Analyzing Voice Calls Using Forensic Voice Analysis System

Subadraa Mahendran
School of Technology
Asia Pacific University of Technology and innovation (APU)
Kuala Lumpur, Malaysia
tp064421@mail.apu.edu.my

Dr. Intan Farahana Binti Kamsin
School of Computing
Asia Pacific University of Technology and innovation (APU)
Kuala Lumpur, Malaysia
intan.farahana@staffemail.apu.edu.my

***Abstract*-The human voice has various distinct characteristics that are referred to as voiceprint. The frequency employed in forensic phonetics to highlight the crime is called a voiceprint. Give us a dependable piece of evidence that can be utilized to determine guilt or identify a perpetrator. There are statistical and mathematical methodologies available, as well as artificial intelligence menthid. On the sound, a visual and auditory analysis is done, and then an assessment is produced using several criteria. Speaker identification algorithms have a success rate of 85 percent to 99 percent. Misrecognition occurs at a rate of 3%, whereas non-recognition occurs at a rate of 10%. In forensic voice comparison, there is a growing emphasis on combining automated and phonetic approaches to increase the validity and reliability of speech evidence presented to courts. To analyze the degree to which long-term measurements of the speech signal collect complementing speaker-specific information, we give a comparison of them. The best-performing system's output was utilized to assess the value of auditory-based voice quality analysis of subpharyngeal (filter) and laryngeal (source) voice quality in system testing. The results imply that the (semi-)automatic system's problematic speakers may be predicted to some degree based on their subpharyngeal voice quality profiles, with the least distinctive speakers giving the weakest evidence and the most misclassifications. The misclassified couples, on the other hand, were readily distinguished using aural analysis. Thus, the quality of the laryngeal voice may be valuable in resolving problematic pairings for (semi-)automatic systems, thus boosting their overall performance.**

Keywords—forensic phonetics, speaker identification, voice disguise, investigation

I. INTRODUCTION

Forensic voice analysis is a thorough examination of audio from a variety of sources, including a phone call, voicemail message, or secretly recorded conversation. This is to create an accurate phonetic profile of a speaker to establish that person's identity. Enhancement of poor-quality audio recordings, as well as verification of questioned recordings to determine whether they have been modified, edited, or otherwise tampered with, are other related disciplines (Saied abd el-atty, Kaniappan Vivekanandan, 2017).

Given its practical applications in the forensic sciences, speaker identification seems to be gaining favor. In most situations, forensic speaker identification (FSI) means comparing incriminating speech to the voices of one or more suspects to determine whether they were produced by the same person. While FSI may be performed by ordinary listeners (witnesses to a crime) or expert listeners (those who have been specialized in phonetics and acoustics), even under ideal conditions, it can be difficult since numerous variables influence the identification process (Didla, 2020).

The detection of speech indicated by voice disguise is a much more challenging problem for forensic voice analysis. People appear to hide their voices for a variety of reasons. Nonetheless, they may be basically reducing to just two: 1) impersonating someone for the sake of amusement (as in mimicry) and 2) impersonating someone with the illegal motive of hiding one's identity. Kidnapping, threats, extortion, hoaxes, and other crimes are often connected with disguised communication. That is, such efforts are most common when the criminal believes his or her 'identity' must be safeguarded, or when the criminal is aware that he is being videotaped.

As a result, "keeping his identity hidden is advantageous to the speaker in issue."

Voice disguise, without a doubt, has a significant negative impact on speaker recognition. Given the prevalence of voice disguise in the criminal world, as well as its devastating effects on speaker identification, this research aims to provide a solution that can contribute to forensic science by developing a which can identify the essential traits of the suspects based on the audio evidence collected.

A. Forensic Science

Science and the scientific method are applied forensically to the judicial system to enforce laws, government regulations, and statutes. Forensic scientists evaluate and interpret evidence in civil and criminal proceedings using cutting-edge scientific processes. Forensic science may help prove a defendant's guilt or innocence in criminal cases. In civil disputes, forensics may aid in the settlement of a broad variety of legal issues by identifying, analyzing, and appraising physical evidence. (Yuhang Zhao, Ruigang Liang, Xiang Chen, Jing Zou, 2021).

Common forensic science laboratory disciplines include forensic molecular biology (DNA), forensic chemistry, trace evidence examination, latent fingerprint examination, firearms and toolmarks examination, handwriting analysis, fire and explosives examinations, forensic toxicology, and digital evidence. Outside of forensic labs, forensic pathology, forensic nursing, forensic psychiatry, forensic entomology, and forensic engineering are some of the forensic areas that are performed. Practitioners of these professions may be found at medical examiners' and coroners' offices, institutes,

and private practices. (Pietro Colombo & Elena Ferrari - 2019). The number of cybercrimes such as unethical hacking, security breaches involving personally identifiable information (PII), online frauds, and others is constantly rising as nations advance with technology, and cyber security has played a vital role in investigations and prevention. The threat environment has changed in tandem with cybersecurity defenses, notably in malware, from classic file-based malware to sophisticated and mobile malware (Sudhakar, Sushil Kumar, 2020).

On the Internet, there is a vast volume of unstructured cyber security data that is difficult to immediately identify and use by a cyber security system. Cyber security blogs, business forums, and related databases are common sources of this information. The extraction of complicated entities, on the other hand, may be improved. Furthermore, text mining is very important in the realm of cyber security. Many obstacles still exist in the world of cyber security, such as nested entity identification and overlapping connection extraction. (Chen Gao, Xuan Zhang, Hui Liu , 2021).

The research supports the necessity for international agreements on cooperation in the development and sharing of computer and information technologies, with the goal of preventing the destruction of electronic evidence and requiring governments to implement and comply with these accords. (Moussa, Ahmad Fekry, 2021). According to several studies, as technology and electrical devices advance, the discipline of digital forensics will continue to grow in popularity and demand in different businesses.

B. Voice Analysis

The frequency of phonation as well as the fluctuation of the voice are measured in voice analysis. The vibrations of the vocal folds of the larynx produce voiced sounds. The feeling of hoarseness, roughness, and breathiness is linked to the fluctuation of the voice signal associated with opening and shutting of the vocal folds.

The area of forensic speech and audio analysis encompasses a vast range of operations, the much more prominent of which is, without the need of a doubt, speaker identification. Using voice analysis, the examiner has two ideas about how to make an identification. Voice parametrization is the process of converting a voice signal into a set of selected features that emphasize the speaker unique characteristics (Shivangi Rao, 2021). In many criminal cases, sound recordings are offered to the court as evidence. These expert evaluations determine if the sound recording is a work of fiction or a result of manipulation. For these records to be used as evidence, there should be no precise fiction and manipulation in the expert examinations (Karakoç, 2017).

The human ear is always more accurate than automated approaches. We're talking about voice printing, sometimes known as spectrogram matching, in which a human observer matches the spectrograms of a suspect's speech to the identical word said by an intercepted speaker. (Michele Catanzaro, Elisabetta Tola, Philipp Hummel, Astrid Viciano, 2017). In most cases, forensic speaker identification (FSI) entails comparing the incriminating speech to the voices of one or more suspects to see whether they are generated by the same person. While FSI may be performed by ordinary listeners (witnesses to a crime) or expert listeners (those who have been educated in phonetics and acoustics), even in the best of conditions, it can be difficult since numerous variables influence the identification process (Dr. Grace Suneetha Didla, 2020)

When interindividual variations are viewed as a signal rather than noise, underlying mechanisms of top-down and bottom-up processes under varying pressures may explain a significant percentage of the discrepancies (Martine Van Puyvelde, Xavier Neyt1, Francis McGlone and Nathalie Pattyn , 2018). The influence of the cross-language issue on these systems is unknown for obvious reasons, but it has also received little attention in published research on auditory perception.

Few studies have been conducted to determine the impact of a bilingual speaker's employment of a specific language on voice recognition. Due to the endless ways individuals may disguise their voices, research on vocal disguise has had limited success in the field of forensic speaker identification (FSI). Given its imposing role in the criminal world, the myriad difficulties affecting voices must be tackled in a methodical manner if the intended outcomes are to be achieved.

C. Cybercrime

The number of people using the internet is steadily increasing. However, as the Internet becomes more integrated into daily life, cybercrime is on the increase. According to the cybersecurity ventures report from 2020, cybercrime would cost approximately $6 trillion per year by 2021. Cybercriminals use any network computer device as a principal method of contact with a victim's device for unlawful operations, allowing attackers to benefit financially, publicly, and in other ways by exploiting system weaknesses. Cybercrime is on the rise every day. Over the previous two decades, the quantity of cybercrime research has exploded. Much of the early research in this field was devoted to determining how cybercrime and cyberspace varied from conventional crime and terrestrial space.

Due to a lack of record maintenance at concerned offices and a variety of factors such as victims' assumptions about police response, lack of awareness of users about IT (information technology) acts on cybercrime, and victims' inability to recognize that they have been victimized, there is currently no foolproof systematic and reliable tool for cybercrime reviews (Rupa Ch, Thippa Reddy Gadekallu, Mustufa Haider Abidi, and Abdulrahman Al-Ahmari , 2020). Promoting and teaching people about cybercrime, as well as identifying and maintaining area-based cybercrime statistics, might all help to reduce and categorize cybercrime.

The increase of cybercrime inside the cyber culture demonstrates the symptoms of altering real-world socioeconomic issues. The borderless, unregulated character of cyber-society crime makes it impossible to trace and has created a perfect environment for social concerns to flourish. Make use of innovative telematics technologies that are difficult to notice and may be used everywhere (M Khairul Basrun Umanailo, Imam Fachruddin, Deviana Mayasari and Rudy Kurniawan , 2019). Cybercrime researchers' research should be the primary source of knowledge for policymakers, the public, security experts, and other academics

interested in reducing different types of cybercrime. Regrettably, there are few evidence-based studies that

evaluate the success of cybercrime regulations (Adam M. Bossler and Tamar Berenblum , 2019).

Both low reporting and problematic measurement of cybercrime necessitate a greater focus on legislation required in this area, and the role of experts in the detection and prevention of cybercrime (Catherine Friend, Lorraine Bowman Grieve, Jennifer Kavanagh , 2020). Theory application to cybercrime as the underlying principle requires the offender and victim to occupy the space as the same time for the crime occur, however do not negate the qualitative differences that intrinsically exits between the spatiality of non-virtual and virtual worlds (Smith, 2020).

While this review was helpful in offering some much-needed insight into the digital realm of cybercrime laws and awareness, there are some suggestions that may be made to better the area's future. Although no viable strategy to cybercrime has yet been discovered, the world as we know it has changed and will continue to alter dramatically over the next 10-15 years as technology becomes more embedded.

II. SIMILAR SYSTEMS

A. *Spectrogram*

A spectrogram is a visual representation of a frequency-time graph. A common approach of visualizing the voice is spectrogram analysis. A sound spectrograph is like a speech picture, including several criteria that are utilized in visual inspections. This method may be used to break down sound waves into their constituent parts. As a result, it's widely utilized to determine voice qualities. Fig 1. shows a spectrogram for the term "digital world" (Mehmet Mehdi Karakoc, 2017).

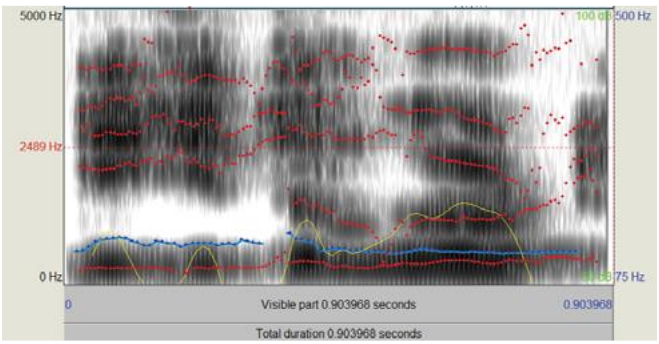

Fig. 1. A Spectrogram

For visual analysis, software such as praat, wave surfer, csl, and tf32 are used. Because the tf32 has a simpler structure, it is better to employ it in the analysis process. Praat has the benefit of annotating speech files and making labelling and analysis operations easier; you can also produce an audio report; wave surfer includes a huge variety of setting changes; csl, mdvp, and so on, all of which are beneficial in certain research and therapeutic situations. It has been shown that depending on the criteria used in the tests, such as male-female, adult-child, and so on, these sorts of software may create erroneous comparisons. As a result, different coding techniques, such as the fast Fourier transform (fft), should be applied and the results compared.

B. *Auditory Analysis*

The technique of evaluating the components of voice that may be detected by the ear is known as auditory analysis. Sound quality, pitch period, speech disruptions, physical interventions, breathing order, ambient acoustics, personal speaking style, volume, mouth, background noises, noise variations, and other acoustic components are all considered throughout this procedure. Experts in comparison procedures consider psychological variables such as anxiousness and enthusiasm that are influenced by the speech (Asaf Varol , 2017).

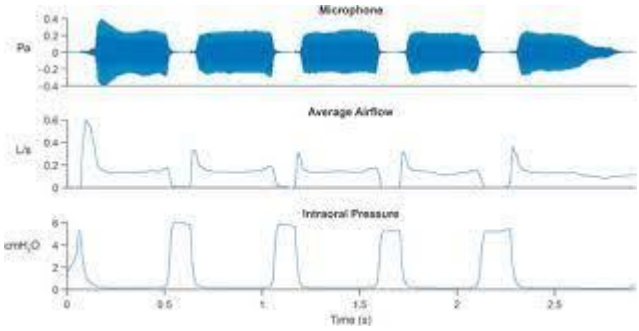

Fig. 2. Auditory Analysis

C. *Speaker Voice Log System*

In a speaker log, the major issues to be resolved are "who is speaking" and "when is speaking." The speaker log is a widely used system. We may utilize an adaptive algorithm for a single speaker using speaker segmentation and clustering technologies to improve speech recognition performance and subsequently increase audio content comprehension. The index and administration of speaker information for multimedia data may also be created in more depth using speaker recognition. In speaker clustering, there is no speaker training data, and the speech characteristics and quantity of speakers are unknown (Sait and Boger, 2020).

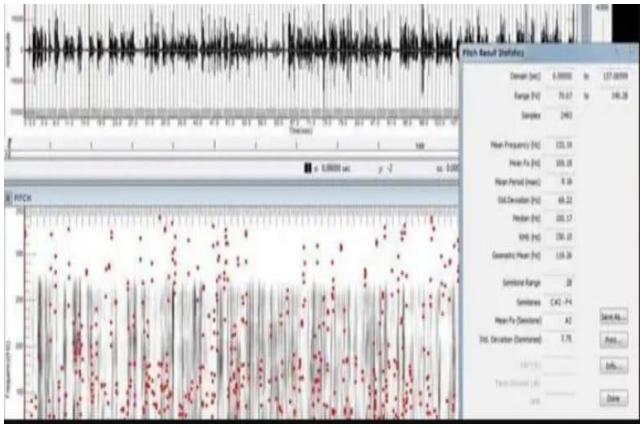

Fig. 3. Speaker Voice Log System

There are a variety of speaker segmentation clustering approaches available today, the most of which are based on hierarchical clustering. However, each technique differs in terms of distance measurement, stop criteria, and clustering model selection. According to the findings of the NIST assessment competition, clustering methods based on Bayes information criterion (BIC) and cross likelihood ratio (CLR) may achieve excellent results for broadcast news corpus. Systems based on HMM BIC or t-test distance outperform those based on BIC in the conference corpus. For telephone conversations, the e-HMM approach performs better. NIST has made significant progress in related technologies in recent years by hosting speaker log system evaluations.

D. System Comparison

Table I demonstrate systems with components that are comparable to those in the proposed system in this proposal. The components that are considered are Sound Waves, Sound Characteristics and Background Cleaning.

TABLE I. SYSTEM COMPARISON

Similarity / Components	Spectrogram	Auditory Analysis	Speaker Voice Log
Sound Waves	✓	✓	
Gender Identification			✓
Sound Characteristics (Frequency, Pitch etc.)	✓	✓	✓
Speaker Quantity		✓	
Background Cleaning	✓	✓	

III. PROBLEM STATEMENT, AIMS AND OBJECTIVES

Harassing, unpleasant, vulgar, or threatening phone calls may happen to anybody. Prank calls, late-night calls, repeated meaningless calls, calls when nothing is said, vulgar calls, calls from past love interests, or threatening phone calls are examples of these. These calls are made intentionally irritate you, either as a form of retaliation or to satisfy the caller's own desires. The expansion of the telephone network and the availability of Voice over Internet Protocol (VoIP) have both contributed to the availability of a flexible and easy-to-use technology for consumers, but they have also led to a considerable rise in cyber–criminal behavior. A lot of interest has been expressed into the analysis and assessment of telephony cyber-threats. A better understanding of these types of abuse is required to detect, mitigate, and attribute these attacks.

Forensic Vocal Analysis is a voice biometrics tool used for criminal identification and law enforcement. It properly analyses audio evidence by using voice biometrics technology in a manner that makes working with audio evidence simpler. It aids forensics professionals and security businesses in correctly performing voice treatment and speaker identification activities. Forensic Voice Analysis aids in the criminal investigation and prosecution of criminals by providing simple identification.

The research focuses on aiming at developing a system for forensic investigators. They can analyze a voice call/recording to identify essential traits of a suspect, such as age and gender and narrow down the suspect in an investigation.

- To speculate the crime's purpose and the primary perpetrator's identity. While the objective of the research is:
- To create an accurate phonetic profile of a speaker to establish that person's identity.
- To increase forensic service providers' capacity so that evidence can be processed rapidly, and investigations may be completed swiftly.
- To improve forensic analyses' dependability so that examiners may present findings with more specificity and confidence

The primary purpose of this system is to identify the essential traits of call evidence collected during an investigation. This speculates the suspect's identity. Creating an accurate profile of the speaker increases the capability of forensic service providers so that evidence may be addressed immediately, and investigations can be finished. This system will also improve the reliability of forensic analysis so that examiners may report outcomes with more specificity and confidence.

IV. METHODOLOGY

A. Sampling

Stratified sampling is a frequently used sampling approach by researchers when they are attempting to derive findings from distinct subgroups or strata. The strata or subgroups should be distinct, and there should be no overlap in the data. While the researcher should use stratified sampling, he or she should also employ basic probability sampling. The population is subdivided according to age, gender, nationality, job profile, and educational level. When a researcher wishes to ascertain the existing connection between two groups, stratified sampling is performed.

The stratified sample approach was used to acquire data in this investigation. The key justification for utilizing this technique is because this study will take a quantitative approach and will include respondents with a range of diverse job experiences. Professionals such as police officers, forensic specialists, and individuals have been used in this study.

B. Identify Respondent

Police officers and forensic professionals served as responders in this research. As a result of their increased likelihood of comprehending how threat calls and forensic services function. On the contrary, the study includes the public as responders since the subject at issue demands anybody who has had an encounter with threat calls to speak out.

C. SAMPLE SIZE

A total of 100 respondents will be surveyed, including members of the government, forensic professionals, police, and people. In the future, the answers evoked from this group of people may be analyzed and improved upon.

D. Data Collection Method

The rationale for adopting surveys as the data collection approach in this research is because the responses received may be utilized to develop the app. Additionally, it is vital to deliver surveys in a very straightforward and accessible style that all respondents from varied backgrounds may readily interpret. The questionnaires will be provided to respondents via email, and the same will be true for the answer retrieval method; this is done to make the survey more cost effective and time efficient for both sides, not to mention highly eco-friendly. The survey form has 10 multiple choice questions and one objective question in which respondents are asked to suggest ways to enhance the app's functionality. An analysis would be done using the facts and figures gathered from the survey to assess the proposed system's features and establish the feasibility of the system, enabling users to have intimate

knowledge with how the system functions. The efficacy, reliability, appropriateness, and usability of the proposed system are all examined.

V. OVERVIEW OF PROPOSED SYSTEM

Fig 4. shows how the system will analyze the voice database collected and process it into 2 categories which is speech and the personalities traits which are programmed to be listed. Based on this we can have a rough idea on who the culprit will be and narrow it down the suspects that is being listed. More efficient results will be provided if we have a voice sample of the culprit.

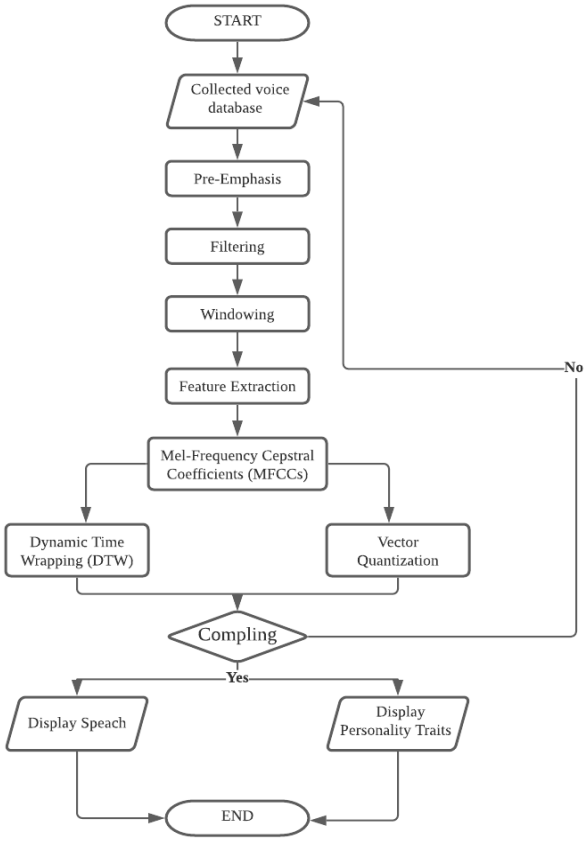

Fig. 4. Overview of Proposed System

VI. CONCLUSION

To sum up this research proposal, it seems as if the amount of work in forensic speech and audio analysis has expanded significantly over the years. Recent advancements in the interpretation of the evidentiary value of forensic evidence seem to be making an impact on the forensic speaker identification community as well. More significantly, there are obvious indicators that people working in forensic speech and audio analysis are becoming more conscious of the need of treating method validation as an intrinsic component of their profession. In a subject that has historically been – and some would say continues to be – rather contentious, these improvements are long overdue, but that does not make them any less beneficial.

REFERENCES

Diarmaid Harkin, Chad Whelan & Lennon Chang (2018) The challenges facing specialist police cyber-crime units: an empirical analysis, Police Practice and Research, 19:6, 519-536, DOI: 10.1080/15614263.2018.1507889

Drygajlo, A. Retrieved from https://link.springer.com/referenceworkentry/10.1007/978-0-387-73003-5_104

Forensic Speech Analysis. Retrieved from https://cyfor.co.uk/digital-forensics/forensic-speech-analysis/

Lian, J. (2020). Retrieved from https://link.springer.com/article/10.1007/s10772-020-09747-2

Martine Van Puyvelde, Xavier Neyt, Francis McGlone and Nathalie Pattyn. (2018). Retrieved from https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01994/full

Mehmet Mehdi Karakoc. (2017). Retrieved from https://www.researchgate.net/publication/320274450_Visual_and_Auditory_Analysis_Methods_for_Speaker_Recognition_in_Digital_

Michele Catanzaro, Elisabetta Tola, Philipp Hummel, Astrid Viciano . (2017). Retrieved from https://www.scientificamerican.com/article/voice-analysis-should-be-used-with-caution-in-court/

Moore, S. Retrieved from https://www.azolifesciences.com/article/Voice-Analysis-in-Forensics.aspx

Rupa Ch,Thippa Reddy Gadekallu,Mustufa Haider Abidi and Abdulrahman Al-Ahmari . . Retrieved from https://www.mdpi.com/2071-1050/12/10/4087/htm

Tade, O. (2022). Retrieved from https://theconversation.com/global/topics/cybercrime-3809

Tushar-ml. Retrieved from https://what-when-how.com/forensic-sciences/voice-analysis/

what-when-howIn Depth Tutorials and Information: Retrieved from https://what-when-how.com/forensic-sciences/voice-analysis/