Computer and Information Science; Vol. 14, No. 3; 2021
ISSN 1913-8989 E-ISSN 1913-8997
Published by Canadian Center of Science and Education

Verifying the Audio Evidence to Assist Forensic Investigation

Hasan Fayyad-Kazan[1], Ale Hejase[2], Imad Moukadem[3], Sondos Kassem-Moussa[4]

[1] Information Technology Department, Al Maaref University, Beirut, Lebanon

[2] AKSOB, Lebanese University, Beirut, Lebanon

[3] Departement of Computer Sciences, Al Maaref University, Beirut, Lebanon

[4] Department of Forensic Sciences, Lebanese University, Beirut, Lebanon

Correspondence: Hasan Fayyad-Kazan, Information Technology Department, Al Maaref University, Beirut, Lebanon.

Received: March 12, 2021 Accepted: June 8, 2021 Online Published: July 5, 2021

doi:10.5539/cis.v14n3p25 URL: https://doi.org/10.5539/cis.v14n3p25

Abstract

Audio forensics is a field in forensics that is used to authenticate, enhance, and analyze audio files to aid in solving different crime investigations. Audio as a forensic evidence must be enhanced and analyzed to be admissible in courts of law. But more importantly, it must be authenticated in order to prove that it is authentic and no manipulations were done to it. In this paper, an overview on audio forensics is presented, previous related work to this topic is shown, and methodologies for audio enhancement and authentication are explained along with audio tampering ways and signatures presentation.

Keywords: audio, authentication, enhancement, forensics, evidence, tampering

1. Introduction

Audio forensics is the field of forensic science related to the acquisition, analysis, and evaluation of sound recordings that may ultimately be presented as admissible evidence in a court of law. In the early 1960s, the Federal Bureau of Investigation (FBI) in the United States started developing experts in audio forensics to improve the speech intelligibility, enhancement and authentication of recorded files (Koenig, 1990). Imagine a specific scenario when a conversation is recorded between two people planning an illegal act, or even when a fugitive is contacted via telephone and the justice system needs to determine his/her location. The employment of audio forensics in those cases is the essential thing to do for solving them. Audio must be enhanced in order to make it clear so that its analysis would become easier to ultimately identify the persons in the first example mentioned before and locate the fugitive in the second example. Sometimes, such audio files or recordings might be manipulated and tampered with to show false matters as true. Here comes the importance of authenticating audio files to verify their integrity and thus neglect the possibility of tampering so that the content of the audio file becomes reliable in the legal field. In this paper, an overview of audio as a forensic evidence is presented. It starts with a literature review presenting work done by different researchers on this topic. The strategy followed to enhance and analyze audio is explained, and then audio tampering is introduced. Basic methods of manipulation and the signatures such manipulation leaves are talked about. Some well-known and available software for manipulation is shown and finally methods of authentication are presented.

2. Literature Review

When audio files became important for forensic investigations leading to the emergence of a new field in forensics: audio forensics, it was inevitable for researchers to shed light on this field and explain all its aspects. For this purpose, Maher ((Maher, 2009); and (Maher, 2010)) gave an overview on audio forensics as he explained the definition of this field and how it started, its relevance to law, and the ways to work on audio files. He also presented the practices in the field and explained the methodology for interpreting the authenticity of an audio evidence, and the enhancement of audio files. Concerning audio enhancement, Ikram & Malik (Ikram & Malik, 2010) presented a new audio forensics method for speech enhancement based on background noise in the audio signals. A two-step framework was proposed to estimate the background noise with minimal speech leakage signal. In the first step, spectral estimation based on geometric transformation has been used to obtain the initial noise estimation, and the second stage exploits higher harmonic structure characterization of speech signal to remove speech signal from the initial noise estimate. A correlation-based similarity measure was then applied to determine the integrity of speech signal. After evaluating it in various environments, the results showed that this method performed better than the existing speech enhancement algorithms according to the authors' claims.

As for authentication, Zakariah & Malik (Zakariah et al., 2018) dealt with the importance of authenticating recorded audio files especially that many audio tampering software became easily available on the market. According to their belief that determining the authenticity of an audio evidence requires several types of observations – examiners need to perform visual, physical, electrical and acoustic tests – they tried listing all the possible techniques used to successfully detect audio tampering. They categorized the analysis of audio files in the process of detection into container-based analysis that deals with the file structure and metadata of the audio file (such as timestamp and file format), and content-based analysis which handles the actual bits and bytes of the audio file (such as speaker identity and speech transcript). One of the authentication solutions for audio recordings was proposed by Nita & Ciobanu (Nita & Ciobanu, 2018); they provided a watermark solution. A control signal, the TIC-TAC signal, was the case. It is visible and audible on the actual recording, yet ignored by the listener, and its insertion process was performed simultaneously with the speech/audio recording, therefore the integrity of the audio signal was not altered. It was incorporated in an integrity check algorithm and then preliminary tests were performed. The results showed a 100% detection rate for edited recordings by means of deleting/inserting audio fragments longer than 5 ms and 0% false alarm rate for unedited recordings. Moreover, Renza, Ballesteros, & Lemus (Renza et al., 2018), aiming to verify if an audio proof has been tampered i.e. to guarantee the chain of custody, and to locate the segments where the signal was modified, the authors worked on a fragile watermarking method for digital audio authenticity. They embedded a text encoded by orthogonal variable spreading factor (OVSF) codes and spread into the entire signal using automatic adjustment. Then different tests were performed to check for tampering detection against different attacks. The results demonstrated that even if a small number of samples is modified, the system correctly labels the audio proof as manipulated and even locates the start and end of the manipulation. Another solution for authentication was proposed by Bhangale & Patole (Bhangale & Patole, 2019) , the method offered by the authors is based on the reverberation component embedded in the audio recording. The statistical features of the Mel-frequency cepstral coefficient (MFCC) and the decay rate distribution were used for tampering detection. The results showed that the MFCC features of the reverberant component show a significant difference compared to DRD features to edition in the recording for both insertions and deletions.

Actually, most of the times, the work done by researchers on audio forensics addresses the problem of audio tampering. Ross, Banerjee, & Chowdhury (Ross et al., 2020) explained content- based audio manipulation techniques and the corresponding forensic detection methods, due to their direct impact on the biometric utility of the audio data that might affect the security in smart cities. They clarified how Digital audio manipulation can be deployed at multiple stages of audio production from audio recording such as tampering the sensor to audio editing where certain audio editing techniques such as butt splicing and copy-move forgery can be applied, thereby, compromising audio integrity. They stated that manipulation can occur by audio mixing like when audio from multiple sources is mixed to form one combined audio track (maybe adding noise), or even occur via audio mastering by subjecting the audio file to a series of subtle audio processes such as compression. Different researchers tended to find solutions for detecting this manipulation. For example, Bianchi et al. (Bianchi et al., 2014) proposed a method to test for the presence of double compression artifacts in a MP3 audio file and to find its location if present as a way to uncover possibly tampered parts. They worked on an algorithm based on a simple statistical feature that measures the effect of double compression thus allowing to decide whether a MP3 file is singly or doubly compressed. Experiments done by the researchers confirm the good performance of the proposed scheme, thus what they proposed offers a new tool for the forensic analysis of MP3 audio tracks. Also, Korycki (Korycki, 2014) addressed the problem of tampering detection in compressed audio files. He discussed new methods that can be used for authenticity analysis of digital recordings. These methods consisting in application of statistical features of MDCT coefficients as well as investigation of MP3 file structure allow for successful detection of forgeries by forensic experts. The effectiveness of tampering detection algorithms was tested on a predefined large music database consisting of nearly one million of compressed audio files. The achieved results contribute to the development of scientific tools for forensic audio analysis. The value of detection ratio depended on the particular case, however, in most cases varied between 90% and 100% thus providing a robust assistance in authenticity investigation process. Moreover, Reis, Da Costa, Miranda, & Galdo (Reis et al., 2017) proposed a technique to detect adulterations in digital audio recordings through exploiting abnormal variations in the Electrical Network Frequency (ENF) signal eventually embedded in a questioned audio recording. These abnormal variations are caused by abrupt phase discontinuities due to insertions and suppressions of audio snippets during the tampering task. First, they proposed an ESPRIT-Hilbert ENF estimator in conjunction with an outlier detector based on the sample kurtosis of the estimated ENF. Then, they used the computed kurtosis as input for a Support Vector Machine (SVM) classifier to indicate the presence of tampering. The proposed scheme, designated as SPHINS, outperformed – according to the authors claims - related previous tampering detection approaches in the conducted tests.

Furthermore, Imran, Bakhsh, & Akram (Imran et al., 2017) focused on blind forgery detection which is challenging because it does not embed a watermark or signature in an audio that is unknown in most of the real-life scenarios. Therefore, they proposed a method that performs blind detection of forgery by inspecting the content of an audio recording. It detects the words through the voice activity detection (VAD) module and computes the histograms by one-dimensional local binary pattern operator application. The results obtained had an accuracy of 96.56% with this method.

A different approach was explained by Lin & Kang (Lin & Kang, 2017), the detection of audio tampering was done by developing an effective detector for multiple types of audio forgeries. Based on the fact that tampering with an audio leads to anomalous variations of the underlying ENF signal, a wavelet filter to the extracted ENF was applied, followed by an autoregressive modeling of the detail ENF signal. A supervised classification framework which exploits the statistical autoregressive features was introduced to identify whether an audio is a tampered one or not. According to their results, Lin & Kang claimed that their proposed method significantly outperformed the state-of-art methods in the context where moderate or high levels of noise are present.

In (Liu & Lu, 2017), Liu and Lu proposed a fast method to detect copy-move forgery in digital audio. At first, the audio was segmented into syllables, then discrete Fourier transform (DFT) on each audio segment was applied for similarity computation. An additional sorting step was done to reduce comparisons and thus time. Finally, unlike other mainstream algorithms which compare the similarity between every two segments of the audio file, they just compared one audio segment with some adjacent ones in the sorted list of audio segments. The experimental results demonstrated that the proposed method is more effective, better, and faster than other methods regarding time consumption.

Further solutions were tested. In fact, Khan, Zakariah, Malik, & Choo (Khan et al., 2018) presented a digital audio forensic data-set, designed to facilitate evaluation of audio forensic algorithms. They also evaluated; based on this data-set, the performance of three existing audio forensics algorithms; Gabor filter coefficients and MFCC and PLP coefficients. Experimental results showed that the selected approaches achieved promising results. Also, Wang, Yuan, Wang, & Unoki (Wang et al., 2019), working on speech tampering, they proposed a tampering detection scheme for speech signals based on spectral manipulations since spectral envelope and formants are important indicators of tampering since tampering of a speech will unavoidably modify the shape of the spectral envelope and the locations/magnitudes of the formants. The Robust Principal Component Analysis (RPCA) was employed to decompose the original speech into sparse component and low-rank component then the sparse component was selected for information hiding/embedding via spectral manipulations to ultimately obtain a manipulated spectral envelope. The evaluation results suggested that the method proposed could not only satisfy inaudibility but also robustness. Furthermore, it was also fragile against tampering and capable of detecting tampering in both the temporal tampering and acoustic feature-based tampering with reasonable accuracy.

The research in audio forensics domain even went further to steganalysis. In (Luo et al., 2018), Luo, Li, Yan, Yang, & Huang designed an improved audio steganalytic feature set from both the time and Mel-frequency domains and then combined these features together to detect some typical steganogrpahy in the time domain and various typical audio operations. These features were very powerful in both steganalytic and forensic applications. Thus, the results showed that the proposed feature set is able to capture the modifications introduced by different audio operations.

3. Audio

3.1 Audio as Forensic Evidence

The important role that video-recordings play to assist the forensic investigations does not minify the contribution of audio recordings in the same role. Though an audio lacks the visual part supplied by a video, it can provide the law enforcement and the investigators with significant clues and missing proofs to solve, analyze and judge the case under investigation.

An audio is any sound that can be heard by the human ear within the acoustic range and most commonly exist in the mp3 audio file format.

Figure 1. Audio Recording (*Best Ways to Record Audio on MacBook Pro*, 2016)

The impact of audio recording in facilitating the interpretation of large-scale crimes and events exponentially expanded over the past decades. An example of how audio recording can act as an efficient link to complete the puzzle of forensic investigation is the case of Dr. Richard Kimble in the movie The Fugitive (National Forensic Service, 2012). When he was running from the police, he called Chicago Police Department to proclaim his innocence. Though he ended the call before the police could trace it, Dr. Richard was not aware that the sound of the L-train was so loud and can be obviously heard in the phone call. That little detail in the background of the recording which he did not anticipate helped the police to track him and find his location in Chicago by analyzing the audio of the phone call (National Forensic Service, 2012). That is one of the instances in which audio recording renders a valuable evidence in the prosecution of any tough case.

Figure 2. Dr. Richard from The Fugitive movie (*Fugitive*, 1993)

In addition to the previous example, an audio recording when rightly collected, enhanced and analyzed can potentially serve a lot of information in many other ways. One of these include identification of particular weapon types. When an audio is recorded in real-time shooting during a crime, fight or war, it can help through weapon signature analysis methods to detect the type used and then establish the order of firing as well. Likewise, an audio can specify the make and model of vehicle used in a certain crime or event via vehicle signature analysis of an available audio recorded at the time of occurrence. Moreover, an audio is very useful throughout the investigation of air accidents whereby the cause of the crash can be established by the analysis of aircraft engine noise.

Despite all the above paramount roles of audio evidences, foremost its primary advantage is in speech recognition and talker identification. This is done by analyzing phone calls between criminals, voicemails recordings containing threats or planning for a crime, telephone answering machines, police /911 calls and audio files containing immoral content. These audio sources and many others help the investigators to identify the suspects and collect other useful information like their number and location.

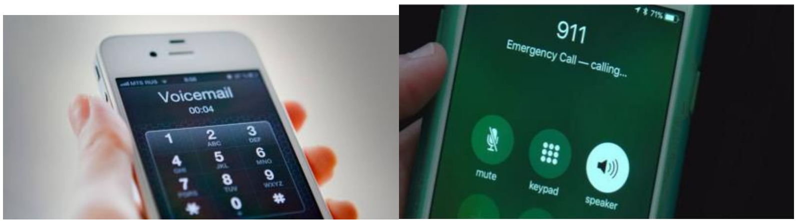


Figure 3. Audio Evidence Types (Andersen, 2018)

Various types of audio recordings can be analyzed to serve as a potential forensic evidence. These include:

- Recordings made on smart phones
- Answering machines' recordings
- Voice mails
- Audio content of videos
- Digital handheld recorders
- Cassette tapes
- 911 police calls
- Courtroom recordings

Regardless in any format an audio evidence is, its authenticity must be established before submitting it to the court to reveal if any tampering method has been applied.

3.2 Audio Enhancement and Analysis

For an audio recording to bring out specific aspects of an event and facts concerning a particular incident its content must be enhanced prior to any interpretation. This process is important to clarify the audio recording content in order to be intelligible to the investigator, attorney and judges in the court. The techniques of forensic audio enhancement must be applied aiming at reducing the unwanted noise and increasing the desired sounds in order to improve the overall quality of the audio content. In this way the enhanced audio will provide an accurate representation of the events.

Similar to the case of video (Hasan et al., 2021), audio enhancement and analysis is specific to each case and varies with the type of recording device and whether an audio is digital or analog. Despite the uniqueness of each audio recording, there are basic steps that must be followed in the process of audio enhancement and analysis.

3.2.1 Critical Listening

After creating a duplicate to preserve the original copy of the audio, the early step of audio enhancement and analysis is critical listening whereby each section of the audio is identified. The purpose of this step is to detect the depth of enhancement required by each section of the audio content based on the recorded characteristics including signal to noise ratios and frequency range (Maher, 2009).

For example, an audio recording might contain three different portions that require different processing. The first one represents an audible conversation between two people with traffic jam noise from a distance. In a second section the speech clarity decreases due to significant amount of background noise. The third section contains no background noise but rather the people were talking quietly with low voices. Each of the three sections requires specific type of processing and enhancement that is identified by an expert in critical listening.

3.2.2 Noise Reduction

As the main purpose of audio enhancement is to remove unwanted sounds, this step is substantial. The techniques applied for this process are the most effective in the context of audio enhancement. Usually, the two common methods used for this aim are compression and echo cancelation.

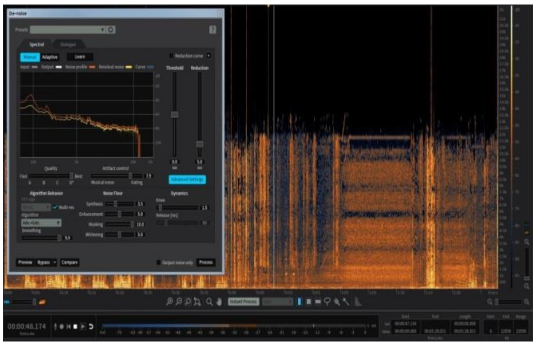
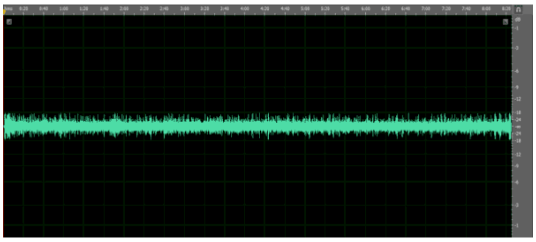


Figure 4. Noise Reduction Step (Audio Forensic Expert, n.d.-a)

Compression is applied to boost the faint sounds and reduce the background noise in the audio content. This is done by leveling the signal so that the dynamic range of the material is reduced making soft sounds more apparent (National Forensic Service, 2012). On the other hand-echo cancelation is applied for decreasing the reflection of noise and reverberance of space that reduce the clarity of audio content. Hence, these techniques ameliorate the intelligibility of the recording and hinder the presence of noise that might become an issue in further processing.



Waveform of a recording made at low volume with significantly loud ambient noise that is masking the speech content of the recording (National Forensic Service, 2012).

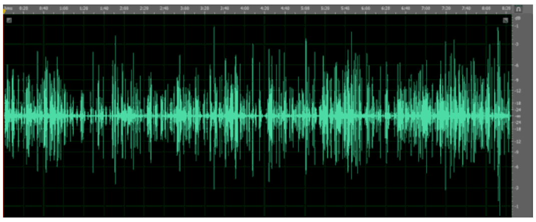


Figure 5. Audio Recording before and after enhancement step (National Forensic Service, 2012)

The same recording after enhancement. The noise is attenuated and the volume of the speech is increased (National Forensic Service, 2012).

3.2.3 Frequency Equalization

Now after the audio has become less noisy, frequency equalization is applied for additional help in boosting the desired sounds. Using specific and highly accurate equalizers to cut specific ranges of frequency bands the speech content becomes clearer. Usually, the frequency bands containing speech content ranging between 200Hz and 5000Hz are isolated and amplified (National Forensic Service, 2012). Whereas the frequencies corresponding to loud background noises and unwanted overtones will be reduced making them less noticeable. Every specific audio content needs specific number of processors to achieve the desired results.

3.2.4 Forensic Transcript

Forensic audio experts demand the formation of forensic transcript in combination with critical listening and frequency analysis in the process of audio enhancement. A forensic transcript is the scientific observation of words under controlled conditions (*Forensic Audio Transcription*, n.d.). It is important for verification of spoken dialogue that occur in the recording as the latter is the main wanted sounds to be retrieved. Forensic audio transcription adds certainty to the enhancement and analysis of the audio as it appends a signed scientific document that is sworn by the expert witness as to the words that are spoken ending by:

‘Based on the pains and penalties of perjury, I testify that the above forensic transcript is an accurate representation of the dialogue (or events) spoken during the time of this recording so help me’ then we sign it.' (Forensic Audio Transcription, n.d.)

As the forensic transcript produced by an expert who is able to conclude to a reasonable degree the unintelligible words that were spoken, it becomes an essential part that must accompany other methods of audio enhancement and analysis.

3.3 Audio Tampering

"Words fade but voices remain”. Yet, the widely spread forgery techniques can turn the rigorous evidence that an audio recording supply to an unreliable one. Currently there are hundreds of software that are highly qualified in manipulating any audio recording at ease and in a very short time. In spite of the diverse tampering techniques that can alter the recording in various ways, all of them are based on four methods of falsification.

3.3.1 Basic Methods of Audio Manipulation

- Deletion: stopping the tape to remove wanted words or sounds by over-recording unwanted areas.
- Synthesis: using artificial means and specific algorithms to add words or voices to the tape.
- Obscuration: masking certain waveform patterns to mix voices or sounds which will be reflected in the recording as sudden stops and starts in inappropriate places.
- Transformation: changing the content of the recording by altering the arrangement of words.

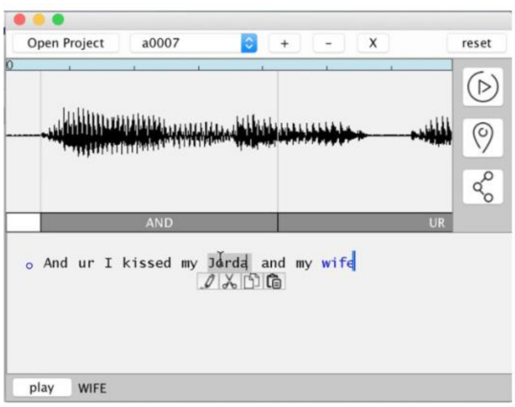
Though the four basic types of tampering have potent capability of manipulating the audio content to any desired form, they leave detectable magnetic signature in the tape that can be identified through audio inspection by an expert. Some of the electromechanical indications - signatures - left by tampering software include gaps, fades, transients and equipment sounds.

- Gaps: represent segments in the recording containing unexplained content's changes like sudden silence, humming or buzzing.
- Fades: represent gradual loss of the volume of sounds (when complete inaudibility occurs fades will become gaps).
- Transients: represent abrupt and short sounds like pops and clicks.
- Equipment sounds: represent inconsistencies in the audio content caused by the recording equipment itself like whistles or varying pitches.

3.3.2 Voice Morphing and Cloning Software

Adobe ProjectVoco

Among all the audio editing and content altering software, Adobe ProjectVoco is one of the most interesting to shed the light on. As the well-known Photoshop applications tweak digital images, ProjectVoco affects the audio in the same way. The codenamed software ProjectVoco is designed to be one of the most potent audio editing applications. Beyond the familiar features of standard speech editing and noise cancelation that ProjectVoco provides, what makes it a subject of attention is yet so unique. The standout feature of ProjectVoco is the capability to generate new words in the recording using the speaker's recorded voice (Lardinois, 2016). This means that it is able to add words to the audio file that were not originally found before. In fact the software targets at least 20 minutes of the desired recorded speech (Vincent, 2017). By understanding the makeup of a person's voice, ProjectVoco is able to replicate it generating sound-alike voice saying words that were not said in the original speech. The insertion of newly spoken words is activated by simply typing the desired words and the algorithms of ProjectVoco does the rest of the process resulting in a modified content of the audio recording via inserting unsaid words into the voiceover.



And ur I kissed my Jorda and my wife

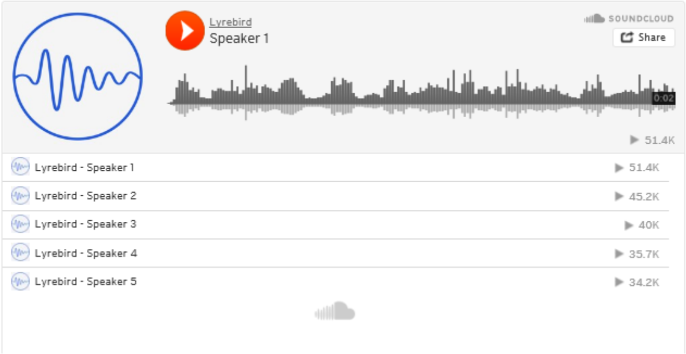
Adobe

Figure 6. Adobe ProjectVoco Algorithm (Lardinois, 2016)

ProjectVoco, Wavenet which is another software for generating realistic human-like voices and other similar software appear to be very beneficial for audio engineers when working on clips, TV shows, dialogues and news where people may desire to add or change a word or phrase in their speech. Though there are several ethical sorts of applications facilitated by employing ProjectVoco, its unlawful dark side is really worrisome. Imagine if someone records an audio by your own voice saying words that you have never said. The idea itself is so creepy. What if the audio contains a speech mentioning people names that you have not ever known? What if the audio manipulates the time you said that you were planning to go to a certain event? What if the audio contains indecent and abusive verbalism expressed by your voice? What if any of these and other manipulated forms of speech by ProjectVoco are submitted and considered a genuine evidence in the court! Since ProjectVoco has the potential to falsify entire sentences using a person's voice in a near-perfect replication of the talker's, awareness must be raised to general publicity and law enforcements regarding the doctored audio clips that could be generated by such malicious technique.

Lyrebird

The outcomes generated by audio editing software like ProjectVoco are really impressive but nothing compared to these of Lyrebird. Lyrebird is a voice cloning software unveiled by artificial intelligence (Vincent, 2017). It consists of a set of algorithms that are able to create incredibly human-like synthesized voices by listening to a very short segment of the person's recorded audio. Unlike ProjectVoco that requires at least 20 minutes of the sample audio to alter the voice content, Lyrebird needs a very small segment of the recording cutting this requirement down to just one minute (Vincent, 2017). This means that using only one minute of sample audio, Lyrebird is able to create any voice which is a process that was seen impossible a few years ago. To add more truthfulness to the voice output created, Lyrebird algorithms are able to infuse emotions with the speech it creates as making the target person's voice sound sympathetic, cheery, angry, or stressed out.



Lyrebird - Speaker 1	▶ 51.4K
Lyrebird - Speaker 2	▶ 45.2K
Lyrebird - Speaker 3	▶ 40K
Lyrebird - Speaker 4	▶ 35.7K
Lyrebird - Speaker 5	▶ 34.2K

Figure 7. Lyrebird Software (Vincent, 2017)

Similar to ProjectVoco, Lyrebird also has an advantageous role when employed in the right places. With the help of this advanced software, it is now feasible to read audio books with famous voices and synthesize speeches for people with disabilities or for animation movies (Vincent, 2017). These and many other uses are identified for this powerful tool that can create thousands of sentences in less than seconds after the voiceprint is already detected. Besides, there are other troubling uses of Lyrebird as well. Since the results generated are indistinguishable from human speech, it is possible now to produce a whole audio content, not just a word, by someone's voice who has not spoken what is said in the recording. Through this synthetic voice generator one can steal the identity of anyone and record an audio by his/her voice stating incidents that did not happen, confessing fake truths or replying to questions that he/she has not even answered. These dangerous consequences are very misleading especially when such recordings are used as forensic evidences in the courts. The veracity of audio recordings is brought to question due to the perfectly voice synthesizing software Lyrebird with hardly detectable tampered outcomes. This problem is solved by releasing this technology publicly so that it becomes available to anyone and everyone is aware of its results as they are aware now of the manipulated photos by Photoshop. By this way the forensic investigations should go deeper when analyzing an audio recording subjoining it with other evidences to testify its reliability and originality.

3.4 Audio Authentication

The verification of audio recording originality and integrity is a necessity to neglect the possibility of tampering and make its content reliable in the legal field. Because manipulation of the recording has become an easy task, evaluating the authenticity of the audio is done before it is considered a valuable evidence furnishing credible facts. Several methods are developed to check if any malicious or accidental tampering has been applied to the audio under investigation.

3.4.1 Basic Methods of Audio Authentication:

1) Electronic Measurement

The main concept of electronic measurement is to check the frequency ranges of the voices in the audio. If any sudden shift in the available frequency to a smaller or larger one occurs, this will be a sign that certain alteration was done. In a similar way any sudden unexplained changes in the background noise or the sudden appearance of a new source of noise are also signs of manipulation. The tools used for measuring this parameter are spectrograms, level meters and frequency analysis panels that are helpful is observing minute changes.

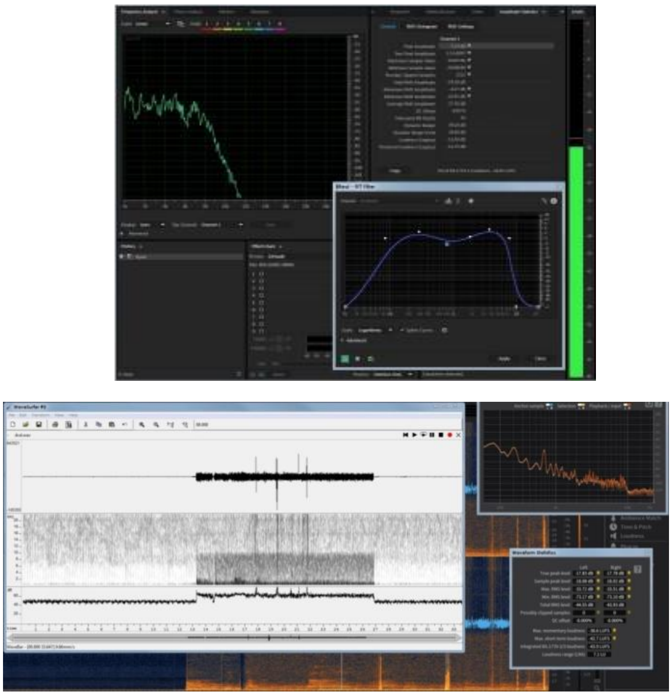


Figure 8. Electronic Measurement Step (*Authentication of Digital Audio Recordings*, 2014)

2) Visual Inspection

Visual inspection goes hand in hand with electronic measurement whereby analysis of frequency information and physical wave property are done. The interpretation of the waveform is a critical step in audio authentication. Since waveforms are smooth and continuous, any sudden break in the recording waveforms indicate that a certain edit has been done. Another example of how waveforms' characteristics sign for an alteration is the case when a waveform is inverted. Waveforms sudden breaks (Maher, 2009), inversions and other similar signs of edits are visualized in a full frequency spectrum shown by the spectrogram.

3) Analyzing Metadata in Digital Audio Evidence

This method is exclusively applied on digital audio recordings but not for analog ones like standard audio cassette. When an audio is digitally recorded, an extra information can be interpreted while performing audio authentication. Information about how the recording is made and the type of equipment used to create the audio are revealed in what so called metadata of the digital audio recording. Any alteration in the audio content leaves specific footprints in the recording hexadecimal information indicating that the audio was loaded into a software program for audio editing. When analyzing metadata of a certain audio an exemplar - a recording made in the same kind of recorder and in close environment of the original - must be established to check any inconsistencies in the data signing for tampering.

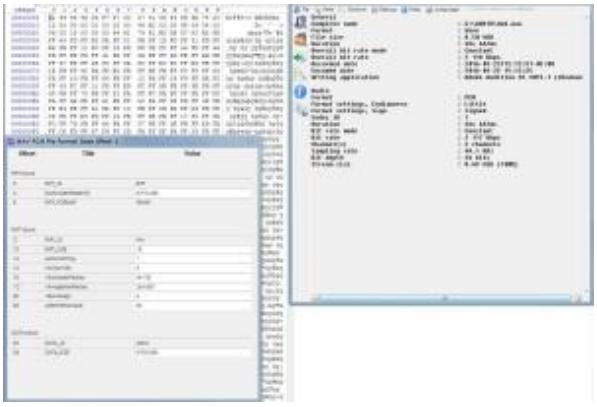


Figure 9. Metadata Analysis (Audio Forensic Expert, n.d.-b)

3.4.2 ACUSTEK for Audio Authenticity

Figure 10. ACUSTEK designs and manufactures unique Professional Technical Surveillance, Countermeasure Solutions and Audio Forensic Solutions (*ACUSTEK*, 2009)

Acu-expert has released its powerful software, ACUSTEK, for audio tampering detection and authenticity analysis. It is made of a set of instruments that are developed to expose the traces left by audio manipulating techniques (*ACUSTEK*, 2009). What make it really beneficial are its various useful characteristics such as producing repeatable results, being fast and easy to use, supporting all audio formats, and having scientific fullness since all signals authenticity approaches are covered. ACUSTEK has solved the issue of authenticity analysis and so that it becomes accredited by many forensic audio labs and investigation centers on the world (*ACUSTEK*, 2009). Some of the diverse features of ACUSTEK include:

- Revealing traces of editing via audio software functions and compression types representing results in levelograms (ACUSTEK, 2009).

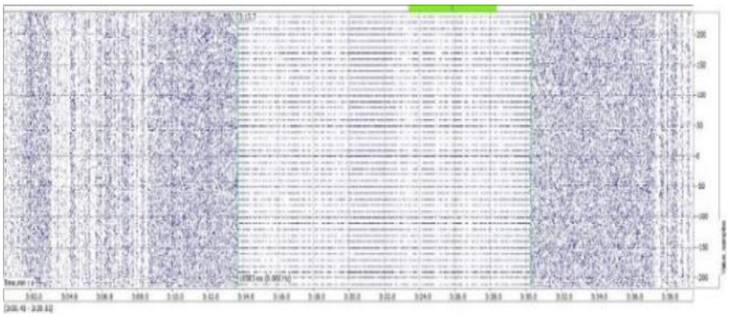
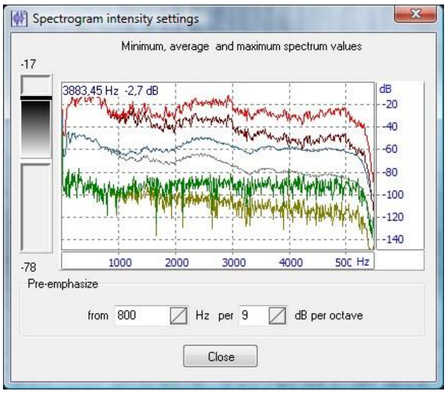


Figure 11. A levelogram (Levelogram, n.d.)



Spectrogram intensity settings

Minimum, average and maximum spectrum values

from 800 Hz per 9 dB per octave

Figure 2. A view normalizer (View Normalizer, n.d.)

- Revealing traces left by unexplained start/stop modes that sign for intentional or accidental alterations.
- Revealing matching signal fragments repeated several times to fraud the audio evidence.
- Identifying the speaker’s identity through voice biometrics and unique vocal elements of the speaker’s voice after significantly improving the quality of the speech.

Analyzing audio compression type to detect any mismatch between the audio recording and the recording device via a samples order specification (*ACUSTEK*, 2009).

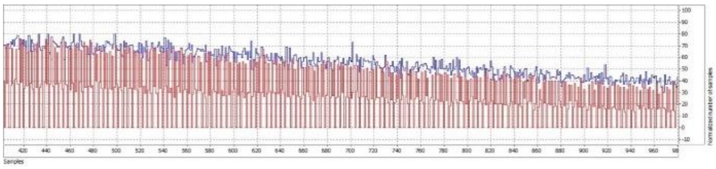


Figure 13. A samples order (Samples Order, n.d.)

Further specifications are valid in ACUSTEK to help reveal any basic audio forgery. Though ACUSTEK does not provide a definitive answer when dealing with cases of tampering like audio cloning as previously discussed, it helps in the following processing once the audio recording is considered doubtful. Regarding other common tampering methods, ACUSTEK is the fundamental solution for unmasking any audio manipulation. Forensic audio experts and researchers are looking forward to develop more advanced technological procedures to depend on for testing the authenticity and reliability of any audio recording.

4. Conclusion

Audio forensics is essential for investigations as most of our speech and conversations nowadays can be recorded in a way or another. These evidences are adapted in courts if well authenticated, enhanced, and analyzed. However, their authentication is harder now that tampering is more sophisticated and easier than ever due to the fact that currently available technologies allow doing so. Therefore, solutions must be found and worked on to improve tools that can be used by audio forensic experts to easily evaluate different audio files. In this paper, work done by many researchers is presented to show different methods proposed to authenticate audio files. Indeed, there is a need for further studies in order to come up with a solution that can level up to the developed manipulation methods.

References

ACUSTEK. (2009). https://www.acustek.com/en/

Andersen, R. (2018). *May I Leave a Voicemail Message? (The Conundrum Continues)*. Retrieved from https://www.ontariosystems.com/may-i-leave-a-voicemail-message/

Audio Forensic Expert. (n.d.-a). *Forensic Audio Enhancement*. Retrieved from https://www.audioforensicexpert.com/forensic-audio-enhancement

Audio Forensic Expert. (n.d.-b). *What is Audio Forensics*. Retrieved from https://www.audioforensicexpert.com/what-is-audio-forensics/

Authentication of Digital Audio Recordings. (2014). Retrieved from https://www.audioforensicexpert.com/authentication-of-digital-audio-recordings/

Best Ways to Record Audio on MacBook Pro. (2016). Retrieved from https://showmore.com/record-audio-on-macbook-pro.html

Bhangale, T., & Patole, R. (2019). Tampering detection in digital audio recording based on statistical reverberation features. In *Advances in Intelligent Systems and Computing* (Vol. 900). Springer Singapore. https://doi.org/10.1007/978-981-13-3600-3_55

Bianchi, T., De Rosa, A., Fontani, M., Rocciolo, G., & Piva, A. (2014). Detection and localization of double compression in MP3 audio tracks. *Eurasip Journal on Information Security*, *2014,* 1-14. https://doi.org/10.1186/1687-417x-2014-10

Forensic Audio Transcription. (n.d.). Retrieved from http://www.forensictranscript.com/services.html

Fugitive. (1993). Retrieved from https://www.empireonline.com/movies/reviews/fugitive-review/

Hasan, F. K., Hejase, H., Moukadem, I., Rkein, H., & Kobeissi, K. (2021). Anti- Forensics: The Tampering of Videos. *International Journal of Forensic Sciences*, *6*(1), 1-12. R https://doi.org/10.23880/ijfsc-16000219

Ikram, S., & Malik, H. (2010). *DIGITAL AUDIO FORENSICS USING BACKGROUND NOISE Sohaib*. 106-110. https://doi.org/10.1109/ICME.2010.5582981

Imran, M., Ali, Z., Bakhsh, S. T., & Akram, S. (2017). Blind Detection of Copy-Move Forgery in Digital Audio Forensics. *IEEE Access*, *5*(c), 12843-12855. https://doi.org/10.1109/ACCESS.2017.2717842

Khan, M. K., Zakariah, M., Malik, H., & Choo, K. K. R. (2018). A novel audio forensic data-set for digital multimedia forensics. *Australian Journal of Forensic Sciences*, *50*(5), 525-542. https://doi.org/10.1080/00450618.2017.1296186

Koenig, B. E. (1990). Authentication of Forensic Audio Recordings. *Journal of The Audio Engineering Society*, *38*, 3-33.

Korycki, R. (2014). Authenticity examination of compressed audio recordings using detection of multiple compression and encoders' identification. *Forensic Science International*, *238,* 33-46. https://doi.org/10.1016/j.forsciint.2014.02.008

Lardinois, F. (2016). *Adobe's Project VoCo lets you edit speech as easily as text*. Retrieved from https://techcrunch.com/2016/11/03/adobes-project-voco-lets-you-edit-speech-as-easily-as-text/

Levelogram. (n.d.). Retrieved from https://acustek.com/images/stories/equipment/tde/04 Levelogram.jpg

Lin, X., & Kang, X. (2017). Supervised audio tampering detection using an autoregressive model. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2142-2146. https://doi.org/10.1109/ICASSP.2017.7952535

Liu, Z., & Lu, W. (2017). Fast Copy-Move Detection of Digital Audio. *Proceedings - 2017 IEEE 2nd International Conference on Data Science in Cyberspace, DSC 2017*, 625-629. https://doi.org/10.1109/DSC.2017.11

Luo, W., Li, H., Yan, Q., Yang, R., & Huang, J. (2018). Improved audio steganalytic feature and its applications in audio forensics. *ACM Transactions on Multimedia Computing, Communications and Applications*, *14*(2). https://doi.org/10.1145/3190575

Maher, R. C. (2009). Audio forensic examination. *IEEE Signal Processing Magazine*, *26*(2), 84-94. https://doi.org/10.1109/MSP.2008.931080

Maher, R. C. (2010). Overview of audio forensics. *Studies in Computational Intelligence*, *282*, 127-144. https://doi.org/10.1007/978-3-642-11756-5_6

National Forensic Service. (2012). *A Simplified Guide To Forensic Audio and Video Analysis*. 19.

Nita, V. A., & Ciobanu, A. (2018). Tic-Tac, Forgery Time Has Run-Up! Live Acoustic Watermarking for Integrity Check in Forensic Applications. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2018-April*, 1977-1981. https://doi.org/10.1109/ICASSP.2018.8461538

Reis, P. M. G. I., Da Costa, J. P. C. L., Miranda, R. K., & Del Galdo, G. (2017). ESPRIT-Hilbert-Based Audio Tampering Detection with SVM Classifier for Forensic Analysis via Electrical Network Frequency. *IEEE Transactions on Information Forensics and Security*, *12*(4), 853-864. https://doi.org/10.1109/TIFS.2016.2636095

Renza, D., Ballesteros L., D. M., & Lemus, C. (2018). Authenticity verification of audio signals based on fragile watermarking for audio forensics. *Expert Systems with Applications*, *91*, 211-222. https://doi.org/10.1016/j.eswa.2017.09.003

Ross, A., Banerjee, S., & Chowdhury, A. (2020). Security in smart cities: A brief review of digital forensic schemes for biometric data. *Pattern Recognition Letters*, *138*, 346-354. https://doi.org/10.1016/j.patrec.2020.07.009

Samples Order. (n.d.). Retrieved from https://acustek.com/images/stories/equipment/tde/03 SamplesOrder.jpg

View Normalizer. (n.d.). Retrieved from https://acustek.com/images/stories/equipment/tde/10ViewNormalizer.jpg

Vincent, J. (2017). *Lyrebird claims it can recreate any voice using just one minute of sample audio*. Retrieved from https://www.theverge.com/2017/4/24/15406882/ai-voice-synthesis-copy-human-speech-lyrebird

Wang, S., Yuan, W., Wang, J., & Unoki, M. (2019). Detection of speech tampering using sparse representations and spectral manipulations based information hiding. *Speech Communication*, *112*(May), 1-14. https://doi.org/10.1016/j.specom.2019.06.004

Zakariah, M., Khan, M. K., & Malik, H. (2018). Digital multimedia audio forensics: past, present and future. *Multimedia Tools and Applications*, *77*(1), 1009-1040. https://doi.org/10.1007/s11042-016-4277-2

Copyrights