Asian Journal of Advances in Research

Volume 7, Issue 1, Page 369-384, 2024; Article no.AJOAIR.3762

Advancements in Forensic Voice Analysis: Legal Frameworks and Technology Integration

Pragati Jain [a*], Pragna Chinmayee [a], Kamaljeet Kaur [a], Shefali Chaudhary [a], Kulwinder Kaur [a] and S. Karunya [a]

[a] MSc Forensic Science, University Institute of Applied Health Science, Chandigarh University, Ludhiana, Punjab, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://prh.mbimph.com/review-history/3762

Review Article

Received: 20/05/2024
Accepted: 22/07/2024
Published: 25/07/2024

ABSTRACT

Forensic acoustics, specifically forensic phonetics, plays a crucial role in legal investigations. It aids in speaker identification, tape authenticity, and analyzing contested statements. In India, the legal framework for forensic voice analysis has evolved through amendments to the Indian Evidence Act and key judicial rulings, although specific legislation for voice sample testing is lacking. Internationally, voice identification has long-standing applications with significant advancements in voice analysis technology. Technologies such as Layered Voice Analysis (LVA) and Phonexia Voice Biometrics Solution demonstrate high accuracy in identifying individuals and uncovering emotional cues, meeting international standards for court admissibility. AI and machine learning enhance forensic voice recognition by providing rapid and accurate analysis, addressing traditional limitations. Ongoing research, including Muiredach O'Riain's project on machine learning for

**Corresponding author: Email: pragatijain980@gmail.com;*

Cite as: *Jain, Pragati, Pragna Chinmayee, Kamaljeet Kaur, Shefali Chaudhary, Kulwinder Kaur, and S. Karunya. 2024. "Advancements in Forensic Voice Analysis: Legal Frameworks and Technology Integration". Asian Journal of Advances in Research 7 (1):369-84. https://jasianresearch.com/index.php/AJOAIR/article/view/464.*

forensic audio classification, underscores AI's innovative potential in forensic applications. Modern statistical techniques like Bayesian analysis offer more reliable results, despite challenges such as voice alterations due to illness and diverse interpretations of sound analysis methods. Advancements in technology and AI integration present promising avenues for improving the accuracy and reliability of forensic voice analysis in legal contexts. Continued research and development are necessary to maximize its effectiveness in the pursuit of justice.

Keywords: Forensic acoustics; forensic phonetics; artificial intelligence; speaker identification; legal admissibility.

1. INTRODUCTION

Forensic phonetics, applies the scientific study of speech sounds (phonetics) to legal cases. In cases involving speaker identification, tape authenticity, or contested statements, law enforcement and legal professionals often consult phonetic experts and researchers to provide evidence [1].

This field of Forensic Sciences encompasses the domain of forensic communication, which can be further categorized into three sub-disciplines [2]. These sub-disciplines are designed to cater to the requirements of criminal justice, judicial, and intelligence organizations.

- **The first sub-discipline is forensic linguistics**, which includes Psycholinguistics and focuses on the evaluation of language, whether written or spoken, to identify the authorship, individual intention, deception, and other relevant factors. Within this sub-discipline, Forensic Phonetics also plays a role in speech/language decoding [3].
- **The second sub-discipline is known as forensic psychoacoustics,** which signifies the auditory aspect of human perception. This contributes in analyzing audible signals and their impact on people and their actions from acoustic, perceptual, and neurological perspectives.
- **Forensic phonetics** focuses on analyzing spoken communication to achieve specific goals. This includes tasks like identifying speakers, enhancing, and decoding speech, analyzing emotional cues in voices, verifying recordings, and other related activities.

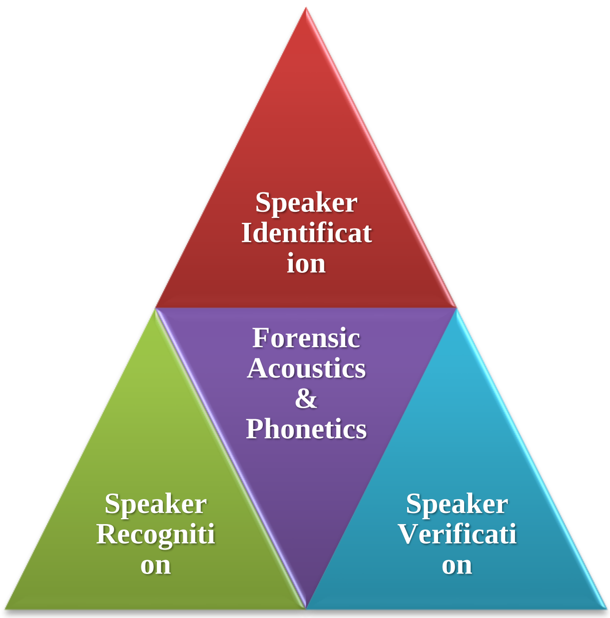

Fig. 1. 3 fundamental aspects of forensic acoustics and phonetics

The 3 fundamental aspects of forensic acoustics and phonetics is:

- Speaker identification: It is defined as the systematic method of identifying a particular speaker, by providing accurate information about vocalization [4-7].
- Speaker recognition: It is defined as the process of identifying person based on the characteristics of his/her voice [8-11].
- Speaker verification: It is defined as the process of establishing the authenticity of the identity of the speaker [12-15].

Since 1960s, various forensic laboratories have been engaged in the examination of audio evidence. The Federal Bureau of Investigation (USA) has led the way in the effort to examine the authenticity of such audio recordings. The field of audio forensics has gone through significant advancements over the last four decades, by leading in its recognition as a specialized area by the government [16].

In the past, analog signals were utilized, but they were associated with several limitations, such as reduced signal stability and quality, as well as limited capacity [17]. To tackle these issues, digital recording has progressively gained importance. While digital recordings yield superior outcomes, they are susceptible to manipulation, as the ability to modify, save, and manipulate the digital bitstream cannot be ruled out.

In 1950s, the first portable magnetic tape recorders were swiftly employed for clandestine purposes such as capturing secret interviews, wiretaps, interrogations, and confessions. Subsequently, the initial examination of audio data occurred with the advent of live recording devices being used outside of recording studios [18].

In the early 1960s, the FBI in the United States started training specialists in audio forensics to improve speech clarity, enhance recordings, and verify their authenticity. Additionally, Ernst Werner Siemens obtained a patent for a dynamic or moving coil microphone in 1874.

The utilization of the dynamic microphone was not widespread during that time. As technology advanced, however, the dynamic microphone came to serve as the fundamental basis for most microphones used today.

In 1930, the esteemed French Nobel Laureate Jean-Baptiste Perrin designed the Perrin Acoustic Array, which garnered recognition by being featured on the cover of Popular Mechanics. Described as an "Airplane Finder Which Locates Ships in Flight and Registers Their Altitude, Speed, and Distance from the Apparatus," this array showcased Perrin's remarkable ingenuity. Moreover, on the British TV show "Wanna Bet?" hosted by Ant and Dec, a fourteen-year-old named Kyle Kirshbaum astutely identified five vacuum cleaners solely based on their sounds. For those seeking to investigate vacuum cleaner noise, the Sim center Test lab Sound Diagnosis serves as a commendable option [17].

The field of audio forensics emerged prominently during the Watergate scandal in 1973. In response to the gaps discovered in President Nixon's covertly recorded Watergate Tapes, a federal court enlisted an audio engineering team to investigate. Their analysis revealed that crucial portions of the recordings had been intentionally removed, leading to the development of innovative magnetic tape analysis techniques [17,18]. Helmholtz resonators, like blowing over an empty coke bottle to amplify specific frequencies, are employed in this process. They harness trapped air to boost tones, akin to how Helmholtz's resonators were initially used to isolate individual musical instrument tones.

Lastly, Annex G was introduced to the ECMA-74 standard in 2018, elucidating an innovative approach for determining the "Tonality" sound meter.

The latest approach draws inspiration from our sense of hearing. Instead of simply assigning values ranging from 0 (no sound) to 1 (full sound) [17], it extends this range to encompass a broader spectrum, factoring in the varying levels of loudness in tonal sounds.

This review paper will explain how forensic phonetics helps solve legal cases by identifying speakers, verifying recordings, and analyzing contested statements. It will highlight the history and advancements in the field, present key methods and technologies used for voice analysis, and provide real life case studies. Additionally, the paper will discuss the role of artificial intelligence in improving forensic analysis and outline challenges and future possibilities.

2. LITERATURE STUDIES

The previous review literature on forensic voice analysis reveals a dynamic and evolving field, which is connected to a particular medicolegal significance. Furui's studies [13,14] has explained the methods of automatic speaker verification and suggested the advancements in the speaker recognition technology, by covering the both methodologies those depend on specific text and those do not. In their study, Gish and Schmidt [7] has focused on the probabilistic modelling techniques for the process of text-independent speaker identification and highlighted their robustness in handling the abnormalities. Campbell [9] has provided a basic tutorial on automatic speaker recognition system, by elaborating their operational modes and the high accuracy achieved in modern systems.

Hsu et al. [15] has presented an efficient algorithm for speaker verification, which can reduce computational requirements and Kinnunen et al. [6] has optimized vector quantization for faster and accurate speaker identification. Howard [3] has explored the intersection of linguistics and legal standards and advocated enhanced forensic linguistic analysis techniques. Jessen [19] has signified the complexities, introduced in the field of forensic speech analysis in comparison to controlled laboratory settings, highlighting the subtle tasks; voice comparison and presentation of voice evidence in the court of law. Grimaldi and Cummins [4] have introduced modern framework of different techniques utilized for speaker identification, and have shown the effectiveness of AM-FM (Amplitude Modulation- Frequency Modulation) speech signal representations compared to traditional MFCC (Mel frequency cepstral coefficients) parameters.

Maher [2], has emphasized the core functionalities in audio forensics, by including authentication and enhancement of audio recordings and projects for future research directions. Hollien [20] has outlined the significance and emergence of forensic phonetics, by defining its scope and importance in the field of forensic science. Hansen and Hasan [8] have compared human and machine speaker recognition, by discussing the advancement in automatic systems and their metrics of performance. Later, Mukattash [1] has signified the origins and methods of forensic phonetics by acknowledging foundational contributions by the important pioneers like Baldwin, French and Nolan and detailing applications such as speaker profiling and voice line-ups.

Recently, Bai and Zhang [5] have reviewed the significant advancements in the field of speaker recognition due to the integration of deep learning mechanism, which offers enhanced representation of speaker characteristics. Ajitprasad (2022) has discussed the ability of Layered Voice Analysis (LVA) for the detection of physiological, cognitive, and emotional deviations associated to deception, which will play a significant role in interview and interrogations. Sheoran and Mahna (2023) has emphasized the significance of noise cancellation in forensic voice analysis to improve the quality and coherence of the speech signals and enabling accurate speaker profiling by the utilization of Phonetic and acoustic analysis toolkit (PRAAT). Gouri et al. (2024) has explored the impact of the social media on acoustic properties, by showing how the effect of different communication medium on vowel sound and speaker identification. Ajitprasad (2022) has discussed the ability of Layered Voice Analysis (LVA) for the detection of physiological, cognitive, and emotional deviations associated to deception, which will play a significant role in interview and interrogations. Sheoran and Mahna (2023) has emphasized the significance of noise cancellation in forensic voice analysis to improve the quality and coherence of the speech signals and enabling accurate speaker profiling by the utilization of Phonetic and acoustic analysis toolkit (PRAAT). Gouri et al. (2024) has explored the impact of the social media on acoustic properties, by showing how the effect of different communication medium on vowel sound and speaker identification.

3. OBJECTIVE

The primary objective of this review article is to provide an overview of the forensic voice analysis. The paper will discuss important aspects of forensic acoustics, legal admissibility of voice analysis procedure, case studies, and the future potential of this field when integrated with artificial intelligence.

4. FORENSIC VOICE ANALYSIS

It is necessary to first provide a general review of phonetics and explain how and why professionals in this field perform the tasks and functions mentioned below. The acoustic, physiological, and perceptual elements of human speech are of particular interest to an Experimental Phonetician. To conduct research, teach, and practice in these areas, an Experimental Phonetician must be educated in various supplementary disciplines, including some aspects of phonology. Among the most significant are engineering specializations related to acoustics [21].

A comprehensive understanding of human behavior, encompassing psychology, physiology, and hearing—specifically psychoacoustics and hearing neurophysiology—is essential for phoneticians. To meet client needs, they must also be capable of developing custom processing methods or modifying existing relevant systems. Additionally, a professional with expertise and interest in various forensic fields is known as a Forensic Phonetician [21].

Forensic phonetics has two subfields. The first involves the analysis of transmitted and stored speech signals, usually using electro-acoustic methods [20].The second focuses on the study of communication acts themselves. These areas aim to enhance recording authenticity, speech decoding (including transcript accuracy), and speech intelligibility, among other objectives. The second subfield includes tasks such as voice recognition, analysis of physical, emotional, and psychological states, and examination of speech for signs of deception [22].

Forensic phonetics intersects with many distinct fields. Two examples are forensic psychoacoustics, which addresses auditory-perceptual issues, and forensic linguistics, which studies language, particularly concerning nonhuman signal processing. These ancillary fields will not be covered in this brief review. Instead, this section will focus on four key areas:

4.1 Quality of Voice

Speaker speech is normal and good when it possesses a good quality of voice, pleasant and intelligibility. A number of harmonics, noise ratio should be animated [23].

4.2 Pitch of the Voice

Male and female voices differ primarily in pitch. Female voices are often noted for their higher and more lyrical pitch, whereas male voices tend to be lower and deeper. Pitch can help us determine whether a voice sounds feminine or masculine. According to research from the NCBI Library of Medicine, the average pitch range for females is typically between 160 and 300 Hz, while for males, it ranges from 60 to 180 Hz. a male voice generally falls within the lower A2 to C3 octave, whereas a female voice typically operates in the higher A3 to C4 range [24].

4.3 Loudness of the Voice

A speaker's loudness should typically range from 60 to 70 decibels. It should not be excessively loud or too soft. A raised voice falls between 65 and 75 dB, a very loud voice is between 75 and 85 dB, and shouting exceeds 85 dB [25].

4.4 Intonation

Pitch variation or fluctuation during speech delivery is important. A good speech uses proper intonation, meaning the speaker should not have a monotonous pitch. Intonation refers to the patterns of pitch variation or tones used in utterances. In normal speech, pitch constantly changes—rising, falling, and sometimes staying steady. These different pitches form patterns of variation that constitute intonation. Intonation is closely linked to stress, as significant pitch changes often occur with stressed syllables [26].

4.5 Rate of Speech

It refers to the number of words uttered/spoken in a particular time interval. Speaker's rate of speech should be 140-160 words per minutes or 300 letters per minute, being too fast in the results of omission of words, articulation be may not be similar [27,28].

4.6 Rhythm

It refers to the smooth and continuous flow of speech. A good speaker delivers speech effortlessly, without struggling, hesitating, or pausing excessively. Stuttering disrupts the rhythm of speech. Rhythm is the pattern of sounds in speech, created by stressed and unstressed words. Maintaining a natural rhythm makes speech sound fluent and engaging. Using only stressed words can make speech sound dull and artificial, and listeners may miss the intended emphasis or meaning [26].

4.7 Articulation

Articulation refers to how well a speaker pronounces words clearly. When sounds are blended or parts of words are skipped, it can make it difficult for listeners to understand. Effective articulation requires using the tongue, teeth, and lips to form sounds accurately [29].

Table 1. Description of the 4 different types of approaches used in Forensic Voice analysis [30]

Approach	Description
Auditory Approach	A forensic expert in auditory phonetics listens to recordings from both suspects and known individuals. They analyze and compare the expected sounds from these recordings to determine if they match or differ, considering factors like regional accents, voice characteristics, speech issues, word choice, pronunciation, and tone. They may also examine recordings from other individuals for comparison. The findings are qualitative and presented as evidence in court [30-33].
Spectrographic Approach	Spectrographic approach involves creating spectrograms of words/phrases from questioned and known speaker recordings. Examiner compares same sounds in same words. Known and foil speakers may provide recordings. Multiple repetitions for intra-speaker variability assessment [30-32,34,35].
Acoustic-Phonetic Approach	The forensic expert compares specific measurements of speech characteristics between recordings from questioned individuals and known individuals. They analyze parameters such as formant frequencies, pitch, and timing of voice onset. This process requires human oversight and considerable effort. The results are interpreted within a framework that assesses the likelihood of a match between the questioned and known samples [30,36,19,37-39].
Automatic Approach	Automatic voice comparison systems, also called Speaker Identification (SID) or Automatic Speaker Recognition (ASR), use quantitative measurements of speech acoustics. They can rapidly compare thousands of test pairs based on these measurements. Spectral measurements made for short segments, transformed into mathematical models. Deep Neural Networks (DNN) used for speaker model creation, significantly improving accuracy and speed. Evaluates likelihood ratio between hypotheses: same speaker vs. different speakers [30,40-42].

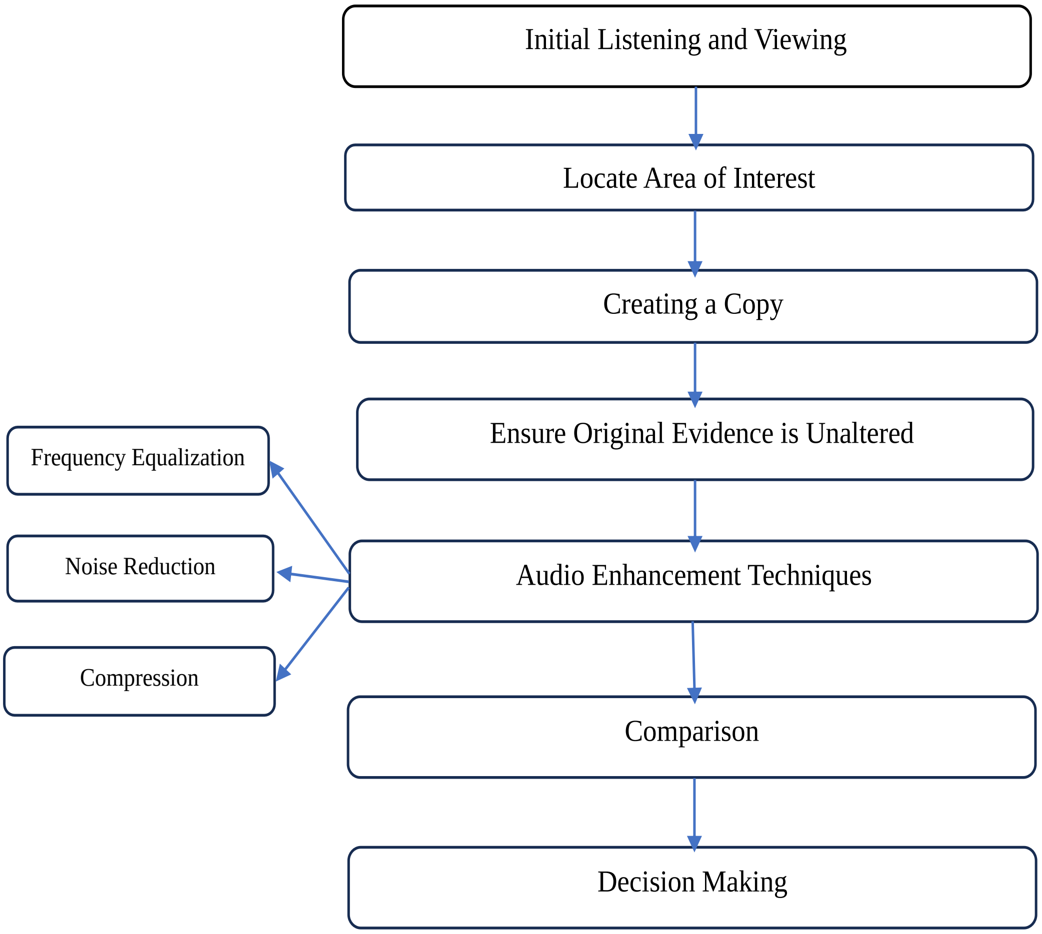

Fig. 2. This flowchart outlines the steps and decisions involved in the process of audio analysis and enhancement [43]

Investigative agencies request court permission to collect a person's voice sample. These samples are recorded in a soundproof room to ensure a controlled, noise-free setting. A voice recorder captures the sample as the person says a specific word from an existing piece of evidence.

An anonymous voice sample is then compared to a list of five suspects. When the speaker's identity is known, both voice samples are directly matched. The International Phonetic Alphabet is used for recording, with the person clearly pronouncing a short segment of the original statement to make analysis easier [44].

5. VOICE SAMPLING METHODOLOGY IN INDIA

Indian forensic laboratories use a semi-automatic spectrographic method for analyzing voice samples. The forensic lab provides a conclusive report to the investigating agency, stating whether the voice sample matches or not. In contrast, some countries use an automatic method that calculates a likelihood ratio for voice samples, improving efficiency [44].

The first step in audio analysis is for the examiner to listen to or watch the recorded footage. The examiner pinpoints the specific section that needs enhancement and closer examination using specialized tools and software. To ensure the original evidence remains unchanged, a bit-for-bit copy is made before processing the audio. Investigators do not change the recorded data; they only enhance what is already there [44].

Audio Enhancement Techniques: Enhancement involves using special techniques to improve the quality of a recording. This might mean making voices clearer, reducing background noise, or boosting the overall sound quality [45].

The techniques utilized for the audio enhancement are:

Frequency Equalization: Precise equalizers are used to adjust specific frequency bands. To improve speech clarity, the frequency range that contains most speech content (200Hz–5000Hz) can be amplified. However, amplifying this range also increases any noise within it, which can limit the effectiveness of clarifying voices.

Noise Reduction: Loud background noises are analyzed using a spectrum analyzer, and their corresponding frequencies are reduced to make the noises less noticeable.

Compression: Faint sounds in the recording are amplified by compressing or leveling the signal, reducing the dynamic range, and making soft sounds more apparent.

Interpretation simply means explaining what the audio evidence shows using words, pictures, stats, and graphs. It helps answer questions during investigations, clarifies what happened with sound events, and helps everyone involved in a legal case understand the importance and limits of the recorded evidence [46].

Based on the result of these process final decision is made for representation of report in court of law [43].

Beyond this, forensic acoustics also tackles issues like whether certain sounds were audible in specific situations, which can be crucial in court cases [46].

Forensic acoustic experts address common concerns with recorded evidence, such as confirming the recording’s authenticity, checking for any alterations or enhancements, and accurately interpreting the captured sounds. Authenticity involves ensuring that a recording hasn’t been tampered with and genuinely represents what occurred. If there is a disagreement about how a recording was made or if anything was changed afterwards, it could lead to legal issues [45,46].

5.1 Voice Cloning

The Cyber Crime Wing of the Tamil Nadu Police has issued a warning about a new scam involving AI voice cloning. They advise the public to be cautious of unexpected calls on their mobile phones. Scammers mimic the voices of trusted individuals to create fake emergencies, urging victims to transfer money quickly. Once the victim believes the emergency, they are asked to transfer money quickly using methods like the Unified Payments Interface (UPI). Victims, driven by concern, often comply without verifying the caller's identity. Later, they realize they’ve been scammed when they contact the real person and find out there was no emergency. This leaves the victim with a financial loss and emotional distress. To avoid falling victim, the public is advised to verify the caller's identity, if already scammed, reporting should be done to the Cyber Crime Toll-Free Helpline 1930 or at www.cybercrime.gov.in [47].

6. ADMISSIBILITY IN COURT OF LAW

The Indian Evidence Act of 1872 was updated by the Information Technology Act to include digital evidence in Indian courts through Sections 65A and 65B. These sections outline the procedure for presenting electronic evidence in criminal trials and its admissibility, aiming to reduce the risk of falsification with specific guidelines. As a result, there are two legal positions: the Anvar case requires strict adherence to Section 65B, while the Shafhi Mohammad vs. State of Himachal Pradesh case allows for bypassing this requirement in the interest of justice. Recently, the Supreme Court ruled that recorded evidence is considered secondary evidence, but in the absence of primary evidence, secondary evidence can be treated as primary evidence in certain situations [48].

Internationally, the US Federal Bureau of Investigation (FBI) first used the technique of voice identification analysis in the 1950s, showcasing its long-standing application in forensic investigations [44].

In 2013, the Indian Supreme Court considered whether collecting voice samples violates the fundamental rights against self-incrimination or privacy. India's criminal procedure laws do not specifically address voice sample testing, as it is a relatively new technology. However, Section 53(1) of the Code of Criminal Procedure allows a medical practitioner to examine the accused at a police officer's request to collect samples for DNA analysis or general body measurements and "other necessary tests." This phrase is interpreted to include the collection of voice samples. In a split verdict in 2013, the Supreme Court acknowledged the lack of a specific law governing the collection of voice samples but later determined that obtaining a voice sample for investigation does not violate the fundamental rights of the accused. The court held that the right to privacy is not absolute and must yield to significant public interest.

Recently, in 2022, a ruling by the Punjab and Haryana High Court likened voice samples to fingerprints and handwriting. It emphasized that voice samples are collected with legal permission and used to compare with previously gathered evidence [44].

A recent study conducted at the University of Madras by Dr. Hemalatha focused on forensic speaker identification of Tamil vowels. The study reported an accuracy rate of 80%, consistent with findings in studies involving other languages. This research marks the first investigation of its kind for Tamil vowels, a departure from more common studies conducted in English-speaking countries. The study's findings suggest that distinct regional accents across Tamil Nadu can aid in identifying suspects involved in cases such as sexual harassment, hoax calls, threats, and bribery. M. Srinivasan, professor and head of the university's criminology department, highlighted the significance of these variations in forensic investigations [44].

In February 2021, India's Special Courts under the Essential Commodities Act and the Narcotic Drugs and Psychotropic Substances Act (NDPS) granted a request by the Narcotics Control Bureau to obtain voice samples from 33 individuals accused in a drug-related case that followed the death of actor Sushant Singh Rajput [44].

While India's legal framework for the admissibility of voice analysis evidence is evolving, it aligns with international practices and judicial rulings that accommodate technological advancements for the sake of justice.

7. CASE STUDIES

7.1 Case Study-1

In United States v. McKeever (1958), two people were accused of extortion as part of a racketeering investigation involving the International Longshoremen's Association. McKeever secretly recorded a conversation with a witness who later testified against him. During the trial, McKeever's defense team wanted to play this recording to challenge the witness's testimony. However, the court only allowed the witness to listen to the recording through headphones, not the jury. The prosecution objected to playing the recording for the jury because its admissibility as evidence was not established yet.

The were many issues faced by the court: the recording was made secretly, the participants were not sworn in, witnesses disputed their voices on the recording, and there was no clear chain of custody for the audio evidence. As a result, the court established the Seven Tenets of Audio Authenticity, which emphasize the effective and ethical use of sound recordings as evidence in court [49,50].

These principles require proving certain facts before admitting a recording into evidence, such as demonstrating the capability and competence of the recording device and operator, ensuring the authenticity and accuracy of the recording, and verifying that it has not been tampered with or altered [49].

7.2 Case Study-2

Tamil Nadu was deeply affected by the murder of M G Ramachandran's foster son-in-law, Vijayan. During the investigation, recordings of conversations involving two key suspects were available. However, when one suspect's voice was recorded at a forensic lab, she spoke in a flat, monotonous tone, resulting in unusable data and her subsequent acquittal.

This incident served as a crucial lesson for analysts. Over time, they began to explore semi-automatic speaker identification techniques that integrate both perceptual and spectrographic analyses. This reliable scientific approach focuses on studying the characteristics of sound. Control voice samples are meticulously recorded in acoustically treated rooms and analyzed using specialized software for formant analysis.

This method yielded a breakthrough in a kidnapping case in 2020. Investigators utilized forensic digital voice analysis to identify the speaker behind a threatening phone call from a pool of six suspects. They observed that one suspect used the term "lanthu," a local expression meaning nuisance, in a distinct Madurai accent. By comparing the formant frequencies of the vowel in this word between the disputed and control recordings, they successfully pinpointed the suspect [51].

7.3 Case Study-3

The Yorkshire Ripper case involved Peter Sutcliffe, who murdered 13 women and attempted to murder seven others. The investigation faced challenges, including false information from a hoaxer Wearside Jack, who was later identified as John Samuel Humble.

Two linguists, Stanley Ellis and Jack Windsor Lewis, were assigned to analyze a hoax audio recording and letters claiming to be from the Yorkshire Ripper. Ellis, an expert in English regional dialects, concluded that the accent on the tape was from Sunderland, an industrial town in County Durham. His analysis was based on previous dialect research and recordings from police officers and students from Sunderland. This led the police to focus their investigation on the Sunderland area.

Ellis used his experience from the Survey of English Dialects to analyze specific pronunciations on the tape, such as the word "I" and "strike," which indicated the speaker's regional background. Despite Ellis' accurate localization of the hoaxer's accent, the police initially disregarded his suggestion that the recording and handwriting were the work of a hoaxer, not the actual murderer.

In 1981, Peter Sutcliffe was arrested, but the hoaxer's identity remained unknown until a 2005 re-evaluation of the case. DNA evidence from the hoax letters matched John Samuel Humble, who was subsequently sentenced to eight years in prison for perverting the course of justice. Although forensic phonetics did not directly identify the hoaxer, Ellis accurately pinpointed the hoaxer's location, which could have expedited Sutcliffe's identification had the police considered his analysis more seriously [52].

7.4 Case Study-4

The Central Bureau of Investigation (CBI) called upon former Congress MP Jagdish Tytler on Tuesday in relation to the 1984 anti-Sikh riots at Pul Bangash. The investigation concerns allegations that three individuals were killed by a mob during that time. Tytler's voice sample was recorded at the Central Forensic Science Laboratory (CFSL) to compare with a purported sting tape released by politician Manjit Singh GK, in which a person claimed to be Tytler and boasted about killing Sikhs.

Despite three previous closure reports, the CBI took this step after new evidence emerged. Tytler, who arrived at the CFSL in the CGO complex for the voice sample collection, denied involvement in the riots and stated the summons were related to a different case.

The Pul Bangash case pertains to the November 1, 1984 riots at Gurdwara Pul Bangash in North Delhi, following Prime Minister Indira Gandhi's assassination. The victims' families had contested the CBI’s closure reports, prompting a re-investigation after a court rejected the closure in December 2007 [53].

8. LIMITATIONS OF FORENSIC ACOUSTICS AND PHONETICS

8.1 Diverse Interpretations of Sound

The scientific community remains divided on the most effective voice analysis methods, despite discrediting some techniques. According to Juana Gil Fernandez, there are two main perspectives. Linguists advocate for semi-automatic techniques that blend computerized analysis with human interpretation, while engineers favor fully automatic systems.

Semi-automatic techniques, known as "acoustic-phonetic" methods, are still the most prevalent. These approaches combine auditory measurements (acoustic) with automated sound analysis (phonetics).

8.2 Automated Systems and False Positives in Speaker Recognition

In the 1990s, automatic speaker recognition (ASR) systems gained popularity for minimizing human judgment. These systems use software to extract features from recordings and match them to a voice database using cepstral coefficients, modeling the speaker’s vocal tract shape. Antonio Moreno of Agnitio, which produces the Batvox ASR system, claims this method is precise and reproducible.

However, experts like Peter French from the University of York argue that ASR often produces false hits due to the similarities in human vocal tracts and advocate combining ASR with human analysis. Sylvia Moosmuller from the Austrian Academy of Sciences criticizes ASR for lacking a strong theoretical basis, as algorithms are trained on studio recordings from the U.S. National Institute of Standards and Technology (NIST) database, which do not reflect real-life conditions.

8.3 Modern Statistical Analyses in Forensic Phonetics

Forensic phonetics needs to adopt Bayesian statistics, a method that considers the rarity of voice features. Geoffrey Morrison explains this using a shoe analogy: rare shoe sizes provide stronger evidence. Similarly, in voice analysis, the likelihood ratio measures how likely a suspect's voice matches a criminal's compared to someone else's, with higher ratios indicating stronger evidence.

Despite recommendations from the European Network of Forensic Science Institutes, few experts use Bayesian methods due to small, limited voice databases. Morrison suggests creating large, diverse databases, but many labs lack any database [54].

9. ADVANCEMENTS AND FUTURE DIRECTION BY THE INTEGRATION OF ARTIFICIAL INTELLIGENCE

9.1 Enhanced Voice Analysis Technology: Layered Voice Analysis (LVA)

Layered Voice Analysis (LVA) stands out as an innovative voice analysis technology due to its unique ability to dissect emotional cues, cognitive processes, and stress levels through methods like "brain trace activity" and "emotional signature." This sets it apart as a potent tool applicable in diverse fields such as criminology, forensic psychology, general psychology, and medicine. LVA effectively correlates psychological aspects with voice characteristics, uncovering subtle emotional cues and contributing significantly to criminal investigations, behavioral screening, and comprehension of intricate emotional and cognitive layers.

Within rehabilitative and correctional facilities, LVA serves to uncover underlying behavioral factors in inmates, aiding in the tailored design of rehabilitation programs. Additionally, it facilitates ongoing program assessment by tracking progress and effectiveness.

While certain studies have highlighted limitations and less-than-optimal performance of LVA, factors such as induced stress simulations, controlled analyses by LVA operators, and the absence of real-time fear responses may have influenced these findings. It's crucial to understand that LVA discerns the speaker's present mental state and internal reactions to deception but does not directly pinpoint falsehoods [55].

9.2 Phonexia Voice Biometrics Solution for Forensic Investigations

Phonexia Voice Inspector 5.1, created by a Czech company for forensic experts, can accurately identify people from just three seconds of their speech, no matter what language they're speaking or what they're saying. This software meets international standards for being used in court and follows guidelines from the European Network of Forensic Science Institutes (ENFSI).

In a 2021 forensic evaluation using the forensic_eval_01 methodology, the technology achieved an Equal Error Rate (EER) of 2.1 percent before calibration and 1.2 percent after calibration. Since then, Phonexia has continued to improve the software.

The software offers several advanced features, including speaker diarization for identifying and labeling individual speakers in mono audio recordings, a phoneme recognizer to detect similar sound patterns, voice activity detection, and a spectrogram for detailed audio analysis. These tools enhance the accuracy and reliability of speaker identification in forensic investigations [56].

9.3 Phonetic And Acoustic Analysis Toolkit

Background noise can affect the accuracy of these analyses. Noise cancellation techniques improve the quality of voice samples, making it easier for forensic experts to identify and profile speakers. Using software like PRAAT, analysts can focus on various features such as vowel duration, pitch, dialect, and voice quality to enhance the reliability of voice evidence. This study shows that noise cancellation metods significantly improve the precision of voice identification in forensic science (Sheoran & Mahna, 2023).

9.4 Impact of Social Media Platforms on Vowel Sound Analysis for Forensic Voice Identification

With the increased use of voice communication on social media, forensic analysis faces new challenges. Astudy examined how platforms like Whatsapp, Instagram, Snapchat, and Telegram affect vowel sounds. Forty participants (20 men and women) were recorded using these apps and a mobile phone. The study analyzed vowel sounds using software and machine learning techniques. It found that specific formant frequencies (F1, F2, F3 and F4) could accurately distinguish vowels and identify speakers. The Random Forest machine learning method was particularly effective in differentiating genders with the accuracy, even with variations in formant frequencies (Gouri et al., 2024).

9.5 AI and Machine Learning in Forensic Voice Recognition

In forensic science, artificial intelligence (AI) is emerging as a critical tool, particularly in voice and speech pattern analysis. This blog post highlights the transformative potential of AI-powered voice recognition in crime-solving.

A person's voice, like fingerprints or DNA, is unique due to physiological and behavioral factors. AI leverages this uniqueness for identity verification, analyzing voiceprints with machine learning algorithms that consider features like pitch, tone, and accent.

AI enhances phone call tracing by quickly and accurately matching voice samples with database records, aiding in suspect or victim identification. It also detects voice disguises, recognizing underlying speech patterns that may reveal a criminal's true identity [57].

Muiredach O'Riain, a B.Sc. Music with Computing graduate from Goldsmiths, University of London, presented his HPCC Systems intern project on "Machine Learning and the Forensic Application of Audio Classification" at HPCC Systems Tech Talk 30. Muiredach's project aimed to create a classification model using HPCC (High Performance Computing Cluster) Systems to accurately identify the recording location of an audio file. This proof of concept lays the foundation for advancements in forensic audio analysis and innovative sound-based information gathering [58].

The project "Machine Learning and the Forensic Application of Audio Classification" focuses on classifying sounds for forensic purposes. The study has shown promising results in linking a room's spectral properties to the classification of sounds recorded in that room. It demonstrated the potential of artificial neural networks to establish these connections and develop a reliable classification model.

The study has certain limitations that need to be addressed. The use of generated data rather than original recordings means all sounds are "recorded" from the same position within the room, using the same equipment. Additionally, the model currently differentiates between sounds in only eight different rooms. To enhance its capabilities, the model would need to train on a significantly larger dataset.

While this technology is not yet ready for real-world forensic audio applications, it lays a strong foundation. The results are an encouraging starting point for further research and development in this field [59].

10. CONCLUSION

Forensic linguistics, with a focus on forensic phonetics, has become a vital element in legal investigations, providing critical insights into speaker identification, tape authenticity, and contested statements. This review article has explored the comprehensive aspects of forensic voice examination, including legal admissibility, notable case studies, and the future potential with AI integration.

In India, the legal framework for the admissibility of voice analysis evidence has evolved significantly, shaped by amendments to the Indian Evidence Act and key judicial rulings. Despite existing challenges, such as the lack of specific legislation for voice sample testing, recent court decisions have facilitated the use of voice analysis in criminal investigations. The Supreme Court recently ruled that recorded evidence is usually considered secondary evidence. However, if primary evidence is unavailable, secondary evidence can be treated as primary evidence in certain cases.

Technological advancements in voice analysis, including Layered Voice Analysis (LVA) and Phonexia Voice Biometrics Solution, show great promise in accurately identifying individuals and meeting international court standards. AI and machine learning are playing a transformative role in forensic voice recognition, providing rapid and precise analysis of voice samples for suspect identification and detection of voice disguises.

Ongoing research, exemplified by Muiredach O'Riain's project on machine learning and audio classification, highlights the innovative potential in forensic audio analysis. These efforts underscore the importance of continuous development in sound-based information gathering technologies.

In conclusion, despite the challenges of voice cloning and limitations that persist in forensic acoustics and phonetics, the integration of advanced technologies and AI presents promising opportunities to enhance the accuracy and reliability of voice analysis in legal contexts. Continued research and development are essential to maximize the effectiveness of voice analysis tools in the pursuit of justice.

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of manuscripts.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Mukattash BK. The role of forensic phonetics in legal investigation: A Case Study of Two Speaker-Identified/Unidentified Recorded Samples, dalam Journal of Literature, Languages and Linguistics, IISTE. 2016;29:31–37.
2. Maher RC. Overview of audio forensics, in Intelligent multimedia analysis for security applications, Springer. 2010;127–144.
3. Howard BS. Comparative and non-comparative forensic linguistic analysis techniques: Methodologies for negotiating the interface of linguistics and evidentiary jurisprudence in the American Judiciary, U. Det. Mercy L. Rev. 2005;83:285.
4. Grimaldi M, Cummins F. Speaker identification using instantaneous frequencies, IEEE Trans Audio Speech Lang Process. 2008;16(6):1097–1111.
5. Bai Z, Zhang X-L. Speaker recognition based on deep learning: An overview, Neural Networks. 2021;140:65–99.
6. Kinnunen T, Karpov E, Franti P. Real-time speaker identification and verification, IEEE Trans Audio Speech Lang Process. 2005;14(1):277–288.
7. Gish H, Schmidt M. Text-independent speaker identification, IEEE Signal Process Mag. 1994;11(4):18–32.
8. Hansen JHL, Hasan T. Speaker recognition by machines and humans: A tutorial review, IEEE Signal Process Mag. 2015;32(6):74–99.
9. Campbell JP. Speaker recognition: A tutorial, Proceedings of the IEEE. 1997;85(9):1437–1462.
10. Hanifa RM, Isa K, Mohamad S. A review on speaker recognition: Technology and challenges, Computers & Electrical Engineering. 2021;90:107005.
11. Leu F-Y, Lin G-L. An MFCC-based speaker identification system, in 2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), IEEE. 2017;1055–1062.
12. Tran D, Wagner M. A fuzzy approach to speaker verification, Intern J Pattern Recognit Artif Intell. 2002;16(07):913–925.
13. Furui S. Cepstral analysis technique for automatic speaker verification, IEEE Trans Acoust, 1981;29(2):254–272.
14. Furui S. An overview of speaker recognition technology, Automatic Speech and Speaker Recognition: Advanced Topics. 1996;31–56.
15. Hsu C-N, Yu H-C, Yang B-H. Speaker verification without background speaker models, in 2003 IEEE International Conference on Acoustics, Speech, Signal Processing, 2003. Proceedings. (ICASSP'03)., IEEE. 2003;II–233.
16. Zakariah M, Khan M, H M-M. Tools and applications, undefined 2018, Digital multimedia audio forensics: Past, present and future, Springer M Zakariah, MK Khan, H Malik Multimedia tools and applications, 2018•Springer. Jan. 2018;77(1):1009–1040.
    DOI: 10.1007/s11042-016-4277-2.
17. Maher RC. Principles of forensic audio analysis; 2018.
    Accessed: Mar. 06, 2024. [Online].
    Available:https://link.springer.com/content/pdf/10.1007/978-3-319-99453-6.pdf
18. Lindley D. Understanding uncertainty; 2013.
    Accessed: Mar. 06, 2024. [Online].
    Available:https://books.google.com/books?hl=en&lr=&id=Tfk8AgAAQBAJ&oi=fnd&pg=PR11&dq=9.%09Lindley,+D.+V.+(2014).+Understanding+Uncertainty.+John+Wiley+

%26+Sons,+Hoboken,+NJ&ots=55DTatVMSQ&sig=2ZweQ5aNQzKK1LBjj5v6hZ342sk

19. Jessen M. Forensic phonetics, Lang Linguist Compass. 2008;2(4):671–711.
20. Hollien H. The acoustics of crime: The new science of forensic phonetics; 2013.
Accessed: Mar. 06, 2024. [Online].
Available:https://books.google.com/books?hl=en&lr=&id=3NBeBAAAQBAJ&oi=fnd&pg=PA3&dq=3.%09Hollien,+H.+(2013).+The+acoustics+of+crime:+The+new+science+of+forensic+phonetics.+Springer+Science+%26+Business+Media.&ots=WezlIJ5He2&sig=06AEpb1JLCzypPVBv9ziKV3-y-0
21. Hollien H, Bahr R, J H-J. of voice, undefined 2014, Issues in forensic voice, Elsevier.
Accessed: Mar. 06, 2024. [Online].
Available:https://www.sciencedirect.com/science/article/pii/S0892199713001318
22. R M-A. Today and undefined 2015, Lending an ear in the courtroom: forensic acoustics, acousticstoday.orgRC Maher Acoustics Today, 2015•acousticstoday.org.
Accessed: Mar. 06, 2024. [Online].
Available:https://acousticstoday.org/wp-content/uploads/2015/08/Lending-an-Ear-in-the-Courtroom-Forensic-Acoustics-Forensic-acoustics-deals-with-acquisition-analysis-and-evaluation-of-audio-recordings-to-be-used-as-evidence-in-an-official-legal-inquiry..pdf
23. Warhurst S, Madill C, McCabe P, Ternström S, Yiu E, Heard R. Perceptual and acoustic analyses of good voice quality in male radio performers, Journal of Voice. 2017;31(2):259.e1-259.e12.
DOI: https://doi.org/10.1016/j.jvoice.2016.05.016
24. Annika Hope, The Vocal Differences between the Genders.
Available:https://www.musicgateway.com/blog/how-to/the-vocal-differences-between-genders
25. How Many Decibels Does A Human Speak Normally.
Available:https://decibelpro.app/blog/how-many-decibels-does-a-human-speak-normally/#:~:text=A%20normal%20voice%20level%20is,voice%20is%20above%2085%20dB.
26. Lucknow University, Dimensions of Speech-Syllable; Accent, Pitch, Rhythm, Intonation.
Available:https://www.lkouniv.ac.in/site/writereaddata/siteContent/202003261644121802kaumudi_Dimensions_Speech.pdf
27. Lynda Stucky, What is the ideal rate of speech?
Available:https://clearly-speaking.com/what-is-the-ideal-rate-of-speech/
28. Caccamise F, Hatfield N, Brewer L. Manual/simultaneous communication (M/SC) research: Results and implications, Am Ann Deaf. 1978;803–823.
29. Lumen, Articulation, Pitch, Rate.
Available:https://courses.lumenlearning.com/wm-publicspeaking/chapter/articulation-pitch-and-rate/#:~:text=Articulation%20is%20how%20clearly%20the,tongue%2C%20teeth%2C%20and%20lips
30. Phonexia, Forensic Voice Comparison: The Essential Guide.
Available:https://www.phonexia.com/knowledge-base/forensic-voice-comparison-essential-guide/
31. Morrison GS. Distinguishing between forensic science and forensic pseudoscience: Testing of validity and reliability, approaches to forensic voice comparison, Science & Justice. 2014;54(3):245–256.
32. Morrison GS. Forensic voice comparison and the paradigm shift, Science & Justice. 2009;49(4):298–308.
33. Jovičić ST, Grozdić Đ. Arguments for auditory-instrumental Aproach in forensic speaker recognition.
34. Morrison GS. Distinguishing between science and pseudoscience in forensic acoustics, in Proceedings of Meetings on Acoustics, AIP Publishing; 2013.
35. Morrison GS. Distinguishing between forensic science and forensic pseudoscience: Testing of validity and reliability, approaches to forensic voice comparison, Science & Justice. 2014; 54(3):245–256.
36. Enzinger E, Morrison GS. Empirical test of the performance of an acoustic-phonetic approach to forensic voice comparison under conditions similar to those of a real case, Forensic Sci Int. 2017;277:30–40.
37. Rohatgi S, Gupta V, Yadav B, Yadav B. Forensic Phonetics: A Linguistic Approach; 2018.
38. Wang H, Zhang C. Forensic automatic speaker recognition based on likelihood ratio using acoustic-phonetic features

measured automatically, Journal of Forensic Science and Medicine, 2015; 1(2):119–123.
39. Huang CC, Epps J. A study of automatic phonetic segmentation for forensic voice comparison, in 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. 2012; 1853–1856.
40. Tirumala SS, Shahamiri SR. A review on deep learning approaches in speaker identification, in Proceedings of the 8th international conference on signal processing systems. 2016;142–147.
41. He X, Huang L. Speaker recognition: Deep neural networks approach, in Advances in Chinese Document and Text Processing, World Scientific. 2017;263–279.
42. Tirumala SS, Shahamiri SR. A deep autoencoder approach for speaker identification, in Proceedings of the 9th international conference on signal processing systems. 2017;175–179.
43. Forensic Science Simplified, A Simplified Guide to Forensic Audio and Video Analysis.
Available:https://www.forensicsciencesimplified.org/av/how.html
44. Drishti IAS, Voice Samples in Criminal Investigations.
Available:https://www.drishtiias.com/daily-updates/daily-news-analysis/voice-samples-in-criminal-investigations
45. B K-J. of the A. E. Society and undefined 1990, Authentication of forensic audio recordings, aes.org.
Accessed: Mar. 06, 2024. [Online].
Available:https://www.aes.org/e-lib/online/browse.cfm?elib=6053
46. Malik H, Acoustics HF, Processing SS., undefined 2010, Audio forensics from acoustic reverberation, ieeexplore.ieee.orgH Malik, H Farid2010 IEEE International Conference on Acoustics, Speech and Signal, 2010•ieeexplore.ieee.org; 1988.
DOI: 10.1109/ICASSP.2010.5495479
47. Sivaraman R, TN. Cyber Crime Police issue advisory on new scam involving AI voice cloning.
Accessed: May 31, 2024. [Online].
Available:https://www.thehindu.com/news/national/tamil-nadu/tn-cyber-crime-police-issue-advisory-on-new-scam-involving-ai-voice-cloning/article68113216.ece
48. Aneja P, Parashar S, Choudhary SK, Mathur S, Sharma P. Audio forensics & jurisprudence: Evidentiary value of voice in investigation, identification & legal admissibility, European Chemical Bulletin. 2023;12(7):518–527.
49. M P-P TF I S. on and undefined 2010, Explaining the likelihood ratio in DNA mixture interpretation, cybgen.comMW PerlinPromega's Twenty First International Symposium on Human Identification, 2010•cybgen.com; 2010.
Accessed: Mar. 06, 2024. [Online].
Available:https://www.cybgen.com/information/presentations/2010/ISHI/Perlin_Explaining_the_likelihood_ratio_in_DNA_mixture_interpretation/paper.pdf
50. Koenig B, D L-AESC. 54th, undefined 2014, Forensic authenticity analyses of the metadata in re-encoded WAV files, aes.org.
Accessed: Mar. 06, 2024. [Online].
Available:https://www.aes.org/e-lib/online/browse.cfm?elib=17321
51. Ram Sundaram & Sindhu Kannan, TN cops add voice analysis as forensic tool to crack crime.
Available:https://timesofindia.indiatimes.com/city/chennai/tn-cops-add-voice-analysis-as-forensic-tool-to-crack-crime/articleshow/103623936.cms
52. Gallego Balcells C. An Overview of Forensic Linguistics and its Application in Real-Life Cases; 2023.
53. Express News Service, 1984 riots: Tytler appears before CBI to give voice sample, The Indian Express.
Accessed: May 31, 2024. [Online].
Available:https://indianexpress.com/article/india/1984-anti-sikh-riots-jagdish-tytler-cbi-voice-sample-8550161/
54. Catanzaro M, Tola E, Hummel P, Viciano A. Voice Analysis Should Be Used with Caution in Court.
Accessed: May 31, 2024. [Online].
Available:https://www.scientificamerican.com/article/voice-analysis-should-be-used-with-caution-in-court/
55. Srivastava S, Hussain AA, Gupta S. A review article on layered voice analysis: Forensic Utility and Limitation, International Journal of Indian Psychology. Aug. 2022;10(3):278–288.
56. Borak M. Phonexia launches voice biometrics product for forensic investigations, Biometric Update.Com.
57. Accessed: May 31, 2024. [Online].
Available:https://www.biometricupdate.com/202401/phonexia-launches-voice-

biometrics-product-for-forensic-investigations

58. O'Riain M. Machine Learning and the Forensic Application of Audio Classification, HPCC Systems. Accessed: Jul. 14, 2024. [Online]. Available:https://hpccsystems.com/resources/machine-learning-and-the-forensic-application-of-audio-classification/
59. Terri, Harnessing the AI Power in Forensic Voice and Speech Analysis, Forensic Science Academy Southern California.

Peer-review history:
The peer review history for this paper can be accessed here:
https://prh.mbimph.com/review-history/3762