of detection of the remnants of these modifications a difficult and non-invasive task, the modifier tries to or attempts to cover up or remove the detectable signals and traces of tampering by applying modifications like cutting and pasting at the exact zero-crossing or low energy points providing along the slope compatibility, and by interpolating certain softwares that can smoothen the detectable marks. However, the above processes result in loss of natural/original aural features and also the anti-forensic techniques applied to reduce or eliminate the footprints of detection of any tampering themselves lead to another footprint well in accordance with the Locard's Exchange Principle.

Apart from defining and categorizing the tampering methods, there are also types of the tamper detection techniques namely active detection techniques (also known as non-blind techniques) and passive detection techniques (also known blind techniques)^{15,18}. The difference between these two techniques lies in the fact that non-blind technique requires the original file or its representation for tamper detection i.e. it is majorly a comparison technique whereas the latter does not require the same¹⁹. Active methods include techniques like watermarking and digital signatures which have some key information embedded to the content of the file which is used to authenticate the said file. Any changes to this embedded information, signals the presence of certain tampering in the file. But these active methods of tamper detection face certain drawbacks in the most realistic scenarios. Firstly, not every device has the watermark or digital signature embedding capability to the audio file captured, secondly it is a costly and complex affair in case of audio files and thirdly this technique definitely fails in scenarios where tampering is performed even before inserting the watermark or digital signatures. This method is basically and most

commonly used in copyright protection and security applications. So here comes the importance of passive methods when dealing with complex realistic scenarios where the prior metadata information about the audio file might be missing. These methods make the usage of features of the audio files which are extracted and analyzed to verify the integrity and authenticity of the file. The passive methods basically exploit the footprints left by the editing operations in the audio file like noise residues, abnormalities in acoustic frames, quality compromises, etc. However, all such methods have themselves become the challenges of the present day forgeries because no human factor is involved in the creation of such manipulations but the technology itself is doing it making it next to impossible for the Human based interpretation of the signal which leaves no or very minimal traces of variation between the signal of Human and Artificial origin.

Many researchers have been working in this field since the mid-90s, and have come up with many breakthrough methods of authentication, detection and localization of forgeries in multimedia audio files particularly which will be discussed in depth in this systematic review. The most common sources of such media files are mobile phones, digital cameras, CCTV cameras and online platforms/sites. The early and the foremost method generally applied to authenticate such data is to compare the file under question with the original file or to the output of the original recording device. But, the presence or seizure of the original media file or the recording device is generally not possible in the real time crime cases. The original devices are not recovered in almost 80% of the cases. Thus, it becomes very important to research and find out the ways to deal with multimedia forgeries with the real time applications. Although there are numerous tools available that can create, encrypt and extract data, there are somewhat minimal tools that can establish integrity, authentication and provenance. This paper presents overview of the methods or techniques used for integrity verification and authentication testing of the audio files and signals in the past and present along with the possible future projections of research in this field.

This paper aims to provide a comprehensive vision about the past, present and the future of determination of authenticity and verification of integrity of digital audio recordings and their localization. It is designed to give the overview of the both active and passive techniques that exist for determination of authenticity and verification of integrity of digital audio recordings in the literature. Besides highlighting the various anti forensic measures and the forensic techniques to counter and detect these, the paper defines the pros and cons along with the limitations of each of these techniques in real life scenarios. This review brings out the picture that “hearing is no longer believing”.

Examples of Audio Authentication cases

The history of audio forensics can be traced back to the late 1940s, the time when the recording systems got an upgrade from the studio recordings to the recording systems that could be used for recording outside the studio. The Federal Bureau of Investigation (FBI), the leading intelligence and security agency had already developed its expertise in audio forensics the early 1960s. There are certain noteworthy cases wherein the admissibility of audio recording as an evidence is evident like the *United States v. Schanerman*, 1945 case in which the Court considered the inclusion of mechanically recorded conversations as evidence, the *Monroe v. United States* case, 1956 the Court opinionated upon the admissibility of conversations between defendants and a police officer as evidence provided the officer testified as to the accuracy and reliability of the recordings²⁰. The case - *United States Of America v. Thomas McKeever and Lawrence M. Morrison* (defendants), 1958 witnessed the usage of the audio forensics for the recorded conversations as evidence in the court of law and the concept of authentication of audio recording was highlighted. The evidence presented in the court of law were the two tape recorded conversations between McKeever and the Ballsons with Jack Ball and one with George Ball that were recorded by the defendant McKeever or by persons acting in his behalf. The ruling was reached that established a foundation for sound recordings concerning their admissibility as an evidence in the Court of Law and cited this as seven requirements which are as follows:

- i. That the recording device was capable of taking the conversation now offered in evidence.
- ii. That the operator of the device was competent to operate the device.
- iii. That the recording is authentic and correct.
- iv. That changes, additions or deletions have not been made in the recording.
- v. That the recording has been preserved in a manner that is shown to the court.
- vi. That the speakers are identified.
- vii. That the conversation elicited was made voluntarily and in good faith, without any kind of inducement.

These requirements are now assumed by most state and federal courts in the United States and referred to as “The Seven Tenets of Audio Authenticity”²¹.

Another noteworthy and highly publicized forensic audio case is the assassination of former US President John F. Kennedy in 1961 at Dallas which is considered to be the most enduring and controversial mysteries. The President was shot with bullets while he was on his way near Texas School Book Depository in his car. The recordings connected with case contained were those of the air traffic from Air Force originally recorded on 22 November 1963 and the conversation between the pilots and other on flight individuals and individuals in Washington DC. The recordings also include conversation about the disposition of

the President's body by the head of Secret Service. These audio tape recordings are available at GPO and the National Archives.^{22,23}

In 1973 Forensic Audio Engineering explicitly gained momentum with the famous "Watergate Scandal" investigation. The then President Richard Nixon got audio tapping system installed in the White House and in the Executive Office Building which was known to only the Secret Service and a selected group of individuals. Amidst the ongoing Watergate Scandal investigation, the White House aide Alexander Butterfield exposed the presence of secret audiotape recordings of conversations between the President and his advisors in his testimony before the Congress and Federal Court. Upon this revelation, the President Nixon agreed to release edited transcripts of the various conversations recorded and eventually the tapes thereafter. In 1974, interest turned to a particular recording of a conversation between President Nixon and his Chief of Staff H. R. Haldeman recorded in 1972 in the EOB, which the investigators believed to contain the remarks about the Watergate cover-up. During the course of investigation when the recordings were turned over to the court for transcription an unexplained gap of 18 and a half minutes containing a buzzing sound that completely obscured speech was discovered in the tape of a conversation between Nixon and Chief of Staff H.R. Haldeman on June 20, 1972. Investigators suspected that it could be an intentional attempt to coverup the original conversation. The Learned Chief Judge seek the detailed examination and analysis from the advisory panel of the experts about the cause of the gap in the tape recording i.e. whether it was caused accidentally or deliberately or whether it was a result of some malfunctioning at the time of recording. After a comprehensive study of the tape and the electrical and magnetic signals on it and the properties of the recording equipment, the advisory panel reached the conclusion that multiple overlapping start and stop erasures on the tape by the magnetic erase head of a specific model of tape recorder were different from the device that produced the original recording. The panel's methodology was a breakthrough in the field of authenticity of audio recordings and is still widely used and accepted procedural model to test the authenticity of audio recordings.²⁰The procedural examination carried out by the Advisory Panel which is listed in four steps became the basic standard for audio forensic investigators. These were the first few cases that laid the foundation for audio authentication and helped in shaping the techniques and build the reputation of the field.

Methods

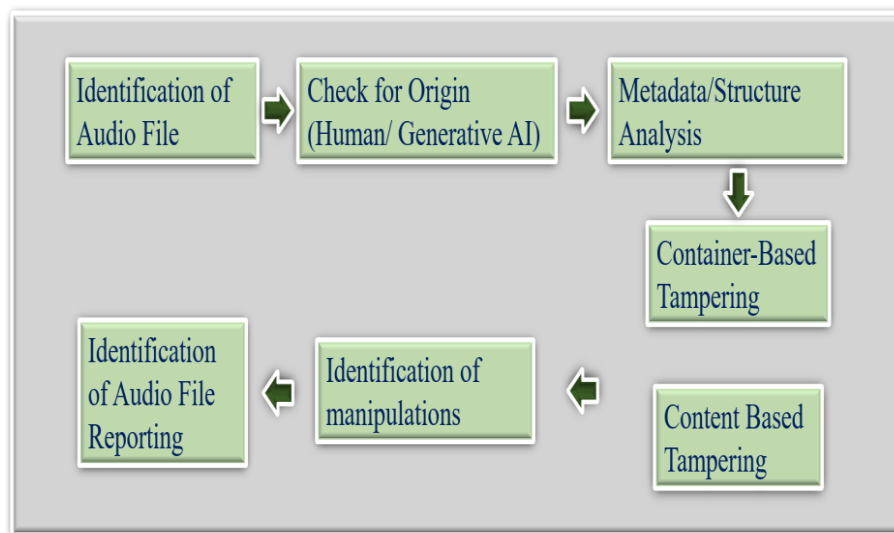


Figure 2: Block diagram of the general method followed for audio authentication

In the past era when the magnetic tape recordings were the medium of audio recordings, the authenticity determination and integrity verification required the analyst to basically observe the physical integrity of the recording medium, analyze the quality of the audio file and the consistency check of the magnetic signatures that are present on the magnetic tape. But with the plethora of the advanced recording mediums and digitization, the authenticity determination and integrity verification also required an up gradation from the traditional practices of forgery detection and localization. Initially, most of the efforts by the researchers and academicians had been majorly concentrated on identification/determination of the ownership of multimedia files and upon authenticating the user of the multimedia files²⁴. This is very well evident from the literature that most of the approaches in the past focused on ownership and copyright protection¹⁵. But from the past 6-7 years determination of content manipulations and forgeries in multimedia files has attracted the interest of the research community. Although in this direction also, major work has been focused on the images and video file contents. The audio domain is still in its infancy stage as only a few approaches with their limitations exist in this direction that identify and classify the basic audio features that could be

employed to counter, identify and localize the content manipulations. This section brings forth the different techniques that have been explored and exploited for verification of the integrity and authentication testing of the content of audio files. Figure 3 gives the overview of the different approach and Techniques utilized while authenticating audio signals.

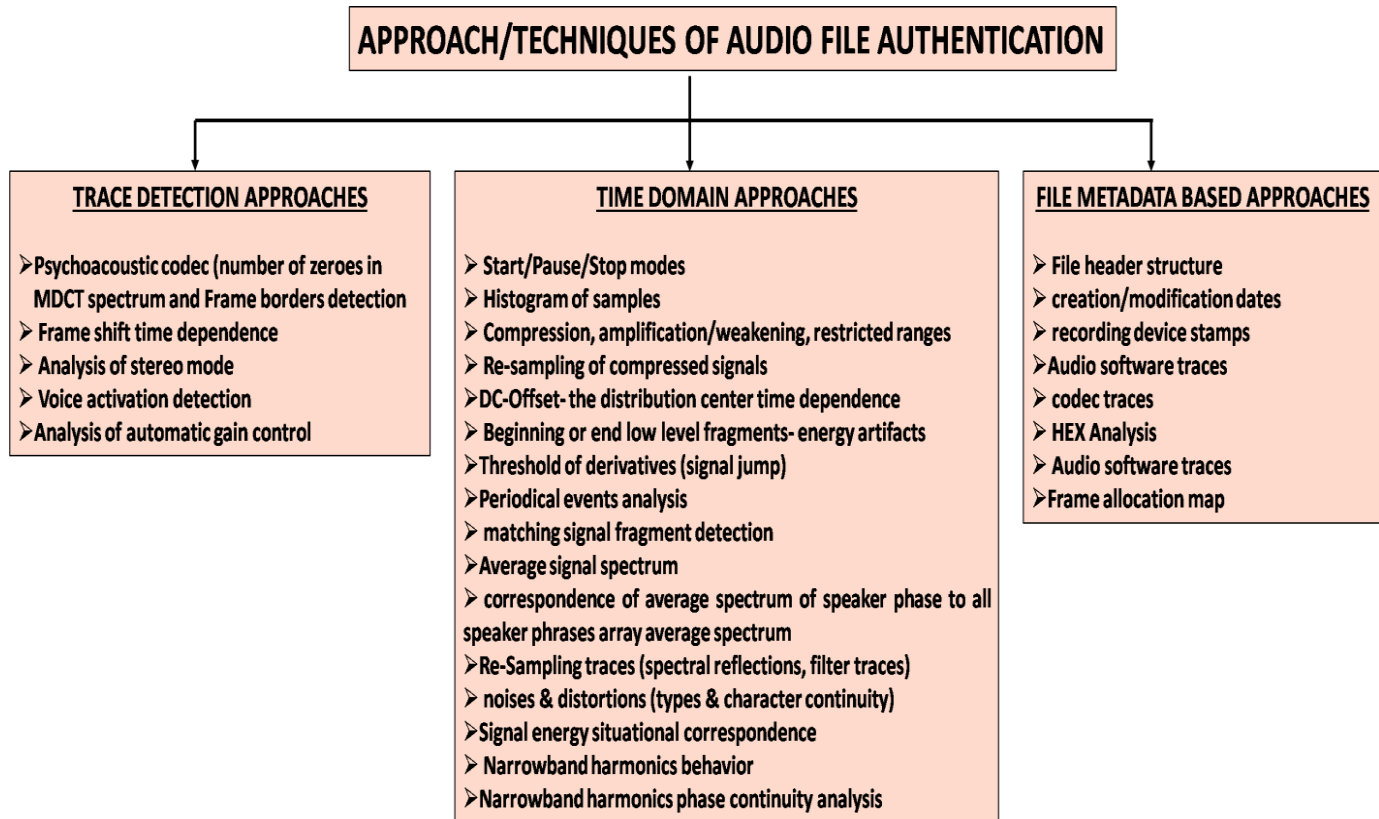


Figure 3: List of different approaches and techniques used for audio file authentication.

I. Critical Listening method

Critical listening is considered one of the very first and basic techniques for integrity verification and authentication testing of an audio file. The examiner does the depth analysis of each segment of the audio signal by repeatedly listening to the audio content and looks out for any deviations in the segments of the audio signals from the recorded characteristics. However, the audio recordings captured in real time are on an average lengthy with background noise and other acoustic artifacts embedded into it and comparatively the tempered regions in audio recording may be very short. This lends a drawback to the technique of critical listening as repeated hearing of any forged part becomes cumbersome. In addition, this form of examination requires a lot of time and sometimes render low accuracy rates of detection of manipulations due to the post audio forgery operations that are used to hide the traces of forgery. No doubt critical listening still remains as an important part of integrity verification and authentication testing procedure, but it can merely be used as a preliminary examination and act as an aid/base for instrumental examination of audio file(s) ²⁵.

II. Pause Detection method

The pause detection is the very basic concept of examining the forgeries in the audio recordings. In this threshold approach certain feature are chosen which act as the indicators for e.g. the loudness measure to classify and segment the pauses using three parameters: (i) an adaptive threshold (to determine if feature frame represent silence or not), (ii) a minimum duration (distinct the semantic levels by capturing varying levels of background noise) and (iii) a maximum tolerated interruption constraint (it influences only segmentation to increase the accuracy and does not influence the semantic level. Thus, it is all the work of algorithms which is improved further using a two-call Hidden Markov Model (HMM), which uses viterbi decoding to determine maximum likelihood classification, results in better segmentation depending upon the training of HMM and as such

differently trained HMMs are trained to arrive at different semantic pauses level. Apart from the silences there are three types of acoustic pauses (i.e. silent, filled and breath pauses), syntactic pauses (i.e. punctuation marks in speech transcripts) whose patterns and properties can help in correlation and determination factors^{26,27,28}.

Another very important factors used in forgery determination is the offset method using cues¹⁹. Offset refers to the number of frames at the pause end. Depending upon the modifications the cue shifts either to the left or right. If the whole pause is cut the cue at pause end will be created, thus the offsets can be determined to analyze the pauses. Apart from pauses the start and stop modes in the audio files are also an important clue to determine the basic audio forgeries.

III. Sampling Points Corelation method

The audio file is made up of segments which can be further divided into sample points. The position of these sampling points acts as a determinant for the detection of tampering or modification in audio files due to various operations like cutting, splicing, inserting, duplication, replacing, etc. Such audio content modification operations result in the proximity of sampling points in the time domain which results in decrease or disappearance of correlation between the sample points and its neighbors which leads to certain abnormal mutation points called Singular points ²⁹.

IV. Gibbs Effect method

The literature highlights another time domain feature i.e. Gibbs effect which has been deployed majorly as a supporting method since 1998. Gibbs effect occurs at the tampering positions and as such causes discontinuity in the underlying original audio generation process. The Gibbs oscillations i.e. the traces of manipulations can be visible in both waveform and Fast Fourier Transform spectrograms and are audible too as a metallic sound. However, the major counter forensic that is found is that the Gibbs oscillations can be camouflaged in between the signals or can be smoothed out with the use of certain software's leaving the trails of tampering undetected and also certain Gibbs oscillations get disappear in the electronic background white noise³⁰. The authors proposed supplementary Enhanced Audio Files Examination Method (EAEM) to enhance and highlight the visibility of Gibbs Effect which was developed in the Forensic Audio Laboratory of Istanbul University Forensic Sciences Institute. The drawbacks of this technique include the time-consuming process, low success rate in forgery detection when editing software's smoothen out the manipulations or when the segments are lost in the background noise.

V. Digital Fingerprinting method

Digital fingerprinting is a computerized form of security measure used to protect multimedia files and information and ensures that the integrity of multimedia files and transmitted information and messages. Digital audio fingerprinting is a content-based compact signature for audio recording. Besides Content-based integrity verification or watermarking support, there are many more applications to audio fingerprinting but this review primarily focuses on manipulation detection³¹. Audio fingerprints usually show the feature of being invariant to inaudible or even moderate signal transformations, e.g. lossy compression or microphone recording. However, in case that major transformations of the content or malicious tampering are applied, the current audio fingerprints do not match the original audio fingerprints, which, in effect, indicates major integrity infringements. An audio file is a sequence of basic sounds which can be decomposed into audio units. These audio units are also modeled with HMM and used as robust fingerprint. Authors have also worked upon extracting audio fingerprint from the Fourier domain and embed it as a digital watermark in the same domain. ³².

VI. Hash Functions method

Digital fingerprinting also works basically by usage of mathematical functions called hash value. Hash method, such as MD5 (Message Digest 5) or CRC (Cyclic Redundancy Checking), SHA are used to obtain a compact representation of the binary file.

If a file is tampered with, even just in terms of the alteration of a single character, the digital fingerprint for that file changes. The hash digital fingerprinting is basically classified into preimage resistance (one wayness), second preimage resistance (weak collision resistance) and collision resistance (strong collision resistance) depending upon the security properties. Majorly cryptographic hash functions are employed for general security purposes and second preimage resistance is used for file authentication ³³.

Two of the most commonly encountered hash functions are MD5 (Message Digest algorithm 5) and the Secure Hash Algorithm (SHA) family. With time in around 2004 the collision resistance vulnerabilities of MD5 were in light, and subsequently the credibility of MD5 was effectively terminated in the field of digital signatures and was then considered adequate only to check consistency of files and for other non-secure application. At present the SHA family is the predominantly used algorithm which

employs avalanche effect and its most widely used versions include SHA-256 and SHA-512 which are based on Merkle-Damgård construction. Basically the character of digital fingerprinting technique is preventive rather than tracking-based and has value for comparative analysis. The only drawbacks being that the digital fingerprinting technique only works with media that has a reference file, but in real case scenario most of the times don't have any reference file for forensic examination and analysis require the blind detection of authentication and integrity verification in the absence of any reference media file¹⁸. And second drawback being that even the technique is too sensitive towards the tolerable manipulation like loss transcoding, format conversion also experiments show that even if it is robust against severe compression, but bit errors do occur³⁴.

In this approach since the algorithm compares the compact signatures instead of the whole files, therefore this method is not robust to compression or distortions of any kind, and might not even be considered as content-based identification of audio, since it is not based on an analysis of the content (understood as perceptual information) but just on manipulations performed on binary. An important factor about hashing method is that hash values are very fragile and even a single bit flip results in complete change in the hash of a file. But still if we compare hashing with watermarking, the former is ideally less vulnerable to attacks and distortions and does not compromise with the quality of the sound and does not require any kind of modification to the audio content unlike watermarking. The only limitation of hashing when compared to watermarking is that the complexity of hashing is higher.

a. Audio File Encryption using crypto-hash method

Crypto-hashes are of course are not only sensitive to malicious tampering, but also to inaudible signal transformations of media data that might be tolerable in many scenarios. The scientists have proved and proposed many algorithms for audio file encryption like Dr. Ekhlas a five level of encryption method based on a combination of block cipher and chaotic maps³⁵. Ardalan Ghasemzadeh and Elham Esmaeili proposed a novel audio signal encryption method based on a mixture of three chaos functions³⁶. Monther and Azman suggested a method to create a cryptographic public key to encrypt the data. Since cryptographic hashes rely on the avalanche effect wherein the slightest change in the source possibly produces widely different hash results³⁷.

b. Perceptual Hashing method

Perceptual hashing a sub category of hashing functions maps the source data into hashes while maintaining correlation and thus measure the similarity index between the source data. Perceptual hash functions are analogous if the features of the multimedia are similar. This method is widely used in copyright infringement cases, detect multimedia copy by identifying similar contents and for authentication of malicious manipulations for multimedia authentication in digital forensics³⁸. The perceptual hashing method has been discussed and proposed for image and video copy move forgery detection, but we get very little or no evidence of its usage for the forgery detection in audio files³⁹. This is basically a blind authentication scheme widely utilized for content identification, authentication and retrieval. Perceptual algorithms like aHash, dHash, pHash, etc generate fingerprints for digital multimedia which can be used for comparison purpose⁴⁰. The robust perceptual hashing incorporates acoustic features in the hash function for audio characterization like loudness, energy, zero crossing rate, spectral centroid, pitch, spectral flatness, harmonicity and Mel-Frequency Cepstral Coefficients (MFCC's), to bring out the final hash code robust to audio manipulations as long as the content is preserved⁴¹.

The ideal fingerprinting system should fulfill certain basic requirements as follows⁴²:

- i. regardless of the level of compression and distortion or interference in the transmission channel it should be able to accurately identify an item.
- ii. It should be able to deal with other sources of degradation like pitching, equalization, background noise, D/A-A/D conversion, speech and audio coders, etc.
- iii. It should be computationally efficient.

VII. Digital Watermarking method

Digital Watermarking technique works on the basic principle of embedding payload data i.e. a watermark or signature to the cover work i.e. the digital media contents through methods like direct embedding and seed-based embedding. This technique was originally focused on limited application of copyright protection and ownership detection, but overtime it has been exploited and accepted as a technology with wide range of applications including multimedia authentication and high security applications¹⁵. The digital watermarking can be characterized and defined into five classes based upon the application fields i.e. (i) authentication watermarks (ii) fingerprint watermarks (iii) copy control watermarks (iv) annotation watermarks and (v) integrity watermarks^{43,44}. Whereas robust watermarking schemes are peculiarly used for copyright protection and

ownership identification due to their property of withstanding attacks (like common image processing), the more sensitive fragile, semi-fragile, content-fragile, self-embedding, and invertible watermarking schemes that are fragile to linear and nonlinear transformations and 3D hologram approach are generally approached for multimedia authentication over continuing video and audio frames for integrity verification respectively and extended cryptographic approaches are deployed for security applications. There are lot many publications directed to work for image and video data but with very limited approaches for detection of both malicious and inadvertent changes to digital audio to ensure data authentication for high-security applications^{44,45}.

In Steinebach, M.¹⁵ the authors have conceptualized two content authentication schemes for digital audio data on the basis of two different approaches i.e. content-fragile watermarking and invertible audio watermarking. The first approach presents the basic content fragile watermarking scheme with an advanced approach of using feature checksums instead of the robust features to circumvent the payload problem. This not only decreases the high error rates at retrieval but also reduces the high false error rates by ignoring the content-preserving operations. This watermarking algorithm is not only robust against the cropping attacks but also works out for the cut out samples through its sync compare function. In the second approach presented in the paper, the authors bring out the frame-based solution through invertible watermarking. In this approach the authors used the basic invertible watermarking scheme with the advanced lossless compression algorithm to produce a compressed bit vector and an additional embedding of the security information through Sequence IDs, an Incremental frame ID_T and a Synchronization header Data_{sync}. The first approach is suitable for the scenarios like news data authentication wherein certain operations (like the binary representation of the content and several other postproduction modifications) are acceptable and tolerated whilst identifying content changes and the second approach using invertible watermarking scheme as basis finds its application in surveillance recordings, forensic recordings, etc where even small changes in the audio can effect and thus require high content security.

N.V. Lalitha, Ch. Srinivasa Rao, and P.V.Y. JayaSree proposed a watermarking-based method. They utilized Discrete Cosine Transform (DCT), Quantization Index Modulation (QIM) rule to watermark voice segments and then inserting hash bits detect copy-move Forgery in speech signals. The proposed method has been claimed to be robust against signal processing attacks such as low-pass filtering, resample, compression, jittering, and cropping.⁴⁶

Generally, there is mix up of digital fingerprint and a digital watermark, but these are two very different technologies with somewhat similar goals. The term watermark, typically refers to a completely visible marking on a digital file.

The major general shortcomings of usage of the watermark technique approach are that:

- (i) the audio recordings encountered in real life scenarios are random and do not always carry digital watermarking or signature content for authentication due to the expensive procedure and lack of inbuilt software/hardware support configurations in the most commonly used non-professional recording devices like mobile phones, digital cameras and recorders.
- (ii) Also, another major challenge is the way in which the watermarking algorithmic approaches can differentiate between the friendly and non-friendly attacks or operations i.e. the content-preserving and content-changing manipulations as majorly the known existing techniques use threshold-based approach for content integrity. As seen in literature majorly compression alterations have been addressed and very often the other friendly attacks like filtering, format conversion and scaling still remain unaddressed. This scheme is thus unsuitable in situations like wherein the watermark is embedded in Least Significant Bit (LSB) which is fragile towards common signal processing operations. Therefore, lately the researchers have been more focused on developing active audio forgery detection.
- (iii) Certain watermarks are embedded in public and fixed frequency points, and practically it is a lot easier to get the features and as such tamper such watermark in the frequency domain by the attackers.
- (iv) Also, the attackers concerned are well aware of the embedding and extracting methods of watermarks and thus, possible for them to extract the embedded watermark from some another audio file representing same features and then embed or substitute the frame which will go undetected by the authentication client. Such an attack is called "Feature-Analyzed substitution attack".
- (v) Certain audio watermarking schemes embedding arbitrary message alter the perception of the sound.
- (vi) Also this scheme of authentication fails in cases where the embedded watermark gets corrupted and thus renders unmatched results³².

VIII. Electric Network Frequency based methods

Power grid signals in some recorded signals also act as an evidence to authenticate and verify the integrity of audio/video recordings, Electric Network Frequency commonly known as ENF method is the technique based on evaluating and interpreting the fundamental frequency of the current signal in the electrical network present in a questioned recording and comparing it with an appropriate and reliable reference ENF signal⁴⁷. In the audio recording captured by digital equipment

connected to electrical outlet, the ENF designed to be 50 Hz (e.g., in most of European countries and China) or 60 Hz (e.g., in US), will most likely be embedded in the recorded audio signals. By analyzing the ENF consistency the audio signal can be assessed for alterations or modifications. This approach of authentication, validation and integrity verification of audio/video evidence was initiated, assessed and executed at the Speech and Audio Laboratory of the Institute of Forensic Research in Cracow in 2002 using Grigorias's work⁴⁸. The reference recording for this can be any recording whose authenticity is beyond doubt or that which belongs to particular momentary ENF registered by the power grid operators. Most European countries belong to the Union for the Coordination of Transmission of Energy (UCTE) and their electrical networks are linked to each other to form one network. The asset of this authentication method is that it works not only for mains-powered recording devices but also for battery powered recording equipment's and those in a built up or close proximity to some other mains-powered equipment or transmission cable. In addition works even in the absence of an ideal voltage regulator⁴⁹.

The ENF Criterion is based on two main assumptions: firstly, that the frequency of interconnected networks varies equally at all points of this network at a given moment and secondly these variations in frequencies are non-repeatable over a long range of period. An important consideration for ENF technique to work is to check if the digital recorder signal has a component of electromagnetic field generated by the electric network. Majorly it has been seen that to provide an opinion about authenticity of audio/video recording, the spectrograms visual inspection and comparison of ENF signal with ENF database are enough. Short time window measurements and comparison of ENF signals lead not only to the determination of authentication but also establishes where and when the editing had occurred⁵⁰.

Broadly classifying there are three methods reported to extract ENF in literature-

- i. 'time/frequency domain' spectrograms method which consists of computing spectrograms and visually comparing the ENF of questioned recording with the database ENF. This is the fastest and easiest method to implement which is usually employed for date and time verification. It basically provides the number of different ENF components present in the digital recording.
- ii. 'frequency domain' method which computes FFT over short time windows, extracts the maximum magnitude value around 50 Hz and then compare the ENF of questioned recording with the database ENF. This method is basically for verification of ENF and to identify the recording date and time. This approach can likewise also be utilized on more than one ENF component recordings to isolate and identify the ENF components.
- iii. 'time domain' analysis involving zero-crosses measurement and comparison of ENF of questioned recording with the database ENF. This is the deepest ENF extraction and analysis

In another research work the author proposed a novel spatial domain method to determine editing in audio file and for localization of manipulations based on comparison of the changes in max offsets on block-by-block basis⁴⁹. Another feature is the phase discontinuity check of ENF signals estimated through Fourier Analysis to infer the editing type and locate the concerned editing points⁵¹.

ENF based methods can be credited as one of the most powerful, and reliable methods in the discipline of detection of forgeries in audio signals. Their accuracy of forgery detection in plug-in powered devices is just unrivaled with an accuracy beyond any doubt. Taking into account the focal problem of false positives and the false results in finding the presence and locating the area of tempering, the battery powered devices pose as a major and challenging confront to the ENF based audio signal forgery detection. This is majorly due to the reason that audible hum is the source of the ENF signal that is extracted from the battery powered devices (which). The same has been very well documented through a demonstration and calculation of the accuracy of the results of the audio signal forgery detection in plug-in powered devices and battery powered devices by Savari, M. The author employed a combination of ENF and Phase features for detection of audio signal forgery in battery-powered devices and established a powerful achievement of detection of deletion and insertion forgery with an unbeatable accuracy of 88% and 92 % respectively for mobile phone and laptop.⁵²

In addition, ENF signals act as a promising tool to determine and verify the time stamp of the digital audio and video recordings. Various works explored and examined the anti-forensic operations that can alter or remove the ENF signal and can aid in concealment of any anti-forensic tell tales left behind⁵³. The author in [52] highlighted the potential solutions like inter frequency consistency check, spectrogram consistency check, reference-based statistic matching method and lookout for temporal variations to counter and detect these and presented an envision in the field of dynamic anti-forensic.

The ENF Criterion promises to be a significant breakthrough as a forensic technique that will outshine as a technique for defeating the efforts of the audio/video forger in digital domain. However, there are various challenges or shortcomings that pose a hindrance to this technique as listed below:

- i. The task of establishing the date and time of creation of digital audio recording involves great deal of time for searching. Although the researchers have tried to compensate this with automated systems based on signal correlation methods but their accuracy is still questionable vs the manual method.
- ii. This method requires reference ENF signal database. And this method fails in countries which do not have ENF registered powergrid system.

- iii. Several extraneous phenomenon's like transients, power-line spikes and surges, coding artifacts, etc. due to recording system, inherent to ac transmission or acoustic environments impact the ENF detection performance which needs to be the objective of future research work.

IX. Utilising Metadata for Authenticity

Metadata usually referred to as the data about data. It describes the data by providing a significant amount of information about a file without examining the content of the file. Metadata basically helps in corroborating the details of file with the information of the file which serves as one of the aspects of file authentication, for example, the metadata detail about the content creation device and/or application. There are two types of metadata of the files that aid in authentication of the said files i.e. files system metadata and digital media metadata. File system metadata not only includes the basic information about the file like: size, created/ modified/accessed date and time stamps but also includes MAC times, group ownership, MD5 hashes associated with media files.

James Migletz presented an automated approach of exploiting and capturing the information about the media file⁵⁴. Yet in another work, the authors have focused on the examination and analysis of the file signatures and meta data properties of the audio files recorded using different devices and recorders and edited/tempered using softwares like Adobe Audition, Audacity, GoldWave, Oceanaudio, Wavepad Sound Editor available on open-source platform. The freeware software tools (like Media Tab, Media Info, Hex Data Viewer, Winhex, Hex Workshop, HXD) have been utilized to compare the metadata of original files and the edited files on the basis of (a) Time stamp i.e. the file creation and modification date and time, (b) Recording device or source identification, (c) Identification of editing software using hexadecimal reader and header information of file format⁵⁵.

File signature analysis and metadata analysis can very well help to authenticate the source of the audio file, but there is a certain limitation when it comes to determining any tempering or editing of the audio file in the same device from which it originated. The file headers of the audio file do not change when editing like trimming of the file was done using the same device i.e. the device of origin of the file. However, there is absence of any evidence related to date and time in the edited file v/s the original file. The software's such as EnCase, FTK Imager, Winhex, Hex Editor, Hex Workshop, etc can be used to display the file signatures, metadata and contents of the file in binary, hexadecimal, decimal, or ASCII. These help the analyst to identify and compare the system-generated artifacts of multimedia files created using various devices or downloaded from different social media applications like YouTube, Facebook, Viber, WhatsApp, etc⁵⁶. Reviewing and digging deep into the existing methods of editing and inserting metadata into multimedia files, it has been observed that there is very limited scope and native support for metadata editing in operating systems except ExifTool.

X. Discrete Cosine Transform (DCT) and Singular Value Decomposition (SVD) based approach

The voice activity detection (VAD) algorithm involving DCT and SVD is a blind forensics approach of detection of copy-move forgery in audio signals. In this approach the audio signal is normalized and divided into several segments to obtain syllables which are then processed using DCT and SVD. DCT processing generates DCT coefficients which when processed using SVD give singular vectors. The calculation and comparison of the distance between two singular vectors help to detect and locate copy-move forgery in audio signals.⁵⁷

XI. Comparison of Narrow Band Harmonics approach

In another method of detection and localization of copy-move forgery in audio signals, Qi Yan, Rui Yang and Jiwu Huang compared the similarities of pitch and formants with a threshold. In the procedure the audio signal is divided into voiced speech segments and unvoiced speech segments and feature set of each voiced segment is extracted for pitch sequence and first two formant sequences. Dynamic Time Warping is further utilized which computes the similarities of all the feature sets which are compared to the threshold to detect and localize the copy-move forgery. The proposed method has been found to be robust against background noise and most commonly used post processing operations.⁵⁸

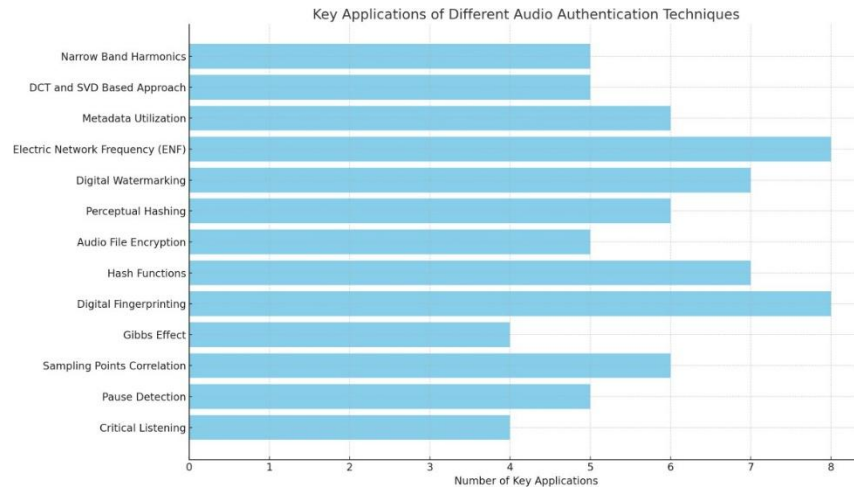


Figure 4: Key Applications of discussed methods

XII. Generative AI Detection Approach

In 2026, 62% of organizations report a deepfake incident (Gartner, n=302), with 41% hit on audio calls⁵⁹. The statistics themselves suggest the need for detection of origin of audio files to verify their integrity and authenticity. The auditory features of generative AI to clone the voice are so realistic that many a times even the known person is unable to recognize the voice with accuracy. Under such circumstances, the forensic experts are in vain because of the lack of the familiarity with the actual voice and additionally because of the veracity of the generative tools available in the market/online form. Further, the aforementioned techniques are also not fully applicable to detect the authenticity if the AI technology is used in the generation of the media. To counter this issue, many efforts are being made but no robust method is yet available which can accurately detect the origin of the voice cloned multimedia files. Yet, the cases like Hong Kong Bank Heist where fraudsters transferred over \$25 million by using cloned voices of company executives and Armenia Prime Minister Deepfake case where fabricated audio clip was circulated online regarding investigation are increasing tremendously.

The most common method of analyzing if the clip is generative or original is to check its spectrograms and the mel spectrograms to look for the spectrographic variations in contrast to the human voices. However, such methods are not objective and depends upon the expert's interpretation. The work of Bhagtani K. *et. al*⁶⁰. in this direction paves the methods for scientific analysis of the generative multimedia files. They had used patchout spectrograms, unsupervised fine grain attribution, diffusion based deepfake detection for the detection of generative media. However, such methods are practically not feasible in actual case analysis due to the limited amount of speech available in the questioned samples. Thus, the source identification of the multimedia files in the present AI era is one of the biggest challenge which yet remain unanswered despite various efforts.

Conclusion

This research review stands here to identify the various methods available and used with their features, capabilities and limitations. None of the methods discussed above like pause detection, audio fingerprinting, hashing, watermarking or frequency-based domains like ENF can be compared completely to one another because as of now there does not exist any universal tool for audio tampering detection that can detect and locate all type of forgeries like cut, splice, copy, move, copy-move, noise addition, *etc.* at a single platform. Moreover, no sure shot tool is available that can clearly distinguish the origin of the multimedia files i.e., if they came from the actual source or is generated using AI. The various research techniques outlined above are promising to establish integrity, authentication and provenance of the digital audio files, but still there is an urgent need for effective and efficient methods and technologies capable of detecting and locating all types or maximum types of forgeries in the digital audio files using a single platform. Not only this, but the present scenario requires their anti-forensic counterparts to be envisioned and explored.

Future Projection

1. Our future work should not only focus on developing new techniques to authenticate and verify integrity, instead more focused to decrease the erroneous detection and increasing the accuracy of locating the boundaries of edited regions in the existing techniques and methods of forgery detection.

2. Metadata are an important subject of research focusing on both the design and automatic extraction approaches.
3. One of the very major and basic challenges faced by the researchers during the course of evaluation of forgery detection algorithms is the scarcity of the ready to use datasets that are available in public domain. Therefore to facilitate the evaluation of the latest algorithms and to check and compare their efficiency, the preparation of dynamic database of audio forgeries can be considered on priority terms.⁶¹
4. As seen in literature majorly compression alterations have been addressed and very often the other friendly attacks like filtering, format conversion and scaling still remain unaddressed. Therefore, another major challenge is to identify audio features appropriate to distinguish between content-preserving and content-changing manipulations i.e. malicious and non-malicious attacks to avoid friendly attacks to hinder the authentication and integrity verification process.
5. Biggest challenge right now that needs attention is the authentication and verification of integrity of AI generated media, social media processed or shared files.
6. Research on counter-forensics needs to be speeded up assess and improve the reliability of forensic methods in situations where intelligent opponents make efforts to falsify evidence.
7. The validity and effectiveness of the developed methods should not only be verified by experiments not only on synthetic composite or ideal signals but also the real world audio signals or case studies must be included to make the work even more promising for real world applications.

Conflict of Interest

“The authors declare no conflict of interest.”

REFERENCES

1. Jackson, W. *Digital Audio Editing Fundamentals*.
2. Lian, S. & Zhang, Y. Multimedia Forensics for Detecting Forgeries. *Handb. Inf. Commun. Secur.* 809–828 (2010) doi:10.1007/978-3-642-04117-4_37.
3. Koenig, B. E. & Lacey, D. S. Audio and Video Files. (1974).
4. K. Sujatha, P. V. Nageswara Rao, A. Arjuna Rao & L. V. Rajesh. Renowned Information Security Algorithms: A Comparative Study. *Int. J. Eng. Res.* **V5**, 216–224 (2016).
5. Thirananant, N., Lee, Y. S. & Lee, H. Performance Comparison between RSA and Elliptic Curve Cryptography-Based QR Code Authentication. *Proc. - IEEE 29th Int. Conf. Adv. Inf. Netw. Appl. Work. WAINA 2015* 278–282 (2015) doi:10.1109/WAINA.2015.62.
6. Messerschmitt, D. G. RSA Asymmetric Encryption. 18–21 (1999).
7. Audible Magic-Baidu Library.
8. U.S.HOUSE REPRESENTATIVES. The Role of Technology in Reducing Illegal Filesharing: A University Perspective. *U.S. House of Representatives Committee on Science and Technology* 1–7 (2007).
9. Zou, L., He, Q. & Wu, J. Source cell phone verification from speech recordings using sparse representation. *Digit. Signal Process. A Rev. J.* **62**, 125–136 (2017).
10. 8f5af5141eca9774abdd5b362f8d2518.pdf.
11. Qi, S., Huang, Z., Li, Y. & Shi, S. Audio recording device identification based on deep learning. *2016 IEEE Int. Conf. Signal Image Process. ICSIP 2016* 426–431 (2017) doi:10.1109/SIPROCESS.2016.7888298.
12. Gural, E. & Pazarci, M. Forensic audio authentication analysis technique of first or higher generation copies of analog magnetic audio tapes. *Med. Sci. J. Int. Med. J.* **1** (2018) doi:10.5455/medscience.2018.07.8764.
13. Boss, D., Gfroerer, S. & Neoustroev, N. A new tool for the visualization of magnetic features on audiotapes. *Speech, Lang. Law* **10**, 255–276 (2003).
14. Begault, D. R., Brustad, B. M., Stanley, A. M. & Associates, C. M. S. Tape analysis and authentication using multi-track recorders. *Audio Eng. Soc. 26th Int. Conf.* 1–7 (2005).
15. Steinebach, M. Watermarking-Based Digital Audio. 1001–1015 (2003).
16. Zakariah, M., Khan, M. K. & Malik, H. Digital multimedia audio forensics: past, present and future. *Multimed. Tools Appl.* **77**, 1009–1040 (2018).
17. Ahmed, R. A. Multimedia Protection, Authentication and Advanced Digital Watermarking Techniques: A Survey. *Int. J. Comput. Trends Technol.* **30**, 162–168 (2015).
18. Imran, M., Ali, Z., Bakhsh, S. T. & Akram, S. Blind Detection of Copy-Move Forgery in Digital Audio Forensics. *IEEE Access* **5**, 12843–12855 (2017).
19. Yang, R., Qu, Z. & Huang, J. Detecting digital audio forgeries by checking frame offsets. *MM Sec'08 Proc. 10th ACM Work. Multimed. Secur.* 21–26 (2008) doi:10.1145/1411328.1411334.
20. Maher, R. C. Lending an Ear in the Courtroom : Forensic Acoustics. **11**, (2015).
21. Maher, R. C. Intel. Multimedia Analysis for Security Appli. **282**, 127–144 (2010).
22. Foley, M., Lennon, J. J., Foley, M. & Lennon, J. J. JFK and dark tourism : A fascination with assassination JFK and Dark Tourism : a fascination with assassination. 37–41 (2007) doi:10.1080/13527259608722175.
23. HEAR NO EVIL.
24. AES E-Library » Analytical Framework for Digital Audio Authentication.
25. Fayyad-Kazan, H., Hejase, A., Moukadem, I. & Kassem-Moussa, S. Verifying the Audio Evidence to Assist Forensic Investigation. *Comput. Inf. Sci.* **14**, 25 (2021).

26. Pfeiffer, S. & Srinivasan, U. Scene Determination Using Auditive Segmentation Models of Edited Video. *Media Comput.* i. 105–129 (2002) doi:10.1007/978-1-4615-1119-9_6.
27. Braunschweiler, N. & Chen, L. Automatic detection of inhalation breath pauses for improved pause modelling in HMM-TTS. *8th ISCA Work. Speech Synth.* 1–6 (2013).
28. Pfeiffer, S. Pause concepts for audio segmentation at different semantic levels. 187 (2001) doi:10.1145/500170.500171.
29. Chen, J., Xiang, S., Huang, H. & Liu, W. Detecting and locating digital audio forgeries based on singularity analysis with wavelet packet. *Multimed. Tools Appl.* **75**, 2303–2325 (2016).
30. Gural, E. & Pazarc, M. A supporting method to detect manipulated zones in digitally edited audio files. *Med. Sci. | Int. Med.* i. *J.* **1** (2017) doi:10.5455/medscience.2017.06.8699.
31. US5918223.pdf.
32. Zmudzinski, S., Munir, B. & Steinebach, M. Digital audio authentication by robust feature embedding. *Media Watermarking, Secur. Forensics 2012* **8303**, 830301 (2012).
33. Hassan, N. A., Al-Mukhtar, F. S. & Ali, E. H. Encrypt Audio File using Speech Audio File As a key. *IOP Conf. Ser. Mater. Sci. Eng.* **928**, (2020).
34. Haitisma, J., Kalker, T. & Oostveen, J. Robust Audio Hashing for Content Identification. *Content Based Multimed. Index. 2001, Brescia, Italy, Sept. 2001* **4**, (2001).
35. Albahrani, E. A. A new audio encryption algorithm based on chaotic block cipher. *2017 Annu. Conf. New Trends Inf. Commun. Technol. Appl. NTICT 2017* 22–27 (2017) doi:10.1109/NTICT.2017.7976129.
36. Ghasemzadeh, A. & Esmaeili, E. A novel method in audio message encryption based on a mixture of chaos function. *Int. J. Speech Technol.* **20**, 829–837 (2017).
37. C., N. & B., S. Efficient Cryptographic key Generation from Fingerprint using Symmetric Hash Functions. *Res. Rev. Comput. Sci. Int. J.* **2**, 2557 (2011).
38. Bertolami.
39. Özer, H., Sankur, B., Memon, N. & Anarim, E. Perceptual audio hashing functions. *EURASIP J. Appl. Signal Processing* **2005**, 1780–1793 (2005).
40. Arefkhani, M. & Soryani, M. Malware clustering using image processing hashes. *Iran. Conf. Mach. Vis. Image Process. MVIP 2016-Febru*, 214–218 (2016).
41. Kivanç Mihçak, M. & Venkatesan, A. A perceptual audio hashing algorithm: A tool for robust audio identification and information hiding. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2137**, 51–65 (2001).
42. Paper, W. Digital Fingerprinting. *Digit. Fingerprinting* (2016) doi:10.1007/978-1-4939-6601-1.
43. Natarajan, M. & Makhdumi, G. Safeguarding the Digital Contents: Digital Watermarking. *DESIDOC J. Libr. Inf. Technol.* **29**, 29–35 (2009).
44. Yadav, U., Sharma, J. P., Sharma, D. & Sharma, P. K. Different Watermarking Techniques & its Applications: A Review. *Int. J. Sci. Eng. Res.* **5**, 1288–1294 (2014).
45. Klinger, K. & Snaveley, J. Advanced Techniques in Multimedia Watermarking : Image , Video and Audio Applications. *Inf. Sci. Ref.* 545 (2010).
46. Lalitha, N. V., Rao, C. S. & Jayasree, P. V. Y. Localization of Copy - Move Forgery in Speech Signals Through Watermarking Using DCT - QIM. **65**, 527–532 (2019).
47. Kajstura, M., Trawinska, A. & Hebenstreit, J. Application of the Electrical Network Frequency (ENF) Criterion: A case of a digital recording. *Forensic Sci. Int.* **155**, 165–171 (2005).
48. Grigoros, C. Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis. *Forensic Sci. Int.* **167**, 136–145 (2007).
49. Lv, Z., Hu, Y., Li, C. T. & Liu, B. B. Audio forensic authentication based on MOCC between ENF and reference signals. *2013 IEEE China Summit Int. Conf. Signal Inf. Process. ChinaSIP 2013 - Proc.* 427–431 (2013) doi:10.1109/ChinaSIP.2013.6625375.
50. Chuang, W. H., Garg, R. & Wu, M. Anti-forensics and countermeasures of electrical network frequency analysis. *IEEE Trans. Inf. Forensics Secur.* **8**, 2073–2088 (2013).
51. Nicolalde Rodríguez, D. P., Apolinário, J. A. & Biscainho, L. W. P. Audio authenticity: Detecting ENF discontinuity with high precision phase analysis. *IEEE Trans. Inf. Forensics Secur.* **5**, 534–543 (2010).
52. Savari, M., Abdul Wahab, A. W. & Anuar, N. B. High-performance combination method of electric network frequency and phase for audio forgery detection in battery-powered devices. *Forensic Sci. Int.* **266**, 427–439 (2016).
53. Chuang, W. H., Garg, R. & Wu, M. How secure are power network signature based time stamps? *Proc. ACM Conf. Comput. Commun. Secur.* 428–438 (2012) doi:10.1145/2382196.2382243.
54. Kristie L. Krebs. Naval Postgraduate. *Security* 79 (2019).
55. Gangwar, D. P. & Pathania, A. Authentication of Digital Audio Recording Using File'S Signature and Metadata Properties. i. *Int. J. Eng. Appl. Sci. Technol.* **5**, 162–165 (2020).
56. Hamdi, D., Iqbal, F., Baker, T. & Shah, B. Multimedia File Signature Analysis for Smartphone Forensics. *Proc. - 2016 9th Int. Conf. Dev. eSystems Eng. DeSE 2016* 130–137 (2017) doi:10.1109/DeSE.2016.22.
57. Wang, F., Li, C. & Tian, L. An algorithm of detecting audio copy-move forgery based on DCT and SVD. *Int. Conf. Commun. Technol. Proceedings, ICCT 2017-October*, 1652–1657 (2018).
58. Yan, Q., Yang, R. & Huang, J. Robust Copy-move Detection of Speech Recording Using Similarities of Pitch and Formant. 1.
59. https://keepnetlabs.com/blog/deepfake-statistics-and-trends Kratika Bhagtani, SPEECH FORENSICS USING MACHINE LEARNING
60. Bevinamarad, P. R. & Shirldonkar, M. S. Audio Forgery Detection Techniques: Present and Past Review. in *Proceedings of the 4th International Conference on Trends in Electronics and Informatics, ICOEI 2020* 613–618 (Institute of Electrical and Electronics Engineers Inc., 2020). doi:10.1109/ICOEI48184.2020.9143014.