Speaker Recognition As A Form Of Biometric

ArunabhaBasak, A J Alex, Jishnuprasad Menon, Aditya Deshpande

Department of Computer Science and Engineering
SRM Ramapuram, Chennai, India

ABSTRACT

Speaker recognition is the process of recognizing automatically who is speaking on the basis of individual information included in speech waves. This technique uses the speaker's voice to verify their identity and provides control access to services such as voice dialling, database access services, information services, voice mail, security control for confidential information areas, remote access to computers and several other fields where security is the main area of concern. Speech is a complicated signal produced as a result of several transformations occurring at several different levels: semantic, linguistic, articulatory, and acoustic. Differences in these transformations are reflected in the differences in the acoustic properties of the speech signal. Besides there are speaker related differences which are a result of a combination of anatomical differences inherent in the vocal tract and the learned speaking habits of different individuals. In speaker recognition, all these differences are taken into account and used to discriminate between speakers. The forthcoming chapters describe how to build a simple and representative automatic speaker recognition system. Such a speaker recognition system helps in the basic purpose of speaker identification which forms a formidable domain in the field of speaker recognition. The system designed has potential in several security applications. Examples may include, users having to speak a PIN (Personal Identification Number) in order to gain access to the laboratory they work inor having to speak their credit card number over the telephone line to verify their identity. By checking the voice characteristics of the input utterance, using an automatic speaker recognition system similar to the one that we will describe, the system is able to add an extra level of security.

Keywords—*Speaker recognition, biometrics, security, mel-cepstral coefficient, DT*

1.INTRODUCTION

At the most basic, biometrics can be best explained by breaking down the word: bio, as in biological; and metric, as in measurement. That is to say, biometrics are biological measurements. Thanks to the unique nature of many of these measurements, biometrics are particularly suited for identification. Visually, face is one most important feature, other unique features, such as finger-prints, iris, are often used. Another way to identify a person is from the acoustic fact that each person’s voice is different, this forms one area of speech processing, automatic speaker recognition. For the past few decades, many solutions have come out. Although many of them have a very good performance, none of them are perfect. The difficulty of this is caused by several reasons based on the nature of speech. First, voice is not so unique as visual cues such as finger-prints, some people’s voice are very similar which results in the difficulty of separation. Another difficulty is that for speaker recognition, we are not only dealing with the variation between people, but also the huge variability of voice from one person, this includes the large number of phonemes one can utter as well as the variation when speaking in different emotions.

The forthcoming chapters describe how to build a simple and representative automatic speaker recognition system. Such a speaker recognition system helps in the basic purpose of speaker identification which forms a formidable domain in the field of speaker recognition. The system designed has potential in several security applications. Examples may include, users having to speak a PIN (Personal Identification Number) in order to gain access to the laboratory they work in or having to speak their credit card number over the telephone line to verify their identity. By checking the voice characteristics of the input utterance, using an automatic speaker recognition system similar to the one that we will describe, the system is able to add an extra level of security.

Speaker Recognition can be divided by two ways. One way is to divide it into speaker verification and speaker identification. For speaker verification, the test is based on the claimed identity and a decision for accepting or rejecting is made. For speaker identification, there is no claim of identity, the system chooses the speaker from the database or in open set system, the identity can be unknown. Another way is to divide speaker recognition based on the text used for test. It can be text-dependent, which use the fixed or prompt sentence for testing, or text-independent, in which any utterance can be used. In this project, the focus is on the text-independent speaker identification in closed set.

2.EXISTING SYSTEM

The function of a biometric technologies authentication system is to facilitate controlled access to applications, networks, personal computers (PCs), and physical facilities. A biometric authentication system is essentially a method of establishing a person's identity by comparing the binary code of a uniquely specific biological or physical characteristic to the binary code of an electronically stored characteristic called a biometric. The defining factor for implementing a biometric authentication system is that it cannot fall prey to hackers; it can't be shared, lost, or guessed. Simply put, a biometric authentication system is an efficient way to replace the traditional password-based authentication system. While there are many possible biometrics, at least eight mainstream biometric authentication technologies have been deployed or pilot tested in applications in the public and private sectors.

A biometric technology that requires an individual to make direct contact with an electronic device (scanner) will be referred to as a contact biometric. Given that the very nature of a contact biometric is that a person desiring access is required to make direct contact with a 9 electronic device in order to attain logical or physical access. Because of the inherent need of a person to make direct contact, many people have come to consider a contact biometric to be a technology that encroaches on personal space and to be intrusive to personal privacy.

A contactless biometric can either come in the form of a passive (biometric device continuously monitors for the correct activation frequency) or active (user initiates activation at will) biometric. In either event, authentication of the user biometric should not take place until the user voluntarily agrees to present the biometric for sampling. A contactless biometric can be used to verify a person's identity and offers at least two dimension that contact biometric technologies cannot match. A contactless biometric is one that does not require undesirable contact in order to extract the required data sample of the biological characteristic and in that respect a contactless biometric is most adaptable to people of variable ability levels.

3.PROPOSED SYSTEM

Speaker recognition is a commonly used biometric today in most of the commercialization that has taken place for control of access to information services or user accounts on computers. Speaker recognition offers the ability to replace or augment the personal identification numbers and passwords with something that cannot be stolen or lost. There are two main factors [Rey02], that make speaker recognition a compelling biometric; (1) Speech is natural signal to produce that is not considered threatening by the users to provide, and (2) the telephone system provides a familiar network of sensors for obtaining and delivering the speech signal.

The applications of speaker recognition technology are quite varied and continually growing. This technique makes it possible to use the speaker's voice for verification of their identity and thereafter enable the control access to services such as voice dialing and voice mail, tele-banking, telephone shopping, database access related services, information services, security control for confidential information areas, forensic applications, and remote access to computers. Speaker recognition technology is expected to create a host of new services that will make our daily lives more convenient.

Speaker identification is further divided into two subcategories, text dependent and text independent speaker identification. Text-dependent speaker identification differs quite from text-independent as in the aforementioned identification is done on the voice instance of a specific word, whereas in the latter the speaker can say anything. Our project will consider only the text-dependent speaker identification category.

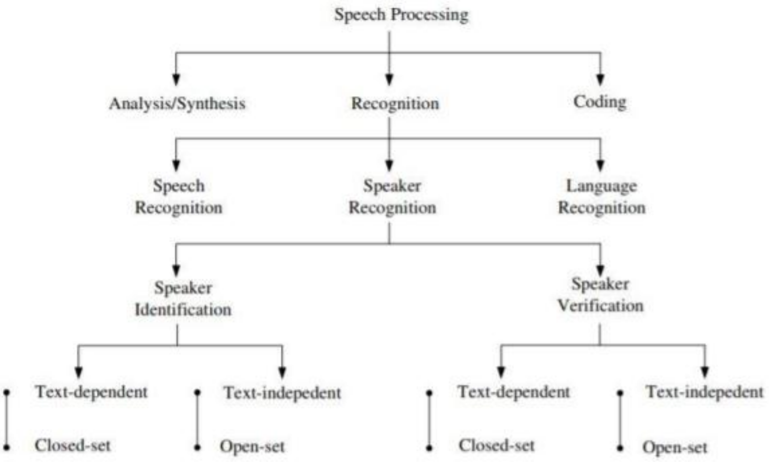

Fig 2.1 Speech Processing Taxonomy

4.SYSTEM ARCHITECTURE

The data flow diagram gives the detailed process flow between the various modules of the system; a module is a separate unit of software or hardware. Typical characteristics of modular components include portability, which allows them to be used in a variety of systems, and interoperability, which allows them to function with the components of other systems. The software and hardware requirements include all data, functional and behavioural requirements of the software under production or development it also includes the hardware specifications required by the project. The software and hardware requirements of the system help in analysing the basic architecture of the system.

Speaker recognition, which involves two applications: speaker identification and speaker verification, is the process of automatically recognizing who is speaking on the basis of individual information included in speech waves. This technique makes it possible to use the speaker's voice to verify their identity and control access to services such as voice dialing, banking by telephone, telephone shopping, database access services, information services, voice mail, security control for confidential information areas, and remote access to computers.

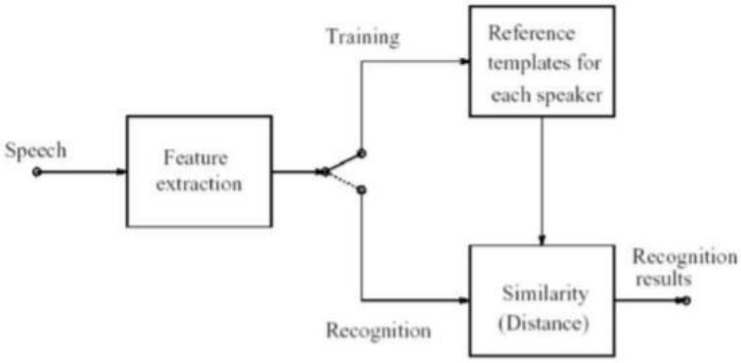

Fig 4.1 Data Flow Diagram

a.Speaker Verfication

Speaker verification (SV) is the process of determining whether the speaker identity is who the person claims to be. Different terms which have the same definition as SV could be found in literature, such as voice verification, voice authentication, speaker/talker authentication, talker verification. It performs a one-to-one comparison (it is also called binary decision) between the features of an input voice and those of the claimed voice that is registered in the system.

Figure below shows the basic structure of SV system (SVS). There are three main components: Front-end Processing, Speaker Modelling, and Pattern Matching. Front-end processing is used to highlight the relevant features and remove the irrelevant ones.

After the first component, we will get the feature vectors of the speech signal. Pattern Matching between the claimed speaker model registered in the database and the imposter model will be performed then. If the match is above a certain threshold, the identity claim is verified. Using a high threshold, system gets high safety and prevents impostors to be accepted, but in the mean while it also takes the risk of rejecting the genuine person, and vice versa.

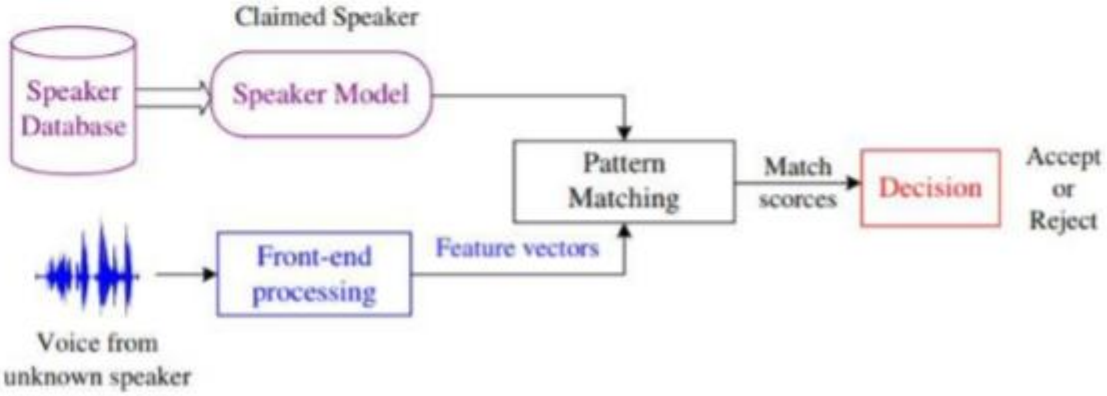

Fig 4.2 Speaker Verification

b.Speaker Identification

Speaker identification (SI) is the process of finding the identity of an unknown speaker by comparing his/her voice with voices of registered speakers in the database. It's a one-to-many comparison [3]. The basic structure of SI system (SIS) is shown in the figure below. We notice that the core components in SIS are the same as in SVS. In SIS, M speaker models are scored in parallel and the most-likely one is reported. In different situations, speaker recognition is often classified into closed-set recognition and open-set recognition. Just as their names suggest, the closed-set refers to the cases that the unknown voice must come from a set of known speakers; and the open-set means unknown voice may come from unregistered speakers, in which case we could add 'none of the above' option to this identification system.

Moreover, in practice speaker recognition systems could also be divided according to the speech modalities: text-dependent recognition, text-independent recognition. For text-dependent SRS, speakers are only allowed to say some specific sentences or words, which are known to the system. In the bargain, the text-dependent recognition is sub-divided into fixed phrase and prompted phrase. On the contrary, as for the text-independent SRS, they could process freely spoken speech, which is either user selected phrase or conversational speech. Compared with text-dependent SRS, text-independent SRS are more flexible, but more complicated.

The core components in SIS are the same as in SVS. In SIS, M speaker models are scored in parallel and the most-likely one is reported, and consequently decision will be one of the speaker's ID in the database or will be 'none of the above' if and only if the matching score is below some threshold and it's in the case of an open-set SIS.

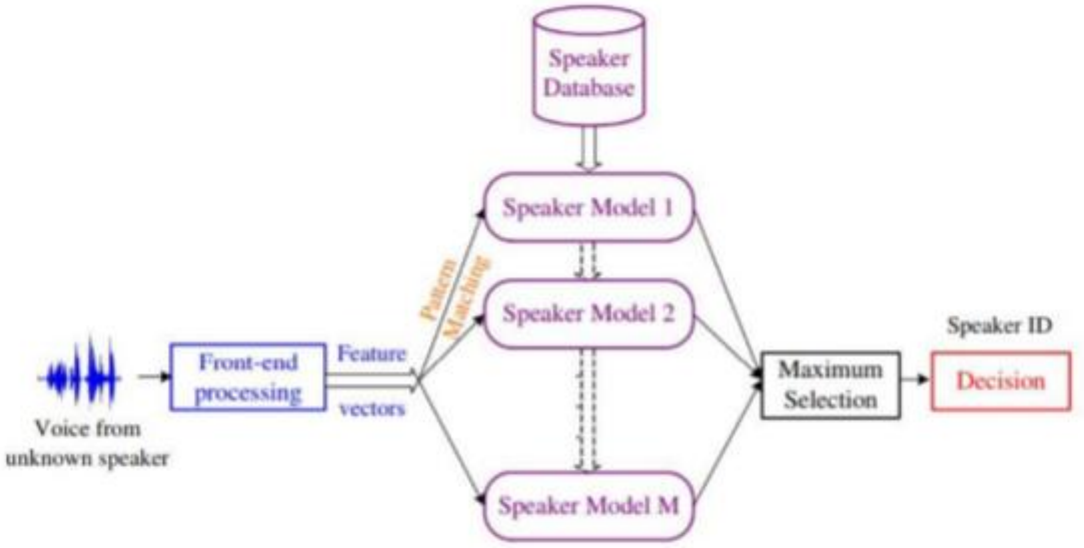

Fig 4.3 Speaker Identification

c.Phases of Speaker Identification

In identification phase, the same method for extracting features as in the first phase is used for the incoming speech signal, and then the speaker models getting from enrollment phase are used to calculate the similarity between the new speech signal model and all the speaker models in the database. In closed-set case the new speaker will be assigned to the speaker ID which has the maximum similarity in the database. Even though the enrollment phase and identification phase are working separately, they are still closely related. The modeling algorithms used in the enrollment phase will also work on the identification algorithms.

Enrollment phase is to get the speaker models or voiceprints to make a speaker database, which could be used later in the next phase, i.e. identification phase. The front-end processing and speaker modeling algorithms in both phases of SIS (SVS) should be consistent respectively.

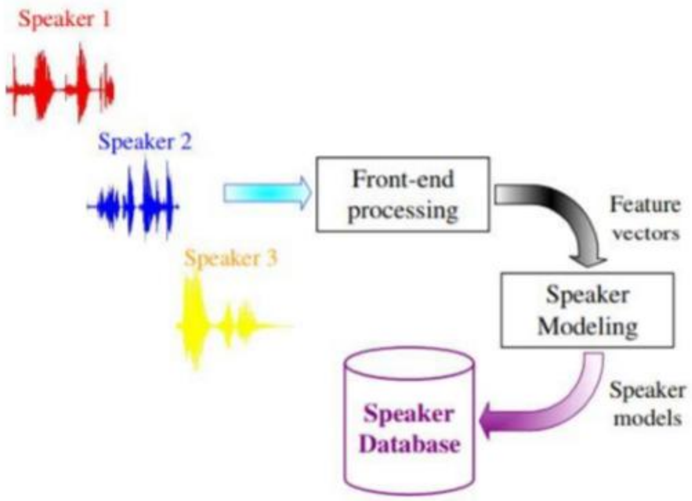

Fig 4.4 Phases of speaker identification

5.SYSTEM IMPLEMENTATION

Our thesis work is based on identifying an unknown speaker given a set of registered speakers. Here we have assumed the unknown speaker to be one of the known speakers and tried to develop a model to which it can best fit into. In the first step of generating the speaker recognition model, we went for feature extraction using two processes given below:-

1. Cepstral coefficients
2. Mel Frequency Cepstral Coefficients

These features act as a basis for further development of the speaker identification process. Next we went for feature mapping using the following algorithm:

1. Vector Quantization using LBG (VQLBG)
2. Dynamic Time Warping (DTW)
3. Gaussian Mixture Modeling (GMM)

A. Matlab

MATLAB (matrix laboratory) is a multi-paradigm numerical computing environment. A proprietary programming language developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, C#, Java, Fortran and Python.

Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the MuPAD symbolic engine, allowing access to symbolic computing abilities. An additional package, Simulink, adds graphical multi-domain simulation and model-based design for dynamic and embedded systems. The MATLAB application is built around the MATLAB scripting language. Common usage of the MATLAB application involves using the Command Window as an interactive mathematical shell or executing text files containing MATLAB code. MATLAB supports developing applications with graphical user interface (GUI) features. MATLAB includes GUIDE (GUI development environment) for graphically designing GUIs. It also has tightly integrated graph-plotting features. MATLAB can call functions and subroutines written in the programming languages C or Fortran.[27] A wrapper function is created allowing MATLAB data types to be passed and returned. The dynamically loadable object files created by compiling such functions are termed "MEX-files" (for MATLAB executable). Since 2014 increasing twoway interfacing with Python is being added.

Libraries written in Perl, Java, ActiveX or .NET can be directly called from MATLAB, and many MATLAB libraries (for example XML or SQL support) are implemented as wrappers around Java or ActiveX libraries. Calling MATLAB from Java is more complicated, but can be done with a MATLAB toolbox which is sold separately by MathWorks, or using an undocumented mechanism called JMI (Java-to-MATLAB Interface) (which should not be confused with the unrelated Java Metadata Interface that is also called JMI). Official MATLAB API for Java was added in 2016. As alternatives to the MuPAD based Symbolic Math Toolbox available from MathWorks, MATLAB can be connected to Maple or Mathematica. Libraries also exist to import and export MathML.

6.CONCLUSION

Automatic speaker recognition is the use of a machine to recognize a person from a spoken phrase. Speaker-recognition systems can be used to identify a particular person or to verify a person’s claimed identity. Speech processing, speech production, and features and pattern matching for speaker recognition were introduced. Recognition accuracy was shown by coarse-grain ROC curves and fine-grain histograms revealed the wolves and sheep of two example systems. Speaker recognition systems can achieve 0.5% equal error rates at the 80% confidence level in the benign real-world conditions considered here.

Recent advances in biometric technology have resulted in increased accuracy at reduced costs, biometric technologies are positioning themselves as the foundation for many highly secure identification and personal verification solutions. Today's biometric solutions provide a means to achieve fast, user friendly authentication with a high level of accuracy and cost savings. Many areas will benefit from biometric technologies. Highly secure and trustworthy electronic commerce, for example, will be essential to the healthy growth of the global Internet economy. Many biometric technology providers are already delivering biometric authentication for a variety of web-based and client/server-based applications to meet these and other needs. Continued improvements in technology will bring increased performance at a lower cost.

Currently, there exist a gap between the number of feasible biometric projects and knowledgeable experts in the field of biometric technologies. This is however, changing as studies and curriculum associated to biometric technologies are starting to be offered at more colleges and universities. A method of closing the biometric knowledge gap is for knowledge seekers of biometric technologies to participate in biometric discussion groups and biometric standards committees. The solutions only need the user to possess a minimum of require user knowledge and effort. A biometric solution with minimum user knowledge and effort would be very welcomed to both the purchase and the end user. But, keep in mind that at the end of the day all that the end users care about is that their computer is functioning correctly and that the interface is friendly, for users of all ability levels. Alternative methods of authenticating a person's identity are not only a good practice for making biometric systems accessible to people of variable ability level. But it will also serve as a viable alternative method of dealing with authentication and enrollment errors.

7.REFERENCES

[1] Campbell, J.P., Jr.; “Speaker recognition: a tutorial” Proceedings of the IEEE Volume 85, Issue 9, Sept. 1997 Page(s):1437 – 1462.

[2] Seddik, H.; Rahmouni, A.; Samadhi, M.; “Text independent speaker recognition using the Mel frequency cepstral coefficients and a neural network classifier” First International Symposium on Control, Communications and Signal Processing, Proceedings of IEEE 2004 Page(s):631 – 634.

[3] Childers, D.G.; Skinner, D.P.; Kemerait, R.C.; “The cepstrum: A guide to processing” Proceedings of the IEEE Volume 65, Issue 10, Oct. 1977 Page(s):1428 – 1443.

[4] Roucos, S. Berouti, M. Bolt, Beranek and Newman, Inc., Cambridge, MA; “The application of probability density estimation to text-independent speaker identification” IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '82. Volume: 7, On page(s): 1649- 1652. Publication Date: May 1982.

[5] Castellano, P.J.; Slomka, S.; Sridharan, S.; “Telephone based speaker recognition using multiple binary classifier and Gaussian mixture models” IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 Volume 2, Page(s) :1075 – 1078 April 1997.

[6] Zilovic, M.S.; Ramachandran, R.P.; Mammone, R.J “Speaker identification based on the use of robust cepstral features obtained from pole-zero transfer functions”.; IEEE Transactions on Speech and Audio Processing, Volume 6, May 1998 Page(s):260 – 267

[7] Davis, S.; Mermelstein, P, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences” , IEEE Transactions on Acoustics, Speech, and Signal Processing Volume 28, Issue 4, Aug 1980 Page(s):357 – 366

[8] Y. Linde, A. Buzo& R. Gray, “An algorithm for vector quantizer design”, IEEE Transactions on Communications, Vol. 28, issue 1, Jan 1980 pp.84-95.

[9] S. Furui, “Speaker independent isolated word recognition using dynamic features of speech spectrum”, IEEE Transactions on Acoustic, Speech, Signal Processing, Vol.34, issue 1, Feb 1986, pp. 52-59.

[10] Fu Zhonghua; Zhao Rongchun; “An overview of modeling technology of speaker recognition”, IEEE Proceedings of the International Conference on Neural Networks and Signal Processing Volume 2, Page(s):887 – 891, Dec. 2003