æ Íx•t Surbhi Mathuræ Íx t1æ Íx t*æ Íx t, Choudhary SKæ Íx0t2æ Í,and Vyas
3

[1]*Assistant Professor, Gujarat Forensic Sciences University, India*
[2]*Assistant Professor, Raksha Shakti University, India*
[3]*Director General, Gujarat Forensic Sciences University, India*

speaker recognition technique in a much better way. It outlines the basic concepts of speaker recognition along with
LWV GLYHUVH DSSOLFDWLRQV ,W DOVR SUHVHQWV DQ LGHD RI VHOHFWLQJ D UR
DWWDLQ WKH DFFXUDWH UHVXOWV OLPLWDWLRQV IDFHG DQG WKH UHFHQW EXL
technological perspective in this important area of speaker recognition.

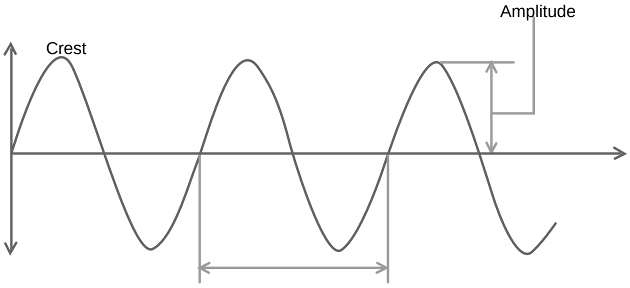
Keywords: Forensic science; Speaker; Recognition; Identification; Verification; Voice; Speech

Introduction

Speaker recognition comprises all those activities which attempt to link a speech sample to its speaker through its acoustic or perceptual properties [1]. Speech signal is a multidimensional acoustic wave (Figure 1), which provides information regarding speaker characteristics, spoken phrase, speaker emotions, additional noise, channel transformations etc [2,3]. The human voice is unique personal trait. For indistinguishable voice, the two individuals should have the identical vocal mechanism and identical coordination of their articulators, which is least probable. However, the some amount variations also occur in the speech exemplars obtained from the same speaker. This is due to the fact that a speaker cannot exactly imitate the same utterance again and again. Even, the signature of an individual also shows variation from trails to trials.

The process of Speaker recognition has two broad application areas explicitly, Speaker identification and Speaker verification. Speaker identification deals with identifying a speaker of a given utterance amongst a set of known speakers. The unknown speaker is identified as the speaker whose model best matches the input utterance (Figure 2).

There are two modes of operation related to known voices: closed set and open set. The closed set mode is considered as multiple class classification modes. Such system assumes that the voice which has to be determined or identified belongs to a set of known voices. While in open set the speaker which do not belong to a set of known speakers, is referred as an imposter. This task can be used for forensic purposes, in which an offender's is used to reveal his or her identity, among several known suspects.



In contrast, Speaker verification is a more direct and converged effort leading to either acceptance or rejection of the claimed identity of a speaker. To be precise, this investigation reveals whether a speake is the one who he claims to be (Figure 3) [4-6]. It can be considered as a true-or-false binary decision problem. It is sometimes referred to as the open-set problem, because this task requires distinguishing a claimed speaker's voice known to the system from a potentially large group of voices unknown to the system. Today verification is the basis for most speaker recognition applications and the most commercially feasible task.

Significance of Speaker Recognition

Security or access control

The voice of a person can be successfully used as a biometric feature as it is well accepted by the users and can be easily recorde using microphones and hardware of low costs [7]. It can provide an

*Corresponding author: Surbhi Mathur, Assistant Professor Jr., Gujarat Forensic Sciences University, Gandhinagar, Gujarat, India, Tel: +919687924168; E-mail: sue.mathur@gmail.com

Received June 07, 2013; Published July 02, 2013

unconventional and more secure means of permitting entry without any need of remembering a password, lock combination etc or the use of keys, magnetic card or any other fallible device which can be easily stolen [8,9].

Although the voice of a person cannot be stolen but it can be copied using some recording devices. Therefore, the voice-based security systems must protect themselves against such flaws. The other concern is voice disguise. An imposter can gain illicit entry by disguising or imitating the voice of a genuine speaker, to access this personal data. Similarly, a valid person may be denied the entry because of some accidental changes in his or her voice due to illness, emotional or physical stress etc.

Forensic or law enforcement

Voice of a person can plays a vital role in forensic examination. In the present era, widely available facilities of telephones, mobiles and tape recorders results in the misuse of the device and thus, making them an efficient tool in commission of criminal offences such as kidnapping, extortion, blackmail threats, obscene calls, anonymous calls, harassment calls, ransom calls, terrorist calls, match fixing etc. The criminals nowadays are more frequently misusing these modes of communication, believing that they will remain incognito, and nobody would recognize them. It is fortunately no longer true. The voice of an individual can successfully recognize him and pin the crime on him [10].

The results obtained through speaker recognition analysis are not easily accepted in the court of law. But with advancements made in this field and with the judges understanding the value of statistical findings, the situation is expected to change in the future [11,12]. But the results in this case also are vulnerable to two types of voice disguise: deliberate and unintentional.

Criteria of feature selection

In a scheme for the mechanical recognition of the speakers, it is desirable to use acoustic parameters that are closely related to voice characteristics that distinguish speakers. It involves selection of those parameters which are motivated by known relations between the voice signal and vocal-tract shapes and gestures. Speaker recognition by and large depends upon both low level and high level information obtained from a person's speech. High level informations include values like dialect, accent, the talking style, the subject manner of context, phonetics, prosodic and lexical information [17]. These features are currently recognized and analysed by humans only. The Low-level features refer to the information like fundamental frequency (F0), formant frequency, pitch, intensity, rhythm, tone, spectral magnitude and bandwidths of an individual's voice [18]. An ideal feature would:

- Have lower intraspeaker variability and high interspeaker variability
- Be robust against noise and distortion
- Occurs frequently and naturally in speech
- Be easily measured from the speech signal
- Difficult to mimic
- Not be affected by speaker's health or long term variations in voice

There are different ways to categorize the features [19]. From the viewpoint of their physical interpretation, following categories have been proposed:

1. Short-term spectral features–These features, as the name suggests, are computed from the short frames of about 20 to 30 milliseconds in duration. They are usually the descriptors of the resonance properties of the supralaryngeal vocal tract.
2. Voice source features – These features characterize the glottal excitation signal of voiced sounds such as glottal pulse shape and fundamental frequency, and it is reasonable to assume that they carry speaker-specific information.
3. Spectro-temporal features-It is very much a rational assumption that the spectro temporal.

   Signal details such as formant transitions and energy modulations contain useful speaker-specific information.
4. Prosodic features-Prosody refers to non-segmental aspects of speech, including syllable stress, intonation patterns, speaking rate and rhythm. These features depends upon the long segments like syllables, words, and utterances and reflects differences in

Recent advances or automatic approach

Switzerland [35,36] are using such methods, which are also being tested in Spain [37] and the United States of America [38]. The FBI recently completed an evaluation project in which four automatic speaker recognition systems were tested on a specially designed forensic database compiled by the FBI. The results confirmed that the performance levels of automatic systems can be quite high when text and transmission conditions are controlled. Deterioration is usually encountered in the conditions related to forensic realm.

Expressing results in forensic speaker recognitionsySis usC /Span <</MCID 440 >>-n t1xt

Citation: Mathur S, Choudhary SK, Vyas JM (2013) Speaker Recognition System and its Forensic Implications. 2: 72.dcdoi: