

PRAAT Software; Utilization of Computerized Approach for Determination of Variation Present in Recorded Audios from Distinct Sources

Smriti Rawat¹, Tasbiha Khan², Dr Amit Chauhan³,

Graduation^{1,2}, Amity Institute of Forensic Sciences, Amity University, Sec-125 Noida, Uttar Pradesh, INDIA-201313.

Assistant Professor³, Amity Institute of Forensic Sciences, Amity University, Sec-125 Noida, Uttar Pradesh, INDIA-201313

Abstract

In Forensic Field, various sources of evidences play a vital role to establish a link between the crime, preparator and the victim. These types of evidences which might be in physical form or material quality, any tangible article, biological form of evidences or chemical etc. Physical evidence is any evidence that have physical or material quality, a tangible article no matter whether microscopic or macroscopic. It encompasses all objects, living or inanimate; solid-, liquid, or gas. Among these physical evidences i.e. fingerprints, palm prints hair strands, blood etc. there is an important class of physical evidences which is audio evidences. Audio forensics involves acquisition, analysis, and evaluation of audio recordings that may eventually be presented as admissible evidence in the court of law. These audio recordings may turn out to be very important evidences in the field of forensic science. They can be encountered most frequently in the cases of ransom calls, cybercrimes, lost and found cases, etc. They may be available from the emergency call helpline, telephone answering machines, voicemail recording, cell phones and computer files. This study was carried out to analyze the authenticity of recoded audios of an individual by various modes of recordings. As a resultant of this study, it was observed that distinct mode of recordings affects the quality of recoding required a repair before the conclusion. This study can be helpful to determine the time duration of recorded audio based on intensity and the length of pitch.

Keywords: Recorded audio, evidences, computerized approach, variation, software etc.

I. INTRODUCTION

In forensic investigation, the scientists deal with various types of evidences that are recovered in distinct

forms i.e. contaminated form, disguised, partial information etc. These evidences provide a valuable link between the victim and suspect/ culprit. It may include any form of the evidences i.e. physical, chemical or biological etc¹. physical evidences mainly deals with the physical form of any evidences or quality, tangible article². It also encompasses all other objects, living or inanimate; solid-, liquid, or gas. Among these physical evidences i. e. fingerprints, palm prints hair strands, blood etc. These evidences can be apparent to human eyes or visibility can be increased by use of microscope or macroscope. Among all these types of evidences, audio and video evidences (form of physical evidences) play a vital role of the involvement of suspect/culprit in a particular crime. These audio recordings can turn out to be very important evidences in the field of forensic science³.

Audio forensics involves acquisition, analysis, and evaluation of audio/ video recordings that can be eventually presented as admissible evidence in the court of law⁴. These types of evidences can be frequently encountered in the cases of ransom calls, cyber-crimes, lost and found cases, or voice notes etc^{5,6}. These can be available from the emergency call helpline, telephone answering machines, voicemail recording, cell phones and computer files⁷.

In the earlier time case audio files were stored in form of an analog format but at present time, the devices are store in form of digital codes. Naturally, sound is a series of analog signals which are converted to digital signals by using ACD (analog-to-digital converter), present in the form of a microchip in laptops, computers and mobile devices. When these sound signals are sent to the speaker to be played, they are converted back to analog signals which is then perceived by our ears. Human hearing power is limited between 20-20,000 Hz, and audio signals outside this

frequency range are not heard by humans. MP3 or MPEG-1 Audio Layer-3 is the most common format of storing audio files⁸.

Audio files have become an inseparable part of modern life and thus these form a special category in the field of forensics as ‘evidences’ which establish an important link between the crime, victim, and the suspect which undoubtedly aids in exonerating the innocent and punishing the guilty. Voice identification often plays a decisive role in cases such as ransom cases, blackmail, threats, obscene calls, extortion, political or other intimidation, bomb or fire hoax, black box voice identification, intelligence collections (identification of spies, saboteurs, etc.), identification of drug dealers (manufacturers, distributors, peddlers, suppliers, etc.) identification of smugglers, their contacts, distributors and sellers of smuggled goods in the country, handling terrorists, their identifications, identification of their contacts, partners, paymasters, supporters, etc. Although at some instances, few variations are encountered⁹. These factors can occur because of genes, gender, environment, ethnicity, age, development of voice box, mood of the person, old age, disease, recording equipment, etc. While audio variations in voice of the same individual are evident in samples recorded at different times. Any voice sample can also be a witness in form of any differences in exhibit when played using diverse equipment’s. It is these variations in a recorded audio that are being measured in this study, using a computerized approach. Sometimes, these evidences can require recovery of information, repair from the storage media without altering the intelligibility of the audio, enhancement of the sound signals to increase intelligibility and improve audibility, interpretation and analysis to identify the speakers. Forensic phonetics and voice biometrics play an important and vital role in the analysis of audio evidences along with various softwares like PRAAT, TF-32, WaveSurfer, CSL, MDVP, etc where the results are displayed as spectrographs providing visual representation of sound signals thus easing the process of comparison¹⁰.

In this present study, we analyzed the variation present in the audio evidences with a computerized approach. We have used “PRAAT” software for the analyses of audio samples. This software enables to analyze an audio recording using multiple features at the same time. It is primarily intended for acoustic analysis of speech. We observed the variations in a single audio recording recorded on two different devices simultaneously. These two devices were used to record the audio are; the laptop using the PRAAT software and an android phone. So according to the research design, we have two audio files in each sample

i.e. (a) recorded on laptop using PRAAT and part (b) recorded on android phone. These files are analyzed on PRAAT by obtaining a spectrogram of the audio recordings then marking the plots of intensity and pitch and finally getting the maximum value of pitch and the maximum value of intensity. The maximum value of pitch and maximum value of intensity is noted in both (a) and (b) part of each sample and results were interpreted. This study also illustrates that even when an audio is recorded simultaneously on different devices it shows considerable variations when analyzed using software PRAAT in its intensity and pitch.

II. METHODOLOGY

This present study was initiated to determine the variation present in the recorded audio of same individual by various sources. For this work, 25 samples including males and females were collected from the distinct subjects of various region in National Capitol of India. All the samples were collected from an age group of 17-30 years old. All the subjects were selected randomly and purpose of the study was disclosed in favor to obtain the consent of an individual. To obtain the exemplar from the subjects, a standard format was prepared. According to this standard, every subjects had to speak “ How are you” at both of the time.

A. Method

During this present study, the recording of the samples was performed on two distinct devices simultaneously. All the recordings were performed in a closed room without any other disturbances or noise. At first recording, an android phone model Le-ECO-2 was used while the second recording was directly exposed to the software installed in Lenovo E-40 Laptop. The recorded audio in the android phone was transferred to the laptop with the help of data cable for further analysis process. To analyze the recorded samples, both audio files were examined in PRAAT software. Before we transfer the recorded audio in the device, we should ensure the format of the stored file because PRAAT software doesn’t accept mp-4 file format. Therefore, we can use the online converter mp-4 to mp-3 convertors. Hence in this study, the online Video Convertor _{v-3.0}’ was used. In all the collected samples, the above-mentioned criteria were applied while the recordings were kept constant.

B. Procedure

For the analysis of all recorded audio’s all the files were inserted in an appropriate manner in the system. Hereafter, samples were run through the fixed parameters on the window that is always present on the right of PRAAT picture and used for plotting graphs.

These plots can be saved in various formats, including an EPS postscript-3 or in a window. Before analyzing speech samples for both (samples & exemplar), it is important to adjust the sound card options properly. The features can be controlled/operated by the volume control window. Now for analyzing the samples, firstly, we opened the file of the recorded audio in system. The sample was analyzed to observe the waveform and the spectrogram of recorded voice/sound in between the range of 0Hz to 5000Hz. Now, intensity button was pressed to determine the intensity of voice and as a resultant of it, yellow plots were obtained. These plots range between 50dB to 100dB of frequency. To obtain the pitch, blue colour plots were appeared that ranges from 75 Hz to 500 Hz.

These observations were obtained in form of the maximum values of intensity and pitch. Thereafter, spectrograms were recorded. Similarly, second audio file was opened obtained from same subject that was recorded on the android phone and the same steps were followed. Now, both the files were compared based on the recorded spectrograms. Many other steps that include filtration of the signal, enhancement in the specific frequency regions etc. can be performed. In this present study, we only described the option of filtering the signal that is necessity during the use of PRAAT software. But if the specific frequency region is focused that is very common in practice. At very few instances, it can be helpful to segment a speech waveform and to attach labels to each segment. The intensity, pitch and spectrogram were observed at the above-mentioned ranges for each and every recorded sample.

III. RESULT & DISCUSSION

Several researchers/ scientists devoted their time to facilitate the forensic sciences investigation and come out with new gigantic array of technologies. In the field of forensic audio- video examination, acoustics characteristics are considered the preliminary study to distinguish one individual from others and sort list the number of suspects. Now a day, computerized approaches have minimized the time and accuracy in the identification to an individual.

According to the existing findings by this time, the voice of an individual differs from time to time and even depends on several facts such as age, disease, temperature or surroundings etc. these differences are even significant to quantify the voice of suspect. During this study, the voice of same individual was analyzed at different paragraph. When the sample no.1 were directed to PRAAT software, the maximum intensity was 87.88 dB while the maximum pitch was 145 Hz.

The graphical presentation is given below in graph no.- 1 (a);

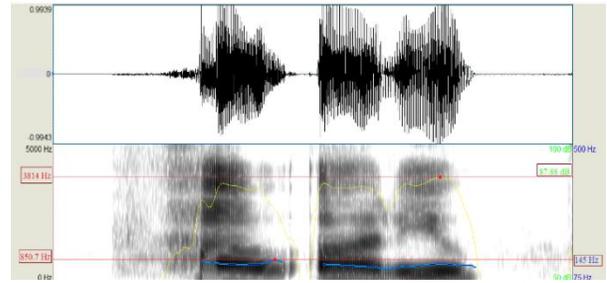


Fig.1(A);Analysis of recorded audio in PRAAT software.

At the same time, when the same individual was directed towards the android phone, the obtained maximum intensity was 77.59 dB while the maximum pitch was 166.6 Hz. the graphical presentation is given below in graph no.1 (b);

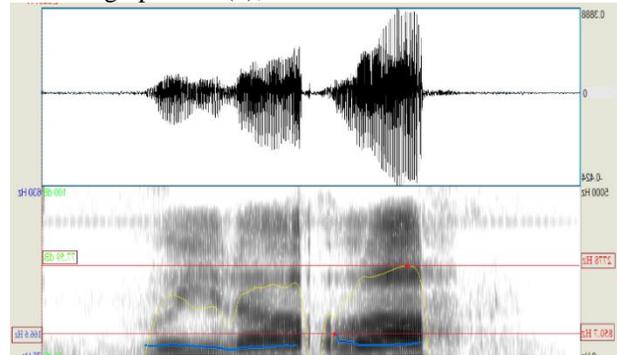


Fig. 1(b) Audio recording taken by android phone.

The obtained result of this sample shows that variation occurs when the recorded audio was transferred from one device to another device. It may be a cause of the present malware, virus or any other functions in the phone/device. Similar like of this process, when other samples were analyzed, the results indicate the variation. According to the obtained results of sample no.2 the maximum intensity was 89.25 dB at a maximum pitch of 326.7 Hz in PRAAT software. While in the recordings of android phone it was 82.44 dB (maximum intensity), Maximum pitch – 322.7 Hz. The analytical graphical presentation is given below in table no. 2 (a) and 2 (b)-

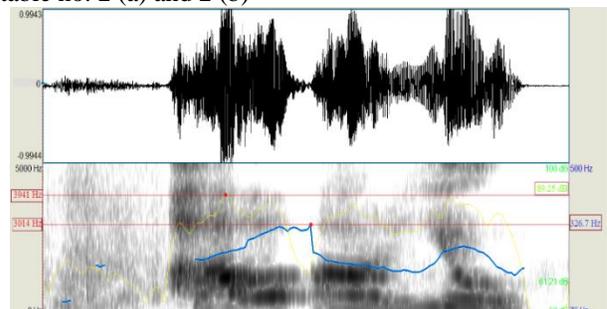


Fig.2(a) Audio recording taken in PRAAT

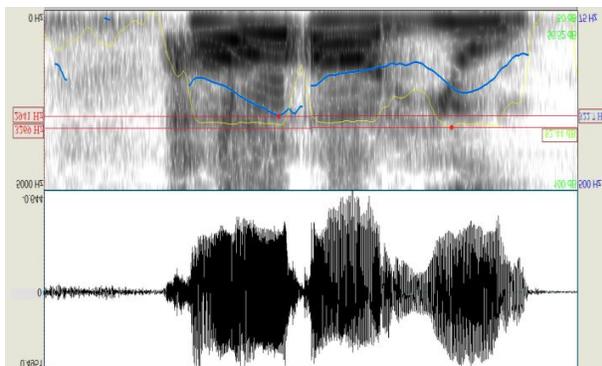


Fig. 2(b) Audio recording taken by android phone.

As per the results of this study, the speaker identification evaluation by using the audio from highly variant devices (software and android phone) helps in the voice activity detection and speaker/individual identification. Therefore, PRAAT proved a significant role in the identification of native speaker among the recorded audios. Along with the identification of variation, identification can be done based on the intonation and track over the time. Moreover, through the use of PRAAT, it requires an additional assistance by verifying the visual feedback that was shown on the praat screen.

IV. CONCLUSION

This study illustrates that whenever any audio is recorded simultaneously on distinct devices, it shows considerable variations. This variation can be observed in PRAAT software measuring its intensity and length of the pitch. Even if a person tries to disguise his/ her voice which might be quite differ or will have spoken by use of any acoustics characteristics, it will provide a clue about an authenticity from the pitch because the intensity of voice and pitch will remain same at any instrument or from distinct devices. But this study shows that the values of maximum intensity and pitch varies along with the device on which it was recorded. This is because of the forensic laboratory criteria; the same device should be used to record the voice of an individual. It has been observed that the same device will produce the variation at same place in both of the samples; questioned sample as well as the exemplar of the suspect. As the outcome of this study, this type of results can be produced in court of law which are unable to identify of suspect/perpetrator.

ACKNOWLEDGEMENT

The authors are thankful to the Honorable Founder president of Amity University for providing us this opportunity to utilize our skills and practicability. We are thankful to the Director of Amity Institute of

Forensic Sciences for his support and cooperation in the conduction of this project.

SOURCE OF FUNDING: NA

CONFLICT OF INTEREST: NA

ETHICAL CONSIDERATION; NA

REFERENCES

- [1] K. Walker and S. Strassel, "The RATS radio traffic collection system," in ISCA Speaker Odyssey, 2012.
- [2] Tim Ng, Bing Zhang, Long Nguyen, Spyros Matsoukas, Karel Vesely, Pavel Mat'ejka, Xinhui Zhu, and NimaMesgarani, "Developing a speech activity detection system for the darparats program," in Proc. of Interspeech 2012, Sept. 2012.
- [3] H. Hermansky, "RASTA processing of speech," IEEE Trans. Speech and Audio Processing, vol. 2, pp. 578-589, 1994.
- [4] Z. Tan and B. Lindbeg, "Low-complexity variable frame rate analysis for speech recognition and voice activity detection," Journal of Selected Topics in Signal Processing, pp. 798-807, 2010.
- [5] P. Boersma and D. Weenik, "Praat: doing phonetics by computer (v. 5.1.05)," <http://www.praat.org>.
- [6] Rainer Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Transactions on Speech and Audio Processing, vol. 9, no. 5, pp. 504- 512, 2001.
- [7] S. Sridharan J. Pelecanos, "Feature warping for robust speaker verification," in Proceedings of Odyssey 2006: The Speaker and Language Recognition Workshop, 2006, pp. 213-218.
- [8] Qi Li and Yan Huang, "Robust speaker identification using an auditory-based feature," in Proc. of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2010, pp. 4514-4517.
- [9] R. Kumerasan and A. Rao, "Model-based approach to envelope and positive instantaneous frequency estimation of signals with speech applications," JASA, vol. 105, pp. 1912-1924, 1999.
- [10] M. Athineos and D. Ellis, "Autoregressive modelling of temporal envelopes," IEEE.
- [11] Ayushi Sharma, Ajit Singh, Implementation of 16X16 bit and 32X32 bit vedic multiplier using FPGA board. International journal of engineering trends and technology. Vol 42 (1) 2016, Pg; 1-5.
- [12] G. Sonia, T. Murli Mohan. An efficient multi keyword search with integrated model. International journal of engineering trends and technology. Vol 40 (4) 2016, Pg; 210-214.