International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 01-03 www.ijaem.net ISSN: 2395-5252

Spectrographic, acoustic and phonetic analysis of voice: A Review Report on Forensic Speaker Recognition

Damanpreet Kaur, Rashmi Sharma, Madhav Gaur, Nishigandha Sawarkar

Central Forensic Science Laboratory, Chandigarh.

Submitted: 10-06-2021 Revised: 23-06-2021 Accepted: 26-06-2021

ABSTRACT: With the uncountable rise in digital crimes (pertaining to digital technology) people now understand that the digital evidence could be a potential evidence. User try to preserve the digital evidence in the form of photographs, audio, video recordings. As far as audio/ video recordings are concerned, forensic laboratories are receiving audio/video recordings for authentication, speaker identification and comparison. In past decades a lot of research had been carried out. In the present paper, authors tried to review research conducted using different advanced tools such as Multi-Speech, PRAAT etc.

KEYWORDS: Spectrographic Analysis, Acoustic, Phonetic, Speaker Recognition.

I. INTRODUCTION

The role of digital evidences in forensics is increased in past decades because of easy availability of the digital media to everyone. The people begin to understand the importance of digital evidences. They now get that they can record the audio and video as a proof against the offender. The same audio and video are being sent to forensic laboratories for authentication and voice matching/recognition of the speaker.

The authors of the present paper are not responsible for the authenticity of the research papers, conference papers, papers and posters present in international meetings etc. present here, as the number of the papers are high and collected from various journals. The work done in the field of forensic speaker recognition is large in number. But in the present review paper, we present detailed review of the papers and try to cover the maximum number of studies from the last two decades.

The paper present with the view point to give information about the spectrographic, acoustic and phonetic analysis using different techniques, using samples recorded through different modes, analysis is independent of language, using disguised and normal voice etc. The law enforcement agencies are in need of forensic speaker recognition to fight against crimes such as bribery, murder, rape, fraud cases. Present paper present an overview of the research conducted in the concerned field.

1. Ton Broeders (2001) did a survey of then Forensic Audio experts from all over the world and the ear witnesses of related cases. The conclusion of the survey was that the reliability of the justice system and cases in forensic speech and audio analysis had been increased.

2. Kunzel Herman J. et al (2004) were made use of natural and disguised voice samples of 100 German speakers recorded 5 times. The voice samples were analyzed using Forensic Recognition System and the unknown samples were matched to the suspected voice samples using likelihood ratios. The result of the study distributed the voice samples into two categories, i.e., disguised and non-disguised voice samples (normal pitch, high pitch, low pitch and nose pinched).

3. Hanson Doug (2004) described the ability of forensic audio experts/analysts in recognizing/matching/identifying the voice samples of a person over a tapped calland discussed about the spectrographic analysis for the voice identification even if the voice is distorted. Engelberg S. et al (2006) used MATLAB based computer program to analyze two (02) samples for voice matching. The authors recorded the two same sentence samples in .wav format and analysis was done using spectrogram and calculating Fast Fourier Transforms (FFTs). The computer generated the report which said two voices are definitely a match to one person.

4. Clark Jessica & Foulkes Paul (2007) took an experiment to see the pattern change in speaker identification using electronic vocal disguise. They used voice samples in four sets (one is natural and three are in increasing order of artificially disguise

done by changing frequency of the voice samples). The study shown notable changes in natural and disguised voice samples but the authors suggest more research is required in the concerned field.

5. Neuhauser Sara (2008) worked on disguised voice by using a different language accent. The study was done on the German speakers who can produce French accent. The study was based on acoustic and non-phonetic features of the voice samples. The voice samples were varied in most of the features, but partially matched with the persons actually from French native capable of speaking German.

6. Gangwar D P et al (2011) recorded ten (10) words using eight (08) different recording modes having eight different formats (.avi, .tsh, .tsh, wav, mp3, mpg, analog and .amr) to see the difference in spectrographic analysis. The software 'Multi Speech' is used to compare the average pitch/fundamental frequency (F0), the different formants (F1, F2, F3 and F4) of Spectrogram and Linear Prediction Coding. This study revealed that different words show a significant difference on the various formants when using different recording modes.

7. Pop-Dimitrijoska V. et al (2012) collected voice samples of 18 female individuals speaking Macedonian language. The recording was done in .m4a file format in a quiet environment. The analysis of vowels used in Macedonian language (a, e, i, o and u) was done using software PRAAT and measurements were formed on the basis of mean value of three formants (F1, F2 and F3) of spectrographic analysis of five vowels. The study was successfully matched the unknown person voice sample with a known sample.

8. Sharma S. et al (2013) recorded speech samples of forty healthy subjects (20 males and 20 females) through Voice Over Internet Protocol (VOIP) in three (03) modes- 1.) Personal Computer (PC) to PC 2.) PC to Public Switched Telephone Network (PSTN) 3.) PSTN-to-PSTN. The auditory and spectrographic analysis was done of all the samples and it was concluded that frequencies of formants were similar to normal mode in PC to PC and PC to PSTN mode. But the frequencies in PSTN-to-PSTN mode shift towards lower region.

9. Dethe Bhanudas K. et al (2016) analyzed the voice samples of four (04) speakers having same loudness. The authors compare the samples on the basis of spectrograph and Linear Predictive Coding using the software 'Multi Speech'. By using the frequency of formants and bandwidth, it was successfully established in the study that even if the loudness & pitch of the two different speakers is same but they vary in another characteristic, sound.

10. Bhall et al (2016) conducted an experiment for forensic speaker identification (auditory & acoustic features and spectrographic analysis) from clue-words of 15 sets of speech samples and the correlation with probability. The spectrographic analysis was done using Frequency formants (F1, F2 and F3) of Linear Predictive Coding of Computerized Speech Laboratory (CSL). The study shown that spectrographic analysis and statistical probability of auditory & acoustic features matched positively.

11. Margi D. Vasan (2018) carried out a study with twenty (20) subjects (10 males and 10 females) using four (04) different recording modes (Analog recorder, Digital recorder, Mobile phone [Samsung] and Voice recording software [Goldwave]) and from two (02) average distances for recording of each sample. The auditory and spectrographic analysis was done using software, Goldwave and Computerized Speech Laboratory (CSL). The study revealed that different recording modes have their specific properties which are unique in their own way and the distance from the speaker's mouth make a considerable difference in speech signal.

12. Zhang Cuiling et al (2018) compared the disguised (whisper voice samples) and non-disguised (normal voice samples). This study is based on auditory, acoustic and spectrographic analysis. The voices of the subjects were compared on the basis of phonetic features such as average syllable duration, intensity, vowel formant frequencies and long-term average spectrum (LTAS). The study shown the significant difference in both the voice samples (disguised and non-disguised) of the same speaker. Hence, it is difficult to match the whispering voice sample of a person with its normal voice sample.

13. Albanesi Alice et al (2020) recorded voice samples (control samples and noisy samples) of six (06) pairs of monozygotic twins to find out the similarities and dissimilarities in both the twins. The spectrographic analysis was done using software 'PRAAT' which shown almost similar distribution in frequency of vowels with slight difference in both the twins.

II. DISCUSSION

From the review of the researches discussed in the present paper we can interpret that the forensic spectrographic analysis was done using different software. But the main characteristics of the voice comparison/matching/recognition are based on acoustic, phonetic and spectrographic analysis. One can use statistics, likelihood ratios as an aided method for more significant results.

III. CONCLUSION

The forensic audio analysts can use different software, such as Multi-Speech, PRAAT, MATLAB (used by Engelberg S.). The use of acoustic features, phonetic features and spectrographic analysis is done to analyze the voice samples. The other features which can be the basis of voice examination are auditory features, articulatory features and non-phonetic features (linguistic). In addition to this, vectors, different formulas are being used for the accuracy and precision.

But as the technology increases day by day, the criminals find new ways to disguise their voice using advanced digital signals. So, to decipher the original voice, more compatible tools are required. Hence, it is the need of hour to develop more forward tools to help the law enforcement agencies.

IV. ACKNOWLEDGMENTS

The authors are thankful to Dr. S.K.Jain (Director cum Chief Forensic Scientist) and Sh. D.P.Gangwar (Head of the division and Assistant director, Physics), Central Forensic Science Laboratory, Chandigarh for their valuable guidance during writing of the paper.

REFERENCES

[1]. Clark, Jessica & Foulkes, Paul. (2007). Identification of voices in electronically disguised speech. International Journal of Speech Language and The Law- INT J SPEECH LANG LAW. 14. 10.1558/ijsll. v14i2.195.

[2]. Neuhauser, Sara. Voice Disguise Using a Foreign Accent: Phonetic and Linguistic Variation. International Journal of Speech Language and the Law, [S.l.], v. 15, n. 2, p. 131-159, feb. 2009. ISSN 1748-8893.

[3]. Pop-Dimitrijoska, V. & Apostolovska, G (2012). Forensic speaker identification through comparative analysis of the formant frequencies of the vowels in the Macedonian language. PhysicaMacedonica, 79-84.

[4]. Künzel, Hermann J. / Gonzalez-Rodriguez, Joaquín / Ortega-García, Javier (2004): "Effect of voice disguise on the performance of a forensic automatic speaker recognition system", In ODYS-2004, 153-156.

[5]. JE. M. Smith, "The Accuracy and Consistency of Spectrographic Analysis for Voice Identification," Paper 6920, (2006 October.).

[6]. Hanson Doug (2004): Sound of Crime: Forensic Audio Analysts can lead law enforcement to a criminal by hearing more than a mere voice on the phone. Law Enforcement Technology Vol. 31, 96,98,101.

[7]. Engelberg S., Saidoff Y. and Israeli Y. (2006): Voice Identification through Spectral Analysis. IEEE Instrumentation & Measurement Magazine.

[8]. Gangwar D P, Thakur Sanjeev, Choudhury S.K. (2011): Codec impact on voice spectrographic features: A Forensic Study.

[9]. Sharma Shivani, Jain S.K., Sharma Rakesh M. (2013): Spectrographic study of speech samples recorded through Voice Over Internet Protocol (VOIP). O-COCOSDA/CASLRE, 2013 International Conference.

[10]. Dethe Bhanudas K., Waghmare Ajit V., Mulik Vitthal G., More B. P, Daundkar B. B. (2016): Forensic Discrimination of Voices using Multi-Speech Techniques, International Journal of Engineering Research and General Science Volume 4, Issue 6.

[11]. Bhall Babita, Singh CP, Dhar Rakesh, Soni Rakesh (2016): Auditory and Acoustic Features from Clue-Words Sets for Forensic Speaker Identification and its Correlation with Probability Scales, Journal of Forensic Research 7(4).

[12]. Vasan Margi D., Choudhary Surbhi Mathur, Dahiya M.S. (2018): The Effect of Diverse Recording Devices on Forensic Speaker Apperception System International Journal for Innovative research in Science and Technology, Volume 4, Issue 8.

[13]. Broeders, Ton (2001), Forensic Speech and Audio Analysis Forensic Linguistics 1998-2001: 13th INTERPOL Forensic Science Symposium, Lyon, France, 2001.

[14]. Albanesi Alice, Cenceschi Sonia, Meluzzi Chiara, Trivillini Alessandro (2020): Italian monozygotic twins speech: a preliminary forensic investigation.