International Journal of Advanced Trends in Computer Applications
www.ijatca.com

Identification and comparison of disguised voices with the genuine voices under various circumstances using spectrographically analysis: A review study

[1]Harsimarpreet kaur, [2]Dr. Ridamjeet Kaur
[1]Students M.Sc. Forensic Science,
Department of Forensic Science, Chandigarh University, Gharuan
[2]Assistant Professor,
Department of Forensic Science, Chandigarh University, Gharuan
[1]harsimarpreet98@gmail.com, [2]ridamjeet.kaur@cumail.in

Abstract: *In this modern era, the misuse of computers and tape recorders results in the widespread use of telephones, cell phones and tape recorders. This makes them an effective tool for the commission of criminal offences, where certain forms of crime are often used by criminals. Any criminal minded individuals might use some strategies and tricks to conceal the voice, assuming they would remain undercover, that some would not identify them. The misuse of voice can be evaluated by using the individualization character of audio. The manipulation or alteration of the speech of individuals is regarded as voice disguise. It can be done intentionally or non-intentionally. Luckily, it is not so easy. Because everyone has their own different and unique voice. By examining the parameters like pitch, frequency, way of talking, focusing on vowels can help to identify the disguised voice produced by someone by changing their voice. Pitch shift is a common method of mask introduced by perpetrators When comparing forensic voices, there is a lot of variation in acoustic properties leads to weaker detection in speaker performance. In most cases, the perpetrator tries to cover his voice until an anonymous or diverse call is sent. That is why it is necessary, before naming a speaker, to research the possibilities for covering the voice. The review paper deals with the studies reported on to analyse and compare genuine with disguised voices using various techniques and softwares.*

Keywords: Forensic science, Disguised Voice, Pitch, Frequency, Audio Forensic

I. INTRODUCTION

T As the vocal folds are brought closer together by airflow from the lungs, voice is formed. The vocal folds vibrate when air is forced past them at a sufficient pressure. Speech could only be produced as a whisper if the vocal folds in the larynx did not vibrate naturally. Your voice is as distinct as a fingerprint. The voice is special to each person. This voice may thus be used to verify a person's identity.[1] In compliance with sections 65 (a) and 65 (b) of the Indian Evidence Act, recording of voice conversations is admissible as evidence in the court of law.

The misuse of the computer and the tape recorders results in the common use of telephones, mobiles and tape recorders in the modern period. This makes them an important weapon for the commission of criminal crimes, where offenders also exploit these types of crime. Communication, assuming they would remain undercover, that they would not be noticed by others. Thankfully, it's really no longer true. A person's voice will identify him effectively and pin the crime on him [2]When there are no immediate crime scenes, such as extortion cases, abductions, extortion, coercion, anonymous phone calls, ransom calls, hoax calls, lewd calls, calls for abuse, fixing matches, and so on, the speaker recognition scenario reverses, and the criminals use phones and mobiles to support them in order to preserve their confidentiality for fear of discovery in these situations. Under some circumstances, a person's voice can be a useful identifier. [3]

The area of forensic speech and audio analysis includes a broad variety of practises, of which speaker recognition is undoubtedly the most impressive. Other field activities include improving the intelligibility of recorded samples of speech, analysing disputed utterances, and examining the authenticity of audio recordings. [4]

In forensic terms, the degree of similarity between a Recording and an accused speaker's speech is referred to as the Proof (Evidence).[5]Speech disguise is an intentional act of a speaker who wants to falsify his or her name or to cover it. [6]Pitch shift is a common method of mask introduced by perpetrators in forensic voice comparison, this presents significant variance in acoustic properties, resulting in poorer speaker detection. [7]The voice may also provide data on the emotional or affective situation of a person. While listening to a third-party debate, lay listeners will usually say whether speakers are worried, pleased, depressed, angry, or suffering from intellectual workloads. [8]It is possible to classify voice disguising into two major groups: deliberate voice disguising and accidental voice disguising. Emotional disorders such as excitement, depression etc. or physical illnesses such as cold, sore throat etc. are caused by unintended changes. Done on purpose speech alterations include shifts in the speech when persons attempt to prevent identification. It is important to further split deliberate deviations into two classes.[1]

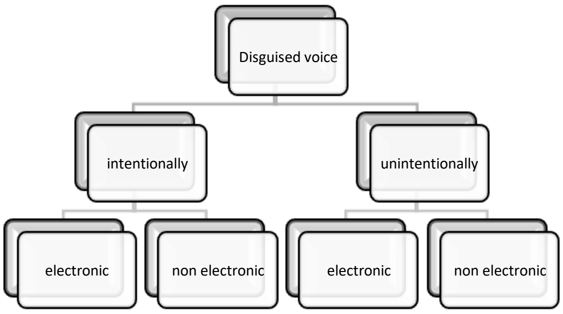

Figure 1: Electronic disguising and non-electronic disguising of voices

By using any computer software, computer speech disguising modifies the speech electronically. In order to adjust the sound, it modifies certain basic parameters such as the pitch, speaking rate, length etc. There is now a wide range of audio editing tools online, such as Audacity, Nice Edit, PRAAT, etc. On the other hand, non-electronic sound disguising affects sound physically by hindering the mechanism of speech processing itself. This involve speaking while speaking with pinched nostrils, whispering voice, using a bite block or handkerchief over the lips, and so on.[1] The digital manipulation methods that are available are specifically designed for musical purposes. They are, however, ultimately added to the sounds of talkers as well. This involves both real-time and post-processing.[9]

Because voices vary widely due to a variety of intrinsic and extrinsic factors (such as fitness, emotion, psychological condition, alcohol, medications, and so on), determining a “normal voice” is difficult (such as environment, noise, reverberation, channel, equipment, etc.). Because speech disguise is the deliberate alteration of a speaker's voice in order to falsify and hide one's identity, "normal voice" is clearly defined as "non-disguised voice," or the speaker's natural voice without any deliberate falsification for the purpose of concealing one's identity[7].

II. STUDY DESCRIPTIONS

While reviewing the data, it has been observed that the methods used for disguising are differently chosen by every researcher which can be common with other but no work is exactly done same. Different methodologies from the paper reviewed are discussed below:

1. An algorithm was given to classify and distinguish obscured voices Voice cover affects the speech signals frequency spectrum and frequency spectral properties can be defined by using *MFCC feature extraction*. The recognition scheme for disguised voices is based on the assumption that values of MFCCs delta, double delta, mean values and correlation coefficients vary from disguised voices. One of essential step involved is feature extraction. They collected speech recordings from students belonging to TKM institute, Kerala. Both electronic and non electronic disguising methods were applied. The audacity software was used to disguise the electronic voice. They used semitones as a masking factor. Disguising factor from +1 to +11 was chosen, and from each of the original voices, 22 different types of voices were created. 40 individuals with 20 males and 20 females were taken from speech recordings. Three form of portrays were used for non electronic (manual) disguising. Those were speaking with pinching nostrils, covered mouth and bite block. The participants were asked to speak normally with parameters to be used along with that. Semitone is also known as half step or half tone. The smallest musical interval typically used in western tonal art is semitone that is known to be dissonant when harmonic sounding. It is known as the interval in 12 tone scale between 2 adjacent tones. They calculated values for MFCC, its delta and double delta and graphs plotted. The whole research focuses on finding hidden voices. The features based on Mel Frequency Cepstral Coefficients (MFCC) are used to differentiate disguised voices from original voices. The theory here is that when a speech is masked, the MFCC statistical moment values change. As a consequence, the MFCC features' mean value and correlation coefficients, as well as their derivative coefficients, are determined. The SVM classifier is used to label a given speech database as 'original' or 'disguised' depending on the acoustic function vector obtained. [1]

2. The difference in F0 value of disguised voice and genuine voice sample was main focus of the analysis.

The value and reliability of F0 speech parameter can also be calculated under various cover conditions. In the course of the study the degree of difference of F0 values between disguised speech and genuine voice samples of each speaker were obtained from 200 different subjects, that were examined and compared using voice spectrograph(CSL-4500). They focused on non electronic mode of disguising. The parameters chosen were keeping hand or cloth on mouth, variation in vocal pitch , stimulating anger, conditions of extreme cold, condition of bad throat, chewing pan or tobacco, constriction of vocal throat, pinching nostrils , pulling cheeks, changing the accent and talking style and mimicry. The F0 parameter for analysis and comparison of disguised and regular person talk has been found more reliable, precise and consistent. The values of F0 were strongly correlated in both males and females with their values of F0 in their control samples in disguised environments, including the constricting of the tract and other parameters. In contrast to male subjects, female subjects have more variance in the values of F0 in their disguised samples.[2]

3. The aim of the project was to detect the likelihood of making definitive opinions on cases involving concealed speech by checking the effect on the personal identity and percentage recognition of speakers of various masked forms for different strategies of voice disguise including elevated pitch, lower pitch of nasality, enhanced mouth covering, constrictive tread, and obstacle in the mouth. This research is exclusively aimed at finding out whether such views can be articulated in cases involving covert expression through the experimental determination of the impact on the personal identity of various disguise formats and the percentage of speech comprehension through different disguise strategies such as higher pitch, lower pitch, enhanced noisy, mouth-capping, constricting tract. This study has taken place at the Speech Section at the Gandhinagar Forensic Science Directorate and the Forensic Science Institute at the Gandhinagar Forestry University in Gujarat. The research contained masks and samples of 200 people, most of whom were of Gujarat descent, of diverse races, faiths and age groups. Of 200 samples of men and 98 of woman speakers aged between 20 and 60 years, 102 were taken. The age range of 25 to 35 years comprised most of the speakers including men and women. Both voice samples were taken from Digital Recorder of high quality. Carefully obtained from each speaker the concealed voice samples have a distinctive condition that applies such variations to the auditory and perceptual parameters of the captured voice sample. All of the dressed and monitored speech samples of each person were subjected to comparative software in order to define their audition and spectrographical parameters in terms of similarities and dissimilitude's. Nearly 22 acoustic parameters, including masked microphones, have been compared. Auditorium features: speech content sample, speech delivery, word use, grammar, accent, speech style, dialect used, speech rhythm, phonation degree, existence & degree of delays, nasality rate and S/T speaking time. Spectrographic parameters: basic frequency; frequency of the forming part; pattern of the forming part; amplitude; energy patterns; timings; loudness; shift, bandwidth. The subjects were asked to make certain adjustments in the original voice to send one of the voice samples.[3]

4. They addressed the uses of statistical algorithms, In order to recognize and diagnose four unique disguises,. The choice of disguises is determined by the most frequent disguises used by criminals. They restricted their implementation in this analysis to non-electronic speech transformations, i.e. to a transition centered on basic techniques that fit those used in cases of crimes. The first move was to determine the influence of the covering voice on the efficiency of the auto-identification of speakers. Four particular disguises were picked based on their use in court cases: mouth side, pinched nose lid, high pitch and low pitch. The automated recognising method theory is split into two parts. Each speaker has been isolated from a 20ms frame (10s overlapped) 12 MFCC (Mel Frequency Cepstral Coefficient) and its derivatives in each section after a silency removal process. The first part consists of a training course aimed at developing new templates for each speaker. The speech patterns of each speaker were modeled by GMM. In various face detection uses, GMMs were commonly used in mathematical models. The principle was that a sufficient number of components approximate some probability density function. The second element was the measure that tests the distance from various models between the question voices. The distance chosen in our method is a chance ratio and the highest value is the good speaker. Experiments and findings focused on various characteristics and separate classification algorithms were presented. The idea was to find a way to detect the mask and to detect it if necessary. The experimental findings indicate that MFCC + its derivatives and QV+SVM grading have fascinating results in case of identification, i.e. in case of normal voice or masked voice being able to tell.[6]

5. Studies were performed on the acoustic characteristics of up and down pitch disguised voices of 11 male speakers in China. In contrast with natural speech, parameters like fundamental pitch, syllables, length, vowel forming frequencies, and long term average range (LTAS) have been calculated and

statistical. The effect of voice masking on human and computer recognition of speakers has also been tested Speech records were obtained at the Chinese Criminal Police University from eleven male students, aged 21 and 24 years of age. Their first language was Standard Chinese (Mandarin). All was asked to mimic the criminal to obtain restitution in an abduction case (role play). The writers are able to reading ten words which in abductions are common in Standard Chinese using usual vocalization and afterwards using shrouded speakers rising and lowering the pitch. For " do not call the police, don't let others know, otherwise child will be killed" The speakers were captured with a handheld digital recorder in a quiet space (Sony ICD-P520). The sampling rate and quantization of the recordings was 16 kHz. The disguised and natural voices of both speakers have been acoustically studied with Praat. Using CSL 3700 (Computer Speech Lab) developed by American KAY Company, only long-term average spectral (LTAS) were analyzed. The findings reveal that the right rate of identification for all listener types is 100% for normal-normal speech pairs. For regular disguised voice pairs the right ID rating for all listener classes is slightly decreased, and disguised voices with increased pitch lead to a low correct ID rate (72% for families and 74% for unfamilies) than disguised voices with a lower pitch (92 percent for familiar group and 90 percent for unfamiliar group). For all types of masked voices, there is no output disparity between two classes.[7]

6. Planwas to increase our perception of changes or the lack of improvement in the signal of speech when people were frustrated. The research attempted in particular to examine the adequacy of using language metrics for detecting deceit. In an interview, ten speakers gave a real, dishonest and control chat. The findings include a group of ten men who speak English between 20 and 30 years old (mean age 1⁄4 24.7 years; SD 1⁄4 3.65 years). Any of the students were from York University students. Much of the speakers originated from northern England and no voice, speech or hearing disabilities were self-reported. The experiment took place in the Department of Linguistics at York University, and the circumstances and the speakers were all preserved continuously. As 'preparatory room' an empty office room was used. The findings reveal that truth tellers and liars are not distinguished on the basis of the speech signals examined in the analysis. The majority of the parameters examined were not only without substantial modifications, but if changes were present, consistencies between the speakers could not be shown. Any parameters could be found with poor trends of discrepancies between Baseline speech on the one side and true/deceptive speech on the other.[8]

7. The investigation of voice captures that have been suspected of being handled by digital signal treatment is among the newest challenges in audio forensics. A number of voice manipulation tools are available in this paper. For musical purposes, several tools are developed. Eventually, however, they also apply to the voices of speakers. This means both after-processing processing in real time. Only maximum bandwidth signals were analyzed and evaluated in this analysis. The study reveals that the various processing applied to the voice recordings has a number of effects on the parameters. If the aim of using voice processing is to provide concealment, these techniques tend to be very accurate in terms of pitch and formants. However, in most cases, a careful inspection will show manufacturing objects. The F1F2 study revealed that the processing-induced variance must be applied to the same-speaker variants. It seems that "undoing" the processing and identifying the features of the undisguised voices is becoming extremely difficult. The voice changes are demonstrated. In conjunction with the forensics and authentication, the influence on acoustic parameters normally examined is presented.[9]

8. tentative normative data for Turkish speaking adults concerning speaker rate, reading rate, articulation rate and articulation rate as well as to compare these 4 gender indicators. In this analysis, it is likely that Turkish participants are taking longer breaks than people who speak other languages because their findings linked to lower speaking rate and high articulation in Turkish in contrast with previous research in other languages. After the informed consent form, the personal information form was given. There were two parts of the form. The first section gathered sociodemographic data such as age, ethnicity, primary language, and educational attainment. The second segment dealt with exclusionary requirements. During the experiment, participants were given a thorough description of the study's intent such that their usual speech performances were not harmed as far as possible. They were just advised that measurements of their speech and reading would be taken in order to determine their speech and reading rates. A total of 84 university students, ranging in age from 19 to 24, were included in the study. The articulation rate was used to measure the power analysis. The analysis reveals that the rate of expression in Turkish is low and the rate of articulation is high.[10]

9. Examination of a user's emotional state classification using voice track analysis, as well as its own solution - the measurement and selection of adequate voice features using ANOVA analysis, as well as the use of PRAAT software for certain voice

factors analysis and the implementation of a custom application to identify a user's emotional state using his or her voice. The results of the created application's tests, as well as the options for expanding and improving this solution, are discussed. To validate the submission, three separate databases were chosen. The measure characteristics Praat script and the Re factor For RAVDESS software were used to retrieve data for the EmoRec recognized case database from the same section of the RAVDESS database that was also used for feature collection. Each database served its function in the testing process. The problem of distinguishing feelings from a user's voice is complex, and the solution was unclear and challenging in the past. Distinct people have different voices, making it difficult to develop general rules for identifying feelings. The best results in this area are achieved by neural networks and systems that integrate emotions determined by voice, facial expression, and other biometrics. Although the current research on methods for identifying emotions is aimed at finding the best one, it is obvious that this field has a lot of future potential. [11]

III. CONCLUSION

Based on the given data point, it indicates that the amount of work performed in forensic speech and audio processing has grown dramatically in recent years. There are indicators that new advances in the understanding of forensic evidence's evidential significance are having an impression on the forensic speaker identity culture. More specifically, there are strong signals that those working in the area of forensic speech and audio processing are becoming more mindful of the value of approaching system confirmation as an integral part of their practice. The hidden voices may decline people’s ear and automated speaker recognition system. Speech mask is commonly used for illicit purposes. A criminal will mask his voice and generate counterfeit proof. Thus, the authenticity of evidence will be adversely influenced. Thus in the world of audio forensics the discovery of disguised voices is unavoidable. In speaker recognition process, the detection of the disguised voices may be used as first step toward determining whether or not the voice test is concealed. The review of literature indicates the future work of voice disguise using several software.

REFERENCES

[1] Lal L T, NJ N A. Identification of Disguised Voices using Feature Extraction and Classification International Journal of Engineering Research and General Science. 2015;3(2)

[2] S M, JM V. Acoustic Analysis for Comparison and Identification of Normal and Disguised Speech of Individuals. Journal of Forensic Science & Criminology. 2016;4(4).). Doi: 10.15744/2348-9804.4.403.

[3] Mathur S, SK C. Effect of Disguise on Fundamental Frequency of Voice. Journal of Forensic Research. 2016;7(3). Doi: 10.4172/2157-7145.1000327.

[4] Broeders A P A. Forensic Speech and Audio Analysis Forensic Linguistics 1998-2001.Netherlands Forensic Science Institute. 2001

[5] Alexander A, Botti F, Dessimoz D, Drygajlo A. The Effect Of Mismatched Recording Conditions On Human And Automatic Speaker Recognition In Forensic Application. Forensic Science International.2004. Doi:10.1016/J.Forsciint.2004.09.078.

[6] Perrot P ,Chollet G. The Question of Disguised Voice. The Journal of The Acoustical Society Of America.2008. Doi:10.1121/1.2935782.

[7] Zhang C. Acoustic Analysis Of Disguised Voices With Raised And Lowered Pitch. IEEE. 2012. Doi: 978-1-4673-2507-3/12/$31.00 © 2012 Ieee

[8] Kirchhübel C, Howard D M. Detecting Suspicious Behavior Using Speech: Acoustic Correlates of Deceptive Speech – An Exploratory Investigation. Applied Ergonomics. Elsevier 2013;44. Doi: 10.1016/J.Apergo.2012.04.016

[9] BrixenB E. Digitally Disguised Voices. Proceedings Of The AES International Conference.2010

[10] Cangi M E ,Işıldar A, Tekin A, Saraç B A . A preliminary study of normative speech rate values of Turkish speaking adults.2020. DOI: 10.32448/entupdates.769051

[11] Magdin M, Sulka J T ,Vozár T M. Voice Analysis Using PRAAT Software and Classification of User Emotional State.(2019). DOI: 10.9781/ijimai.2019.03.004

Author Profile

Harsimarpreet Kaur is students of M.Sc. Forensic science in Department of Forensic Science at. Chandigarh University, Mohali.

Dr. Ridamjeet Kaur is currently working as Assistant Professor in Department of Forensic Science at Chandigarh University, Mohali. She has passed her M.Sc. forensic science from Punjabi university, Patiala in year 2004. She was awarded PhD in year 2013 in subject of Questioned Document examination. She has more than ten years of professional and research experience.. She carries expertise in examination of any type of fake documents such as handwriting analysis, forgery identifications, built up document examinations etc.