A Framework for Speaker Recognition System

Singh N[*1], Agrawal A[1], Chandra S[2], and Khan RA[1]

[1]SIST-DIT, Babasaheb Bhimrao Ambedkar University (A Central University), Lucknow, UP, India
[2]SIST-DCS, Babasaheb Bhimrao Ambedkar University (A Central University), Lucknow, UP, India

[*]**Corresponding author:** Singh N, SIST-DIT, Babasaheb Bhimrao Ambedkar University (A Central University), Lucknow, UP, India, Tel: 9956052275, E-mail: nilu.chouhan@hotmail.com

Abstract

An individual's voice holds the characteristics to identify that individual uniquely. Automatic Speaker Recognition (ASR) is a technique for recognizing an individual by his/her voice. Controlling access privileges and forensic is the major application areas of Speaker recognition systems. To develop a robust speaker recognition system it is required that the system is able to provide acceptable performance with several operating conditions. The research in this area is continued from last six decades. Though various developments have been done in the area but there are still many improvements required. In this paper authors presents a framework for developing a speaker recognition system for improving the system performance. Nowadays research effort made in the area of speaker recognition systems is the focus of intense research.

Keywords: Speaker Recognition; Framework of Speaker Recognition; Guideline for Framework; Speech Features; Feature Extraction; Modeling Method; Matching Techniques; Performance Evaluation

Introduction

Human voice or speech signal encloses with rich information about the individual such as speaker emotion, speaker identity, language, message content, speaker temperament etc. speech processing has the following task for example speech analysis, synthesis, coding and recognition etc. further it is classified as speech recognition, language recognition and speaker recognition etc. [1]. Speaker recognition is the process of extracting voice features for personal identity by the analysis of speech utterances. It is a biometric technology which is used in many security areas for secure access control and forensic investigation. In today's digital era where insecurity is everywhere, speaker recognition technologies provide a secure solution in our daily life. ASR systems provides many services e.g. voice based banking, voice database access, voicemail, remote access to personal computers, voice based access control devices, and many other authentication areas [1,2].

In past few years, speaker recognition technology has many significant developments which are now used in several authentication applications such as physical and logical access control systems [3]. Also Current scenario shows that speaker recognition is a growing research area in speech signal processing [1]. Speaker recognition/voice recognition is a process of identifying of an individual on the basis of his/her voice. Voice has the characteristics of both physiological and a behavioral biometric features. There is a difference between speaker recognition and speech recognition. Speaker recognition is the task to recognize who is speaking while speech recognition is the task to recognizing what is being said [4].The current speaker recognition (text-independent system) are language independent but their performance affected in multilingual trial condition [5]. Prosodic features are robust against technical mismatch hence system performance improved by using prosodic features [6].

Speaker recognition has the main task are speech feature extraction or front-end processing and modeling. Feature extraction is the process of selecting required speech features that is further used for speaker modeling. Several speech features has been proposed to till date, each feature have their advantages and disadvantages. These speech features are used for different type of speech processing e.g. speech recognition, language identification and speaker recognition. Process of developing speaker recognition system has two phase, the training i.e. enrollment phase and the testing phase. During training phase, speech samples are collected and system is trained by the collected speech samples. Whereas in the testing phase, provided speech sample is matched by system for identification or verification of speaker. Speaker recognition categorize in two main task speaker identification and speaker verification. Further it is divided into text-dependent and text-independent speaker recognition system. The recognition is said to be text-dependent when the speaker use same linguistic for both i.e. training and testing else recognition is text-independent.

Furthermore system is either open-set or closed set; if the system has the chances to reject the speaker who is not enrolled in the speaker database else it is closed-set database system [7,8].

Background

Human speech is a natural way of communication with each other. It is a medium by which human express their emotions, thoughts and share messages. Speech is a complex signal and it contains several information of the speaker's and language. For example excitation source information, vocal tract system (while producing voice), linguistic information, emotional states of speaker, supra-segmental information (prosodic features e.g. pitch and energy). Human speech is unique for individual due to differences in shape of vocal cord, size of larynx and other voice production organs [9]. Now a days voice based authentication technology has grown up quickly and it is used for authenticating of individuals. This technology can be used but not limited to crime investigation, forensic, personal authentication, voice based system etc. [10-13].

In recent digital era where everything is going to be digitalized, human authentication is also done by machines. It is fact that human being is able to easily distinguish among voices of different persons. To become a recognizer like human the machine should be robust and reliable. Speaker recognition, speech recognition and language identification are the most commonly used authentication processes. Speaker recognition is emergent area in speech signal processing. It is concerned to the identity of a person, based on his/her voice characteristics. Speaker recognition has many applications in distinguished areas such as personal authentication, forensic, security check in military etc. [14-16]. For example, in digital forensic through voice, a suspected person can be recognized by tapped telephone conversation of criminals/terrorists.

One of the popular areas of speaker recognition is authentication. Process of automatic speaker recognition is based on acquiring speech signal; creating speaker models which are used to compare with available models [17]. In general, speaker recognition is sub divided into speaker verification and speaker identification. Speaker verification is used to confirm (accept/reject) to an identity claimed by a speaker. For example it is useful in case of access control where voice is used as biometric feature. In speaker identification, a speaker is selected from known speaker's set for which speech sample (speaker model) is previously available. This system may also be able to take decision whether the acquired new speech signal matches with the existing stored speech models or this is an unknown speaker [18].

Though, there are numerous researches going on in the area. To development of a robust and accurate speaker identification system is still a big challenge. Lots of efforts have been made to improve the recognition system performance but the progress still needs improvement [19-23]. The framework propose by the researcher is a generic framework for speaker recognition system. It includes complete procedure to design a speaker recognition system. It provides with several choices for its implementation it is implementer's choice to select medium through which speakers signal acquired, to decide on the size of segment of speech signal; to choose the suitable feature extraction technique; to select modeling technique and to select technique for matching score. The framework is applicable for both recognition system i.e. text-dependent/text-independent automatic speaker recognition system.

The framework provides a methodology for speaker recognition. During enrolment speech signal is acquired for each speaker to extract speech features. From speech features equal number of speaker models is created for every registered candidate to create voice training database. Recognition process is done by matching the utterance with each registered speaker's models/template. Speaker recognition system selects the template whose match score matches most closely to the model available in training database. The framework integrates the whole process involved in the speaker recognition system. The purpose of the proposed framework is to identify individual's long utterance of the speech with the help of prosodic statistics. The framework provides both static and dynamic characteristics for creating speaker model.

The Framework

A framework can be defined as, the structure (real or theoretical) supposed to help as a guide for developing software or hardware or anything that uses it to produce something valuable [24].The proposed framework for speaker recognition system is universal in nature i.e. it can be used by anyone to design a recognition system.

Premises

A framework is a structured or a logical way to organize a process to achieve anything [25]. It is reusable set of component which is used to manage system. As every proposal may possess its own premises the framework for improving the performance of speaker recognition system has the following assumptions:

- The framework is designed for text-dependent/text-independent speaker recognition system.
- One can choose a subset of the available techniques and methodology for the development of recognition system.
- To make a robust and accurate system one should keep into account certain things like channel mismatch, background noise, recording conditions etc.
- The framework is not explicitly discussing about noise removal/addition. As it is assumed that it is the part of feature extraction process.

In order to make system more robust one should try to select the speech features which are more robust against noise and have useful information related to speaker. Also a modelling technique which is suitable according to a particular application may be chosen to create speakers model for training and testing voice database.

Guidelines

The aim of the framework is to identify as well as verify speakers. To fulfil the aim, the proposed framework for speaker recognition system has the following phases. These are namely:

- Sample collection and preparation
- Feature extraction
- Model creation
- Feature Matching
- Decision
- Performance evaluation

In the first phase i.e., during sample collection, speaker's voice is collected through the available communication mediums. In the next phase i.e. during sample preparation collected voice samples are broken into small pieces for further process. Further, selection of speech features, feature selection method, modeling technique and matching method is performed. During model creation, speaker voice models for training and testing purpose are created using modeling technique selected in previous phase. Matching is performed by comparing a voice sample with the sample in the database. On the basis of match score it is decided that identity is found or not.

Framework Development

The goal of developing the speaker recognition framework is to recognizing a speaker either in closed-set or open-set. It is supposed that the speech segment is taken by either known or unknown speaker. Proposed framework is shown in the Figure 1. All the phases in the speaker recognition have been discussed in detail in the following subsections:
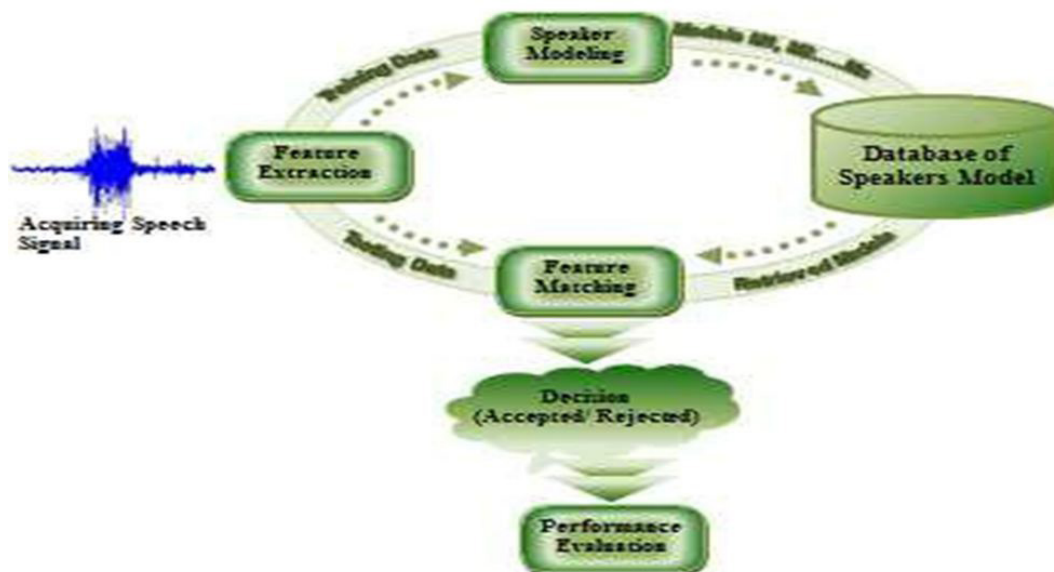

Figure 1: Framework for speaker recognition system

Framing/Windowing: After the acquisition of speech signal framing is performed. During frame blocking the signal splits into equal frames of length N. After framing windowing is performed. There are many types of windowing such as triangular windowing, rectangular windowing Bartlett, Blackman, Hamming, Hanning, Kaiser, Lanczos and Tukey window functions etc. The simplest is rectangular window (no windowing). This window has no discontinuity at beginning and end of frame [27]. Figure 2 shows the frames and window in a speech signal.

Selection of speech frame is an important task and deciding frame length is an essential parameter for spectral analysis of a speech signal. Generally standard frame length 10-30 milliseconds is used for MFCC [27,28].Size of the window is related such that it should be large enough for adequate frequency resolution and short enough to capture the spectral properties.

Windowing of a speech signal is done to find out effect of spectral artifacts in the framing process [29-31]. For windowing several smoothing windows are used such as Rectangular (none), Hanning, Hamming, Blackman-Harris, Exact Blackman, Blackman, Flat Top etc. Hanning window is use for evaluating transients and its shape like a shape of a half cycle of a cosine wave. Modified

version of hanning window is known as Hamming window. Shape of hamming window is like to a cosine wave [32]. In general Hamming window (it gives better spectral performance) is used for calculate window function of speech signal [24,33]. The Hamming window is defined as [27].

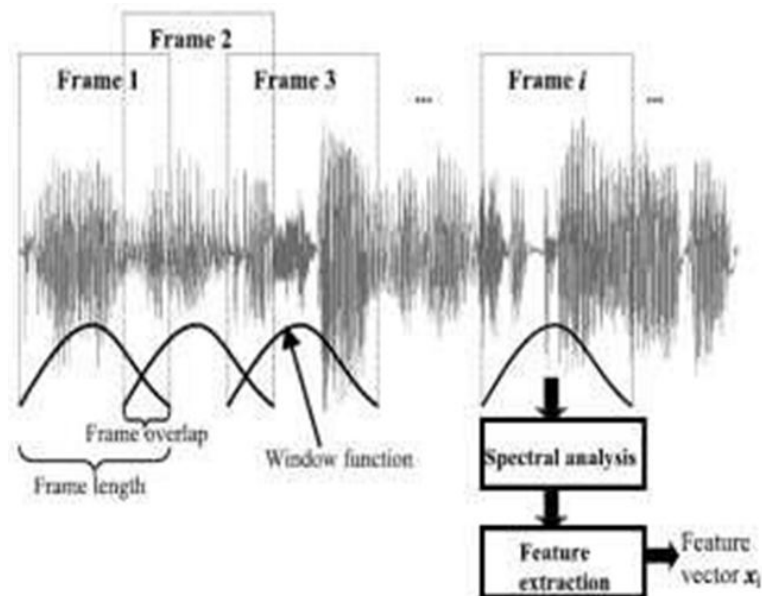$$w[n] = \begin{cases} 0.54 - 0.46\cos\dfrac{2\pi n}{N}, & 0 \le n \le N-1 \\ 0, & \text{otherwise} \end{cases}$$

Figure 2: Frame & window of a speech signal [27]

Speech feature extraction

Feature extraction is the process of converting a raw speech signal into a sequence of acoustic feature vectors which contain the characteristic information about speaker. The following suggestions must be taken into account while selecting speech features [34-36]:

• Speech features should be resistant against voice and channel distortion
• Speech features should not be affected by variation in voice (e.g., by speaker's health or aging)
• Feature extracted from speech signal should be easy to estimate
• Speech features should be able to maintain high inter-speaker discrimination and less of intra-speaker variability
• Feature extraction method should be difficult to mimic against speech of imposters

The above mentioned characteristics of a speech feature extraction methodology are difficult to achieve in individual feature extraction technique. Since some features such as fundamental frequency (F0) is robust against noise but required long speech segments hence prosodic features are individually capable to build speaker recognition system [34].

Selection of Speech Features

To select appropriate speech features and methods to extract selected speech features is known as feature selection and feature extraction [26]. Feature extraction is the main task of speaker recognition or speech recognition. It is well known that speech signal is a complex signal which contains several features of voice. To recognize a speaker it is necessary to extract speech features of the speaker. These features are categorized as physiological and behavioral speech features of the speaker [37]. Physiological features are such as hand geometry, finger print, iris, retina, and face, DNA etc. and behavioral features such as voice, Gait and typing rhythm etc. The next section discuss about the criteria of speech features selection [38].

Criteria for speech feature selection: To develop a robust speaker recognition system there must be some specific criteria to select properties of speech signal after framing/windowing. To create a good system, the speech features selected should possess the following properties [39]:

• It should be robust against noise and distortion
• It should be occur naturally and frequently
• It should not be affected by speaker's health or age

• It should be difficult to mimic
• It should be not affected by speaker variability
• It should not affected by channel mismatch

A single speech feature has not fulfilled the entire above mentioned prerequisite. Therefore selection of speech features depends on the application of authentication such as security level, environmental noise, size of database, type of speakers (co-operative/non co-operative) etc., For example spectral features are extremely discriminative, they calculated from very short segments of speech signal (1-5 sec) but easily affected by noise. F0 statistics require large amount of speech data but robust against channel and noise (technical) mismatches [39].

Analysis and categorization of speech features: A speaker recognition system can be designed by using one or more (combination) of the speech features. The selection of features will depend on the requirement of the system. For example short-term spectral features are highly discriminative and they can be reliably measured from short segments (1-5 seconds) but these features are easily affected by noise (when transmitted over a noisy channel) [37,40]. Fundamental frequency (F0) measurements are robust against channel mismatch but require long speech segments and are not discriminative. In addition, selection of speech features basically depends on the environment where the system is to be deployed such as co-operative/non co-operative speakers, security/ convenience balance, database size, amount of environmental noise etc. There are many speech features available for speaker recognition. These features can be categorized as follows:

• Spectral features
• High-level features
• Supra-segmental / Prosodic features
• Source features
• Dynamic features etc.

The type of authentication system will decide ̱which ̒and ̱how many ̒features are to be selected. Table 1 shows the example of each feature type. Spectral feature are in the form of short-term speech spectrum and describe physical characteristic of vocal tract. High-level features represent the symbolic information e.g. characteristic of word usage. Supra-segmental or prosodic features represent speaking rate, rhythm, intonation pattern and stress etc. Source features represent glottal voice source features. Dynamic features are related with time evolution of spectral features.

Feature type	Examples
Spectral features	MFCC, LPCC, LSF
	Long-term average spectrum (LTAS)
	Formant frequencies and bandwidths
High-level features	Idiosyncratic word usage
	Pronunciation
Supra-segmental/ Prosodic features	F0 contours
	Intensity contours
	Micro-prosody
Source features	F0 mean
	Glottal pulse shape
Dynamic features	Delta features
	Modulation frequencies
	Vector autoregressive coefficients

Table 1: List out the categories of the speech features along with their examples

Speaker Modeling and Database Creation

Speaker modeling involves two phases training phase and testing phase. Speaker models are created by using specific speech feature. There are two types of model creation methods; stochastic models and template models. These modeling methods are used for constructing speaker models using the features extracted from the speech signal. In this phase, a speech model based on the extracted features of speech signal is created and stored. During authentication of a speaker, matching algorithm compares the models of the claimed user. In stochastic models, pattern matching is probabilistic and the result is measurement of likelihood, or conditional probability of the given observational model. The template method can be dependent or independent of time. VQ modeling is an example of time-independent template model. Time-dependent template model are more complicated because it must accommodate human speaking rate variability [26,27]. Stochastic models are more flexible and result is more reliable due to probabilistic likelihood score as compared to template models.

Characteristics of a good speaker model: A good speaker model is one which can rapidly able to adapt voice differences. During construction of speaker models, a number of design goals need to follow. It is very difficult to achieve these goals. However, by choosing a good speaker model these goals can be achieved. Figure 3 shows the characteristics of good speaker modeling technology. Following are the characteristics of a good speaker model [7,41]:

• Consistent within speaker: speaker model for a particular speaker should avoid speker's voice inrta-variability [41]. It should be able to neglect differences occurred in the voice of same speaker over the time.
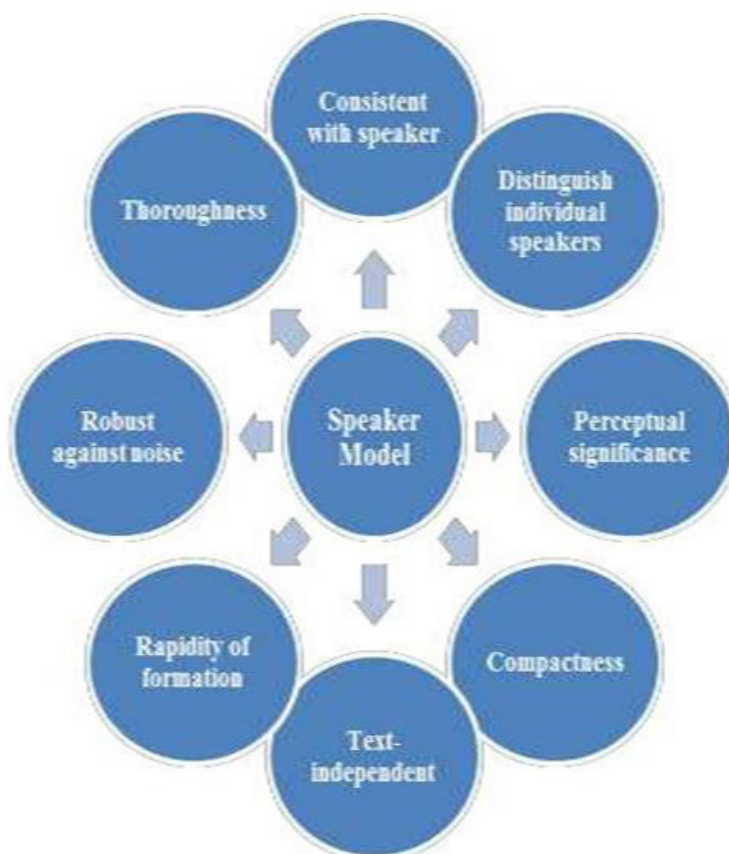

Figure 3: Characteristic of good speaker model

• **Distinguish individual speakers:** Individual speaker should be represented distinctly [41].
• **Perceptual significance:** those speakers voice that are nearby the suspected should be widely separated either similar or different while judged [41].
• **Compactness:** Compactness should be achieved only if the models have low dimension. This allows the model and the application which is using it, bring together new speaker models covered by training speaker models [41].
• **Text-independent:** It should be text-independent. There is no need to utter the same phrase or sentence during training and testing phase [41].
• **Rapidity of Formation:** models should be generated as rapidly as possible by using information from speech signal [41].
• **Robust against noise:** Modeling techniques should be robust against noise. For a given speech signal the model should be free from noise [41].
• **Thoroughness:** the model should contain all the required information about speaker to make conscientious decision [41].

The above are the characteristics of a good speaker model which can enable to achieve the goal of developing robust speaker recognition system. The aim is to develop general models of speakers that can be used successfully to a wide variety of applications in the area of speaker recognition.

After modeling, the speaker's models are stored in a database which can be referred while matching.

Feature matching

Matching is the process of comparing the extracted speech features of a person with stored speaker models/templates. The comparison quantifies the similarity between the voice (record for identification) and a speaker model from voice database. Selection of a matching technique is an important task. There are many prevalent classifications / matching technique such as Hidden Markov Models (HMM), Vector Quantization (VQ) and Dynamic Time Warping (DTW) [42]. Speaker model is used

for comparing with a particular input signal. Speaker models are stored in a database. Two kinds of database are there: training database and testing database. Model comparison method involves the following:

- Matches training database of the target speaker with his/her testing database.
- Match score is calculated.
- If the match score is greater than or equal to the threshold then the target speaker is accepted by the system otherwise rejected.

During matching created speaker models, may be speaker-dependent in case of speaker recognition system and speaker-independent in case of forensic speaker recognition system. There is predefined specific criterion for creating speaker models [17].

Decision Phase

The decision process depends on the kind of the system i.e. closed-set system or open-set system. In case of closed-set identification system, the decision can be made by selecting that model which is most similar to the test sample speech signal. In case of open-set, system requires a threshold to verify that similarity is valid. As there may be chances that a system rejects a registered speaker, hence cost of making an error is considered in the decision process. For example, in case of a bank to allow an imposter will prove to be more costly than to reject a true customer. The Decision is determined by particular matching and modelling algorithms. For example, in case of template matching decision is given by computed distance between speakers models whereas in stochastic matching calculated result is based on the computed probabilities [43-46]. Figure 4 represents the decision process of a speaker recognition system.

Performance evaluation phase

The framework also provides the way to measure performance of the speaker recognition system. For the purpose, various available metrics can be used. The commonly used metrics are the False Acceptance Rate (FAR) and False Rejection Rate (FRR). For making speaker recognition system more accurate, FAR is should be minimum [47]. In addition, performance of speaker identification system is also decided by Equal Error Rate (EER). It is the most common method used to evaluate system performance. EER is a point where probability of False Acceptance (FA) and probability of False Rejection (FR) both are equal [46].

Framework Significance

The proposed framework has the following significance:
1. The proposed framework will help to develop a speaker recognition system using speech signal.
2. The proposed framework is applicable for both; text-dependent or text-independent speaker recognition system.
3. It gives the flexibility to choose any type of speech features for developing speaker recognition system.
4. The proposed framework is independent from the specific modelling technique.
5. During pattern matching there is no compulsion for selecting a specific pattern matching algorithm or technique.
6. The proposed framework is universal framework which includes all the necessary steps which is required to develop a speaker identification/recognition system.
7. In every step, the proposed framework facilitates the user with the option to go back to the previous phase, if required.

Limitation of Developed Framework

The proposed framework, however, has been designed with the following limitations:
• Though it provides step by step solution to develop a speaker recognition system but it may not include exhaustive steps to be performed at any phase.
• In order to improve performance, the proposed framework requires large voice data is required and a broad training has to be performed in advance.

Conclusion

In this paper authors proposed a framework for development of a robust and more accurate speaker recognition system. The proposed framework provides a method for speaker recognition. Speaker enrollment is the first step towards creating speaker recognition system, in this each registered candidate provides a set of utterances. For training, equal number of speech templates is created for each speaker (for individual speaker templates need not be of same duration) but for testing it may be vary. Speaker's template set is used as a model for the individual. During matching process, system selects the template whose match score is more closed to the test template. The proposed framework explains the overall process involve in the development of speaker recognition system. The proposed framework has a credential for long-term statistics. In the framework either static or dynamic speaker characteristics of speech features are used for speaker recognition. The proposed framework has the potential to additional embodiment and modification would be possible as per the requirement.
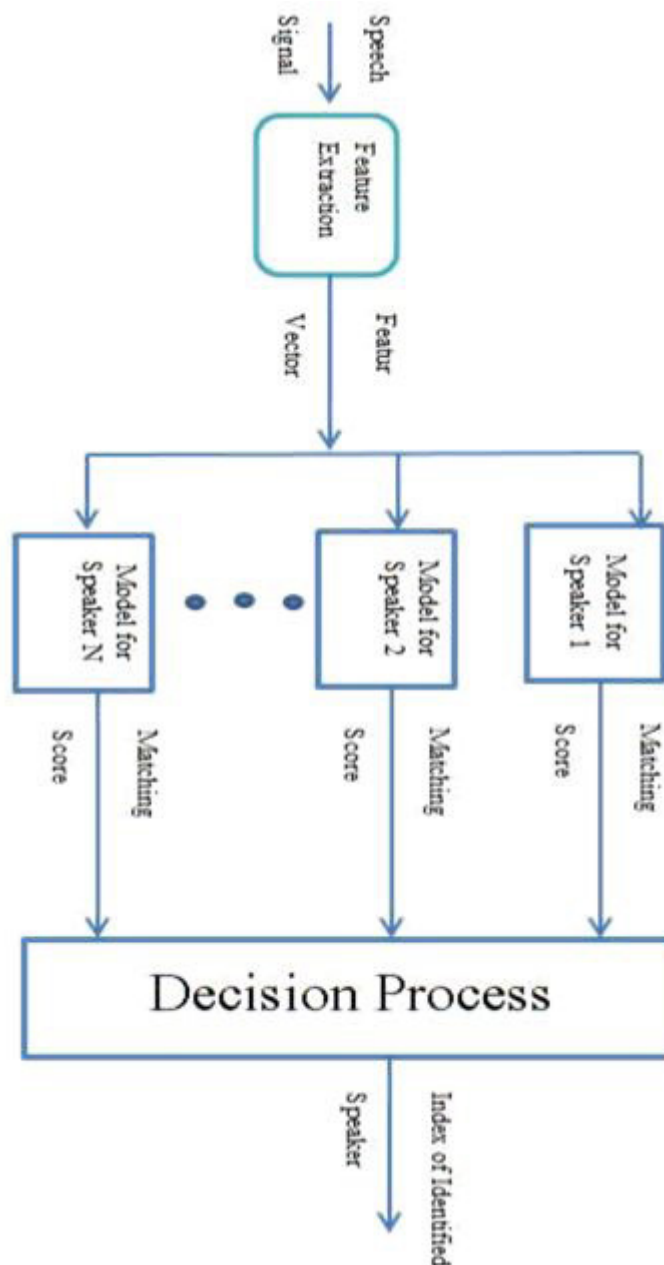
Figure 4: Decision process of speaker identification

References

1. Alsulaiman M, Mahmood A, Muhammad G (2016) Speaker recognition based on Arabic phonemes. Speech Commu 86: 42-51.

2. Singh N, Agrawal A, Khan RA (2017) Automatic Speaker Recognition: Current Approaches and Progress in Last Six Decades. Global J Enterprise Inf System 9: 38-45.

3. Sarkar A, Sahidullah M, Tan ZH, Kinnunen (2017) Improving Speaker Verification Performance in Presence of Spoofing Attacks Using Out- of-Domain Spoofed Data. Interspeech 1-6.

4. Biometry (2014) Biometry, Switzerland.

5. Khosravani A, Homayounpou MM (2017) A PLDA approach for language and text independent speaker recognition. Comp Speech Lang 45: 457-74.

6. Singh N, Khan RA (2014) Extraction of Prosodic Features for Speaker Recognition Technology and Voice Spectrum Analysis. Int J Sci Eng Res 5: 600-5.

7. Singh N, Agrawal A, Khan RA (2017) Gaussian Mixture Model: A Better Modeling Technique for Speaker Recognition. Proceedings of (DIAL-2017) National Conference on Digital India-Altering Landscape 14-8.

8. Ayadi ME , Abdel-Karim SOH, Abdel-Naby A, Elgend OA (2017) Text-independent speaker identification using robust statistics estimation. Speech Commun 92: 52-63.

9. Sandouk U (2012) Speaker Recognition Speaker Diarization and Identification. University Manchester School Compute Science, England.

10. Shaver CD, Acken JM (2017) The development of text-independent speaker recognition technology. J Electr Eng 1-8.

11. Cavazza M, Ciaramella A (1998) Device   for speaker's verification.

12. International Banking (2013) Voice Biometric Technology in Banking. Barclays, United Kingdom.

13. Speaker Identification (2014) Archived from the original on August 15, 2014.

14. Hautamaki RG, Sahidullah Md, Hautamaki V, Kinnunen T (2017) Acoustical and perceptual study of voice disguise by age modification in speaker verification. Speech Commu 95: 1-37.

15. Kollewe J (2016) HSBC rolls out voice and touch ID security for bank customers Business. Guardian.

16. Furui S (1991) Speaker-dependent-Feature Extraction, Recognition and Processing Techniques. Speech Commu 10: 505-20.

17. Speaker recognition (2010) Speaker recognition.

18. Nakasone H (2003) Automated speaker recognition in real world conditions: Con-trolling the uncontrollable. Federal Bureau of Investigation, USA.

19. Kent RD & Read C (1992) The acoustic analysis of speech. Singular Publishing Group, SanDiego, CA, USA.

20. Rose P (2002) Forensic Speaker Identification. London, England.

21. Wolf J (2005) Efficient acoustic parameters for speaker recognition. J Acoustical Soc Am 51: 2044-56.

22. Summerfield R, Dunstone T, Summerfield C (2008) Speaker Verification in a Multi-Vendor Environment.

23. Toh KA (2003) Fingerprint and speaker verification decisions fusion. IEEE ICIAP Conference 626-31.

24. J. Kerin (2004) Biometric passport demand likely.

25. Talkin D (1995) A robust algorithm for pitch tracking (RAPT). Speech Coding Synthesis 495: 518.

26. Boersma P (1993) Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Inst Phonetic Sci 17: 97-110.

27. Gonzalez J, Lopez-Moreno I, Franco-Pedroso J, Ramos D, Toledano D, et al. (2010) ATVS-UAM NIST SRE 2010 System Description. 415-8.

28. Scheffer N, Vogt R (2010) On the Use of Speaker Superfactors for Speaker Recognition. In Proc ICASSP 4410-3.

29. Schwartz R, Shen W, Campbell J, Granville R (2009) Measuring Typicality of Speech Features in American English Dialects: Towards Likelihood Ratios in Speaker Recognition Casework, 5th European Academy of Forensics Science, Glasgow, Scotland.

30. Chen N, Shen W, Campbell J, Torres-Carrasquillo P (2011) Informative Dialect Recognition Using Context-Dependent Pronunciation Modeling. IEEE ICASSP Conference, Prague Czech Republic.

31. Shaver C, Acken J (2010) Effects of Equipment Variation on Speaker Recognition Error Rates. IEEE ICASSP Conference Dallas, Texas, USA.

32. LabVIEW (2011) Characteristics of Different Smoothing Windows.

33. Rodriguez DS (2008) Text-Independent Speaker Identification. Comp Sci Electr 1-121.

34. Gish H, Schmidt M (1994) Text independent speaker identification. IEEE Signal Processing Magazine 11: 18-32.

35. Reynolds DA (2002) An overview of automatic speaker recognition technology. ICASSP 4072-5.

36. Dunn R, Quatieri T, Reynolds D, Campbell J (2001) Speaker recognition from coded speech in matched and mismatched conditions. In Proc Speaker Odyssey 115-120.

37. Kohler T (2010) The NIST 2010 Speaker Recognition Evaluation. Interspeech, Chiba, Japan.

38. Singh N, Khan RA, Shree R (2014) Equal Error Rate and Audio Digitization and Sampling Rate for Speaker Recognition System. Am Sci Pub 20: 1085-8.

39. Kinnunen TH (2005) Optimizing Spectral Feature Based Text-Independent Speaker Recognition. University of Joensuu Computer Science Dissertations 12: 1-156.

40. Witbrock MJ (1996) Learning Models of Speaker Variation. School of Computer Science, Carnegie Mellon University, Pittsburgh, USA.

41. Reynolds D, Quatieri T, Dunn R (2000) Speaker verification using adapted gaussian mixture models. Digital Signal Processing 10: 19-41.

42. Vogt R, Sridharan S (2004) Frame-weighted bayes factor scoring for speaker verification. In Proc 10th Australian Int. Conf. on Speech Sci Technol Sydney, Australia.

43. Yegnanarayana B, Kishore SP (2002) AANN: an alternative to GMM for pattern recognition. Neural Netw 15: 459-69.

44. Kemp T, Schmidt M, Westphal M, Waibel A (2000) Strategies for automatic segmentation of audio data. IEEE ICASSP Conference 1423-6.

45. Reynolds DA, Quatieri TF, Dunn R (2000) Speaker verification using adapted Gaussian mixture models. Dig Signal Process 10: 19-41.

46. Jin Q (2007) Robust Speaker Recognition. Language Technologies Institute School of Computer Science Carnegie Mellon University, Pittsburgh, USA.

47. Furui S (1996) An overview of speaker recognition technology. Automatic Speech Speaker Recog 31-56.