ASR In Pursuit Of Forensics Investigation

Prof. Harshlata J. Petkar **Nikita P.Kakkad**

Abstract-**Automatic speaker recognition technology appears to have reached a sufficient level of maturity for realistic application in the field of forensic science. However, there are key issues to be solved before the forensic community will accept its use as an investigative assistant or as evidence in actual criminal cases. To assess the state of the technology, the Federal Bureau of Investigation (FBI) built a speech corpus that included multiple levels of increasing difficulty based on text-independence, channel-independence, speaking mode, and speech duration. An evaluation of multiple automatic speaker recognition programs indicated that a large GMM model-based recognition algorithm operating with features that are robust with respect to channel variations had the best performance. In this paper we describe (1) the challenges, (2) The FBI's initial Forensic Automatic Speaker Recognition (FASR) program based on these concepts, and (3) a confidence measurement method to indicate the probabilistic certainty level of correctness of each recognition decision. We will also discuss the need and justification for input speech screening and pre-processing to improve the recognition performance of the FASR as applied in a real forensic environment.**

Keywords- **Text-independence, Speaker recognition, Challenges, correctness.**

I. INTRODUCTION

Speaker recognition is the general term used to include all of the many different tasks of discriminating one person from another based on the sound of their voices. Forensics means the use of science or technology in the investigation and establishment of facts or evidence in the court of law. The role of forensic science is the provision of information (factual or opinion) to help answer questions of importance to investigators and to courts of law. Forensic speaker recognition (FSR) is the process of determining if a specific individual (suspected speaker) is the source of questioned voice recording (trace). This process involves the comparison of recordings of an unknown voice (questioned recording) with one or more recordings of a known voice (voice of the suspected speaker). There are several types of forensic speaker recognition. When the recognition employs any trained skill or any technologically supported procedure, the term technical forensic speaker recognition is often used. In contrast to this, so-called naive forensic speaker recognition refers to the application of everyday abilities of people to recognize familiar voices.

The success of speaker recognition system depends highly on how to classify a set of feature used to characterize speaker specific information However, pattern classification from speech signal remains as a challenging problem encountered in general speaker recognition system, including speaker verification and speaker identification. Recent development in classifying speaker data from a group of speakers is still insufficient to provide a satisfying result in achieving high performance pattern classification. There are two main difficulties in pattern classification field; first, how to maintain accuracy under incremental amounts of training data and second, how to reduce the processing time as real time systems regarding efficiency and simplicity of calculation.

II. ASR(AUTOMATIC SPEAKER REORGANIZATION)

Automatic speech-
recognition (ASR) can be defined as the independent, co-
mputer-driven transcription of spokenlanguage into readable text in real time. In a nutshell, ASR is technol-
ogy;that allows a computer-
to identify the words that a person speaks ino a microphone or telephone and convert it to written text.

Having a machine to understand fluently spoken speech driv
en speech research for more than 50 years.

Although ASR technology is not yet at the point where -
machines understand all speech, in any acousticenvironment
, or by any person, it is used on a day-to-day basis in a num
ber of applictions and services.The ultimate goal of ASR res
earch is to allow a computer to recognize in real-time, with
100% accuracy, all words
that are intelligibly spoken by any person, independent of vo
cabulary size, noise, speaker characteristics or
accent. Today, if the system
is trained to learn an individual speaker's voice, then
much larger vocabularies are-
possible and accuracy can be greater than 90%.

Commercially available ASR systems usually require only a short period of speaker training and may successfullycaptur
e continuous speech with a large vocabulary at normal pace
with a very high accuracy. Most commercial-
companies claim-
that recognition software can achieve between-
98% to 99% accuracy if operated under optimalconditions.
Òptimal conditions' usually assume that users: have speech
characteristics which match the trainingdata, can achieve pr
oper speaker adaptation, and work in a clean noise environm
ent (e.g. quiet space). This explains why some users, especia
lly those whose speech is heavily accented, might achieve re
cognition rates- much lower than expected.

III. HISTORY OF ASR TECHNOLOGY

The earliest attempts to devise systems for automatic speech recognition by machine were made in the 1950s.Much of th e early research leading to the development of - speech activation and recognition technology was funded by the National Science Foundation (NSF) and- Defense Department's Defense Advanced ResearchProjects Agency (DARPA). Much of the initial research, performed with NSA and NSF funding, was conducted in the 1980s. (Source: Global Security.Org)Speech recognitio n technology was designed initially for individuals in the dis ability community. For example, voice recognition can help people with musculoskeletal disa bilities caused by multiple sclerosis, cerebral palsy; orarthrit is achieve maximum productivity on computers. During the early 1990s, tremendous market opportunities emerged for s peech recognition computer technology.The early versions o f these products were clunky and hard to use. The early lang uage recognition systems had tomake compromises: they we re "tuned" to be dependent on a particular speaker, or had s mall vocabulary, or useda very stylized and rigid syntax. Ho wever, in the computer industry, nothing stays the same for very long and bythe end of the 1990s there was a whole new crop of commercial speech recognition software packages t hat wereeasier to use and more effective than their predecess ors. In recent years, speech recognition technology has adva nced to the point where it is used by millions ofindividuals t o automatically create documents from dictation. Medical tr anscriptionists listen to dictatedrecordings made by physicia ns and other health care professionals and transcribe them in to medical reports,correspondence, and other administrative material. An increasingly popular method utilizes speech rec ognition technology, which electronically translates sound into- text and creates transcripts and drafts of reports.Transcripts and reports are then formatted; edited for mistakes in transla tion, punctuation, or grammar; and checked for consistency and any possible errors.- Transcriptionists working in areas with standardized ter- minology,such as radiology or pathology, are more likely to encounter speech recognition technology. Use of speechreco gnition technology will become more widespread as the tech nology becomes more sophisticated.Some voi- ce writers produce a transcript in real time, using computer s peech recognition technology. Speechrecognition-enabled v oice writers pursue not only court reporting careers, but also careers as closed captionersand Internet streaming text provi ders or caption provider

IV. HOW DOES ASR WORK?

The goal of an ASR system- is to accurately and efficiently convert a speech signal into a text message transcriptionof the spoken words independent of the speaker, environment or the device used to record the speech (i.e. themicrophone).This process begins when a spe aker decides what to say and actually speaks a sentence. (T his is a sequence of words possibly with pauses, uh's, and um's.) The software then produces a speech waveform, which embodiesthe word s of the sentence as well as the extraneous sounds and pause s in the spoken input. Next, the software - attempts to decode the speech into the best estimate of the se ntence. First it converts the speech signal into asequence of vectors which are measured thrughout the duration of the sp eech signal. Then, using a syntactic decoder it generates a valid sequence of representations.

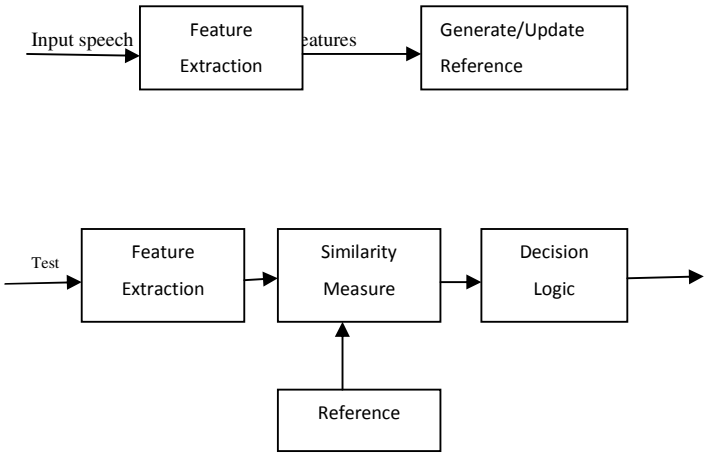

Fig No.1 Working of ASR

V. FORENSICS INVESTIGATION

Here has long been a desire to be able to identify a person on the basis of his or her voice. For many years, judges, lawyers, detectives, and law enforcement agencies have wanted to use forensic voice authentication to investigate a suspect or to confirm a judgment of guilt or innocence. Challenges, realities, and cautions regarding the use of speaker recognition applied to forensic-quality samples are presented. Identifying a voice using forensic-quality samples is generally a challenging task for automatic, semiautomatic, and human-based methods. The speech samples being compared may be recorded in different situations; e.g., one sample could be a general public that it is a straightforward task The article introduced the misleading term, voiceprint identification, which is still in believe that a graphical representation of the voice, via a spectrogram, is just as reliable as the structure of the ridges and minutiae of the fingertips or genetic fingerprints (e.g.,DNA) and that it allows for a highly reliable identification of the original speaker. This misconception complicates the workof those in the forensic speaker recognition domain by introducing a false premise that all voices are unique, and discernibly so,

under most conditions. Combating this mindset has become an ongoing process. No two speakers are absolutely the same, differing some what in anatomy, physiology, and acoustically. Even identical twin scan have similar acoustics and differ in their implementation of a single segment in their linguistic system. In forensics, it is not sufficient to state how similar two speakers are, and typicality must be addressed. To do this, an examiner compares evaluation parameters of the speaker at hand to a larger reference sample of speakers. A measure of typicality helps quantify the strength of the forensic evidence, whichis presented in the form of a likelihood ratio of two probabilities. Automatic speaker recognition systems can aid the forensic examiner in estimating the likelihood ratio. With the developments in automatic speaker recognition over the last decade (e.g., and , there is an increased need to distinguish between its appropriate and inappropriate uses invarious forensic voice authentication contexts and to differentiate between common versus forensic speaker recognition applications. In 2003, several scientific institutions reported on the status of the use of automatic speaker recognition technologies in the forensic field . They concluded by sending a clear need-for-caution message, including statements such as, "currently, it is not possible to completely determine whether the similarity between two recordings is due to the speaker or to their factors . . .," "caution and judgment must be exercised when applying speaker recognition techniques, whether human or automatic . . .," or "at the present time, there is no scientific process that enables one to uniquely characterize a person's voice or to identify with absolute certainty an individual from his or her voice."

VI. PHONETIC

Phonetic symbols are a great help when it comes to learning to pronounce English words correctly. Any time you open a dictionary, you can find the correct pronunciation of words you don't know by looking at the phonetic pronunciation that follows the word. Unfortunately, learning the phonetic alphabet is not always the easiest thing to do. In English, as you certainly know, many words can have the same pronunciation but be written differently with different meanings.

- Articulatory phonetics: the study of the production of speech sounds by the articulatory and vocal tract by the speaker.
- Acoustic phonetics: the study of the physical transmission of speech sounds from the speaker to the listener.
- Auditory phonetics: the study of the reception and perception of speech sounds by the listener.

The study of phonetics grew quickly in the late 19th century partly due to the invention of phonograph, which allowed the speech signal to be recorded. Phoneticians were able to replay the speech signal several times and apply acoustic filters to the signal. In doing so, one was able to more carefully deduce the acoustic nature of the speech signal.

Forensic phonetics: the use of phonetics (the science of speech) for forensic (legal) purposes. Speech Recognition: the analysis and transcription of recorded speech by a computer system.

VII. SPECTRUM

Most of natural signals change over time. A dominant source of change in speech signals is changing shape of a vocal tract that enhances and attenuates individual spectral components of a spectral envelope of speech. Most of the phonetic information is carried in these changes. Spectral components of the logarithm of an appropriately smoothed and rectified signal envelope in a frequency sub-band is what we call in this article the modulation spectrum .The modulation spectrum is related to this logarithmic signal envelope in the sub-band through the linear Fourier transform – both carry the same information. Thus, information-wise, the term "modulation spectrum" is a synonym for shapes of temporal trajectories of elements of spectral envelopes of speech. Therefore, whenever sufficiently long segments of time-frequency patterns of speech are used as inputs tofeature extraction module, we talk about processing of the modulation spectrum of speech.

VIII. TYPICAL CRIMES REQUIRING SPEAKER RECOGNITION

- Homicide
- Drug Matters
- Kidnapping
- Bomb Threat
- Rape
- Physical Assault
- Obscene Calls
- Extortion Calls
- Telemarketing
- False Calls
- White collar crime

IX. TYPICAL RECORDING MEDIA SUBMITTED FOR EXAMINATION

Fig No.2 : Recordings for Examination

Spectrographic Approach by Human Voice identification by use of spectrograms and listening

- Verbatim Samples Required
- Samples must be recorded through similar recording and transmission systems
- Unknown error rates of the human operator
- Automatic Approach by Computer Voice identification by automatic speaker recognition
- Verbatim Samples not Required (text-independent)
- Samples must be recorded through similar recording and transmission systems.
- Error rates and the system performance measures will be available

X. CHALLENGES IN FORENSIC SCIENCE

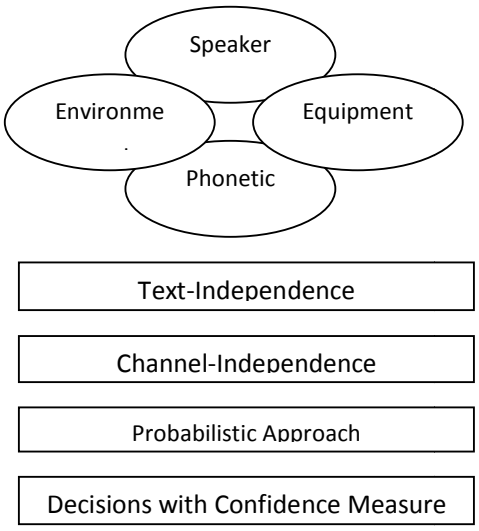

Fig no.3: Forensics application

XI. ERROR ANALYSIS

To improve the speaker recognition performance by the knowledge of forensic environment Separate sources of error into two types:

- Correctable, and Non-correctable
- Discrete tones, stationary or non-stationary
- Wrong recording or playback speed
- Multiple speakers speaking alternately Non-linear distortion
- Overdriving, or clipping both in frequency and/ortime domain
- Voice produced under the influence of alcohol or drugs
- Multiple speakers speaking simultaneously

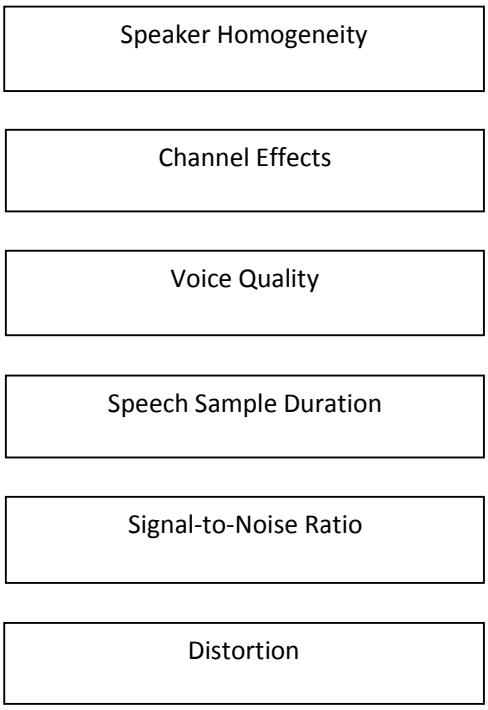

Fig No.4: Error Analysis

XII. CONCLUSION

The current automatic speaker recognition technique is sufficiently reliable in forensics science but remains to be an imperfect system. Speaker recognition is successfully used today in a variety of applications, but must be applied responsibly with caution. Need to address factors in forensic domain that degrade recognition performance, Increase robustness and effectiveness ,Improve calibration and efficiency, Real case data can be extremely challenging,

Opportunities for research., Additional international challenges and opportunities The current automatic speaker recognition technique can report highly accurate recognition decisions, as long as speech input files are produced under similar conditions.

REFERENCES

[1]. Speaker Recognition Workshop, Hosted by National Instituteof Standards and Technology, March 31 – April 1, 1998,College Park, Maryland.
[2]. Siu, M. and Gish, H., “Evaluation of word confidence for speech recognition systems”, *Computer Speech and Language*, 13, 1999.
[3]. Champod, C. and Meuwly, D., “The Inference of Identity inForensic Speaker Recognition”, *Speech Communication*,vol.31, pp. 193-203