Voiceprint Recognition Systems for Remote Authentication-A Survey

Zia Saquib, Nirmala Salam, Rekha Nair, Nipun Pandey
CDAC-Mumbai, Gulmohar Cross Road No.9, Juhu, Mumbai-400049
{saquib, nirmala, rekhap, nipun}@cdacmumbai.in

Abstract

Voiceprint Recognition System also known as a Speaker Recognition System (SRS) is the best-known commercialized forms of voice Biometrics. Automated speaker recognition is the computing task of validating a user's claimed identity using characteristics extracted from their voices. In contrast to other biometric technologies which are mostly image based and require expensive proprietary hardware such as vendor's fingerprint sensor or iris-scanning equipment, the speaker recognition systems are designed for use with virtually any standard telephone or on public telephone networks. The ability to work with standard telephone equipment makes it possible to support broad-based deployments of voice biometrics applications in a variety of settings. In automated speaker recognition the speech signal is processed to extract speaker-specific information. These speaker specific informations are used to generate voiceprint which cannot be replicated by any source except the original speaker. This makes speaker recognition a secure method for authenticating an individual since unlike passwords or tokens; it cannot be stolen, duplicated or forgotten. This literature survey paper gives brief introduction on SRS, and then discusses general architecture of SRS, biometric standards relevant to voice/speech, typical applications of SRS, and current research in Speaker Recognition Systems. We have also surveyed various approaches for SRS..

Keywords: Voiceprint*, SRS, Speaker Recognition Systems, Voice Biometrics, Speech.*

1. Introduction

1.1. Brief Overview of Speaker Recognition

Voice biometrics specifically was first developed in 1970, and although it has become a sophisticated security tool only in the past few years, it has been seen as a technology with great potential for much longer. The most significant difference between voice biometrics and other biometrics is that voice biometrics is the only commercial biometrics that process acoustic information. Most other biometrics is image-based. Another important difference is that most commercial voice biometrics systems are designed for use with virtually any standard telephone or on public telephone networks. The ability to work with standard telephone equipment makes it possible to support broad-based deployments of voice biometrics applications in a variety of settings. In contrast, most other biometrics requires proprietary hardware, such as the vendor's fingerprint sensor or iris-scanning equipment. By definition, voice biometrics is always linked to a particular speaker. The best-known commercialized forms of voice biometrics are Speaker Recognition. Speaker recognition is the computing task of validating a user's claimed identity using characteristics extracted from their voices.

Table 1. Typical applications of speaker recognition systems

Areas	*Specific applications*
Authentication	Remote Identification & Verification, Mobile Bank ing,ATM Transaction, Access Control
Information Security	Personal Device Logon, Desktop Logon, Application Security, Database Security, Medical Records, Security Control for Confidential Information
Law Enforcement	Forensic Investigation, Surveillance Applications
Interactive Voice Response	Banking over a telephone network, Information and Reservation Services, Telephone Shopping, Voice Dialing, Voice Mail

A speaker's voice is extremely difficult to forge for biometrics comparison purposes, since a myriad of qualities are measured ranging from dialect and speaking style to pitch, spectral magnitudes, and format frequencies. The vibration of a user's vocal chords and the patterns created by the physical components resulting in human speech are as distinctive as fingerprints. Voice Recognition captures the unique characteristics, such as speed and tone and pitch , dialect etc associated with an individual's voice and creates a non-replicable voiceprint which is also known as a speaker model or template. This voiceprint which is derived through mathematical modeling of multiple voice features is nearly impossible to replicate. A voiceprint is a secure method for authenticating an individual's identity that unlike passwords or tokens cannot be stolen, duplicated or forgotten.

1.2. Voice Production Mechanism

The origin of differences in voice of different speakers lays in the construction of their articulatory organs, such as the length of the vocal tract, characteristics of the vocal chord and the differences in their speaking habits. An adult vocal tract is approximately 17 cm long and is considered as part of the speech production organs above the vocal folds (earlier called as the vocal chords). As shown in Figure 1.2 (a), the speech production organs includes the laryngeal pharynx (below the epiglottis), oral pharynx (behind the tongue, between the epiglottis and vellum), oral cavity (forward of the velum and bounded by the lips, tongue, and palate), nasal pharynx (above the velum, rear end of nasal cavity) and the nasal cavity (above the palate and extending from the pharynx to the nostrils). The larynx comprises of the vocal folds, the top of the cricoids cartilage, the arytenoids cartilages and the thyroid cartilage. The area between the vocal folds is called the glottis. The resonance of the vocal tract alters the spectrum of the acoustic as it passes through the vocal tract. Vocal tract resonances are called formants. Therefore the vocal tract shape can be estimated from the spectral shape (e.g., formant location and spectral tilt) of the voice signal. Speaker recognition systems use features generally derived only from the vocal tract. The excitation source of the human vocal also contains speaker specific information. The excitation is generated by the airflow from the lungs, which thereafter passes through the trachea and then through the vocal folds. The excitation is classified as phonation, whispering, frication, compression, vibration or a combination of these. Phonation excitation is caused when airflow is modulated by the vocal folds. When the vocal folds are closed, pressure builds up underneath them until they blow apart. The folds are drawn back together again by their tension, elasticity and the Bernoulli Effect. The oscillation of vocal folds causes pulsed stream excitation of the vocal tract. The frequency of oscillation is called the fundamental frequency and it depends upon the length, mass and the tension of the vocal folds. The fundamental frequency therefore is another distinguishing characteristic for a given speaker.

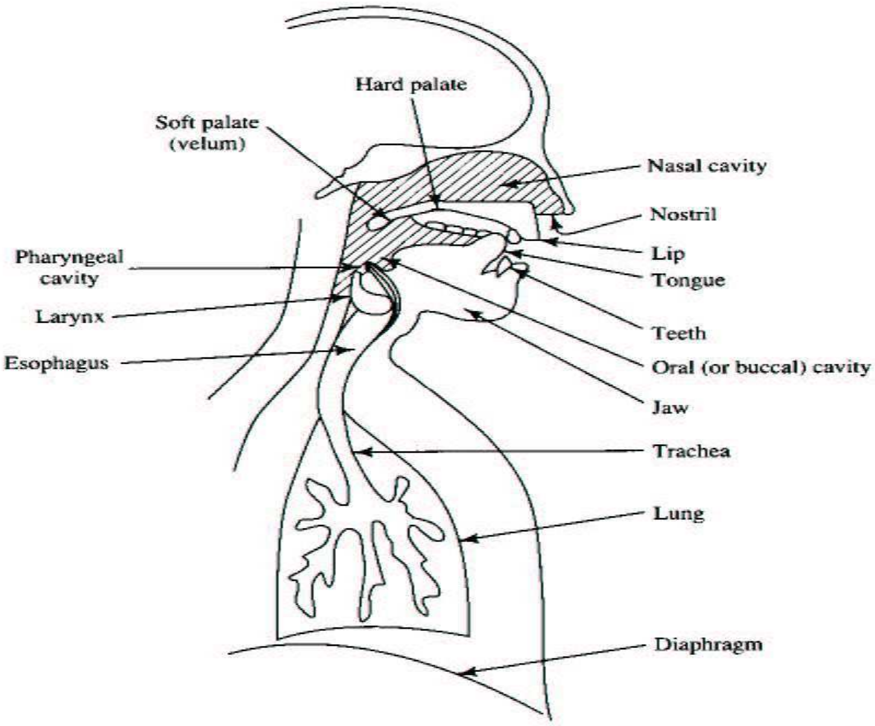

Figure 1. The speech production mechanism [41]

1.3 How the Technology Works

The underlying premise for speaker recognition is that each person's voice differs in pitch, tone, and volume enough to make it uniquely distinguishable. Several factors contribute to this uniqueness: size and shape of the mouth, throat, nose, and teeth, which are called the articulators and the size, shape, and tension of the vocal cords. The chance that all of these are exactly the same in any two people is low. The manner of vocalizing further distinguishes a person's speech: how the muscles are used in the lips, tongue and jaw. Speech is produced by air passing from the lungs through the throat and vocal cords, then through the articulators. Different positions of the articulators create different sounds. This produces a vocal pattern that is used in the analysis.

A visual representation of the voice can be made to help the analysis. This is called a spectrogram also known as voiceprint, voice gram, spectral waterfall, and sonogram. A spectrogram displays the time, frequency of vibration of the vocal cords (pitch), and amplitude (volume). Pitch is higher for females than for males.

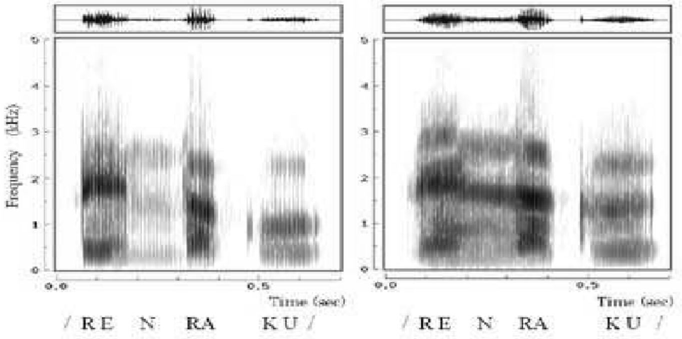

Figure. 2. These voiceprints are a visual representation of two different speakers saying "RENRAKU" [1]

1.4. Methodology

Each speaker recognition system has two phases: Enrollment and verification. During enrollment, the speaker's voice is recorded and typically a number of features are extracted to form a voice print, template, or model. In the verification phase, a speech sample or "utterance" is compared against a previously created voice print

Speaker recognition systems fall into two categories: Text-Dependent and Text-Independent.In a text-dependent system, text is same during enrollment and verification phase .In Text-independent systems the text during enrollment and test is different. In fact, the enrollment may happen without the user's knowledge, as in the case for many forensic applications.

2. General Speaker Recognition System Architecture

There are two major commercialized applications of speaker recognition technologies and methodologies: Speaker Identification and Speaker Verification.

2.1. SIS (Speaker Identification System)

Speaker Identification can be thought of as the task of finding who is talking from a set of known voices of speakers. It is the process of determining who has provided a given utterance based on the information contained in speech waves. Speaker identification is a 1: N match where the voice is compared against N templates.

2.2. SVS (Speaker Verification System)

Speaker Verification on the other hand is the process of accepting or rejecting the speaker claiming to be the actual one. Speaker verification is a 1:1 match where one speaker's voice is matched to one template.

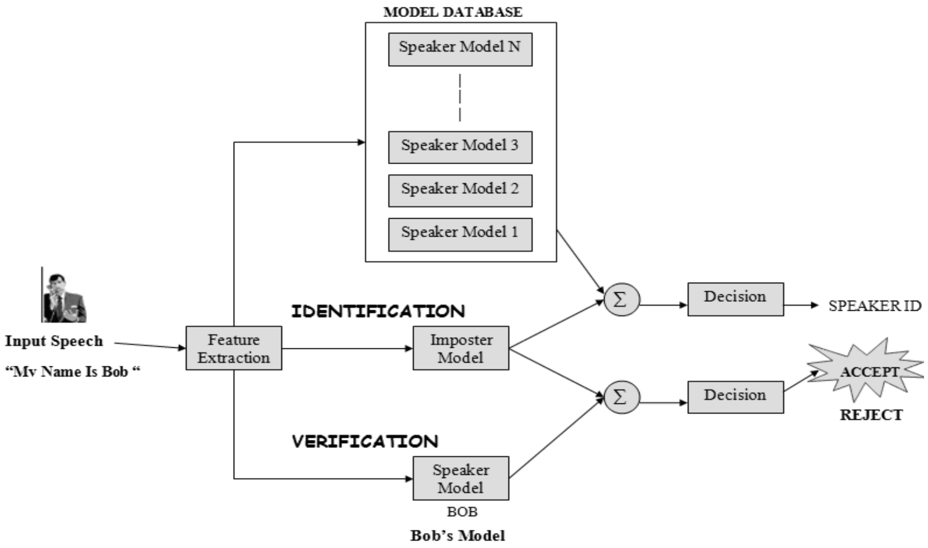

Figure. 3. General SIS and SVS Architecture [2]

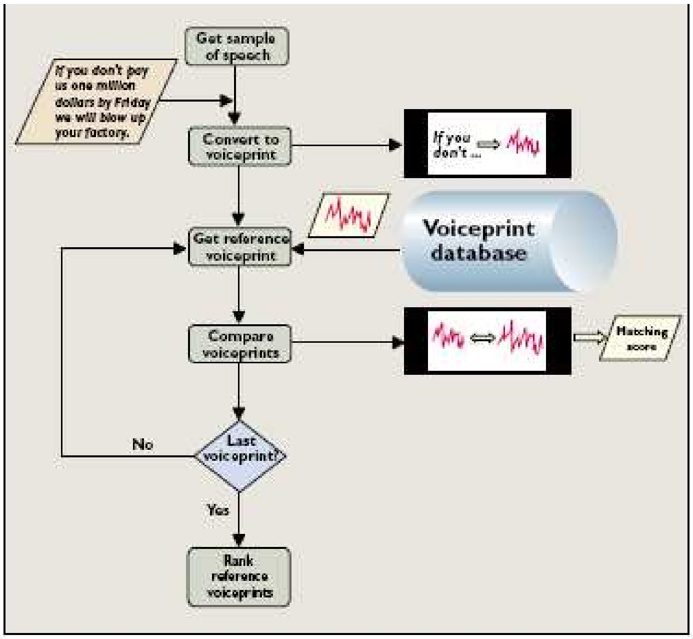

Figure.4. Speaker Identification System[42]

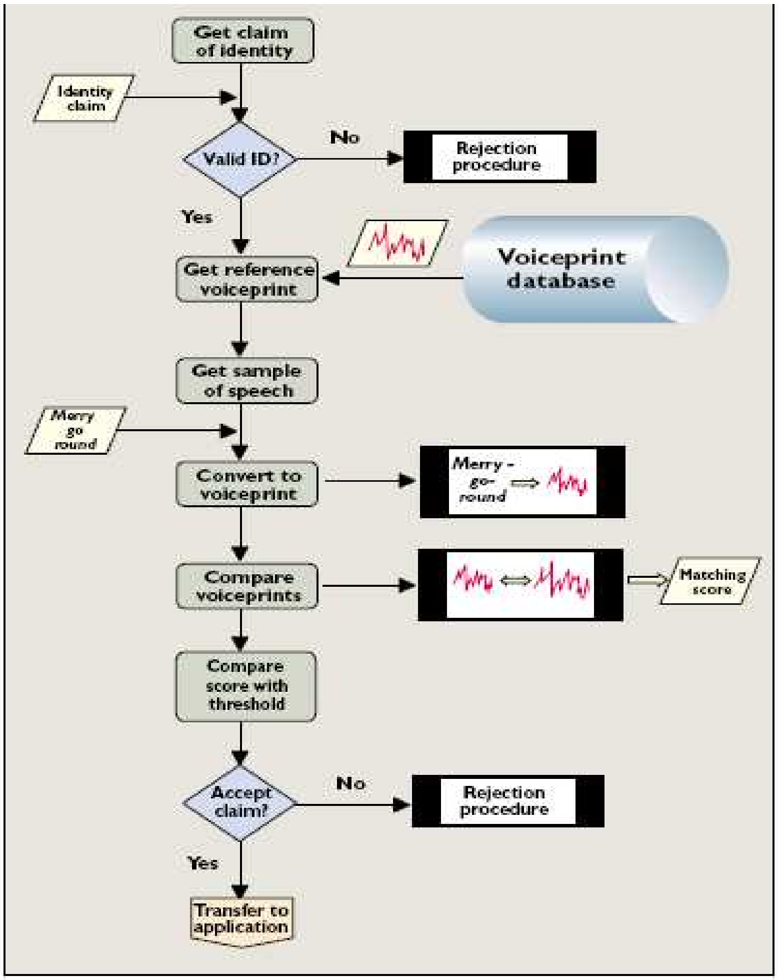

Figure.5. Speaker Verification System[42]

3. Voice Biometric Standards

Standards play an important role in the development and sustainability of technology, and work in the international and national standards arena will facilitate the improvement of biometrics. The major standards work in the area of speaker recognition involves:

- Speaker Verification Application Program Interface(SVAPI)
- Biometric Application Program Interface (BioAPI)
- Media Resource Control Protocol (MRCP)
- Voice Extensible Markup Language (VoiceXML)
- Voice Browser (W3C)

Of these, BioAPI has been cited as the one truly organic standard stemming from the BioAPI Consortium, founded by over 120 companies and organizations with a common interest in promoting the growth of the biometrics market.

4. Commercial Applications of SRS

The applications of speaker recognition technology are quite varied and continually growing. Voice biometric systems are mostly used for telephony-based applications. Voice verification is used for government, healthcare, call centers, electronic commerce, financial services, and customer authentication for service calls, and for house arrest and probation-related authentication.

Table 2. Broad areas where speaker recognition technology has been or is currently used.

S.No.	*Areas/Company using speaker recognition technology*
1.	**Authentication**
	Union Pacific Railroad : Union Pacific moves railcars back and forth across the United States every day. The railcars travel loaded in one direction and empty on the way back. When the loaded railcar arrives, the customer is notified to come and pick up the contents. Once emptied, the customer needs to alert Union Pacific to put the railcar back to work. Union Pacific now has an automated system that utilizes voice authentication to allow a customer to release empty railcars. Customers enroll in the voice authentication system over the phone. When they call back to release an empty railcar, the system authenticates them and allows them to release their railcars. In this case, voice authentication has allowed customers to get off the phone faster, and Union Pacific to guarantee that a customer is not releasing a railcar that doesn't belong to him.
	New York Town Manor: New York Town Manor is a residential community in Pennsylvania designed for senior citizens with technologically advanced features. The residents no longer have to remember passwords. They do carry ID cards that are used in conjunction with voice authentication to allow access to the complex.To enter their apartments, they speak for a few seconds while the system authenticates them. With this approach, voice authentication provides an extra measure of security.
	Bell Canada : Technicians for Bell Canada used to have to carry laptops on the job with them. A technician would dial up using a modem to report the current job as finished and to get the next job. Bell Canada has rolled out a new system that uses voice authentication to verify the identity of the technician through a phone call and give him access to the data. This eliminates the need for a laptop
	Password Journal: Anyone who has ever had a diary has probably worried that someone would read it without permission. One company has solved this problem by adding voice authentication as a privacy measure to their Password Journal product. The journal has its own speaker, raises an alarm if an unauthorized person attempts to access it, and keeps track of how many failed attempts there have been.
	Password Reset: Some companies are allowing users to reset passwords themselves. Users dial an automated system. The system asks questions. When the user answers, the system authenticates his voice and allows him to reset his own password. This saves companies time and money in support costs, and users need not spend time on hold waiting for the next

	available support person

	US Social Security Administration*: The United States Social Security Administration is using voice authentication to allow employers to report W-2 wages online. Used in combination with a pin number, the voice authentication provides system security and user convenience*
2.	**Banking***:* Reducing crime at Automated Teller Machines is an ongoing struggle. Banks have started using biometrics to authenticate users before allowing ATM transactions. Users generally must provide a pin number and a voice sample to be allowed access. Royal Canadian Bank is using voice authentication to allow access to telephone banking.
3.	**Law Enforcement:** In Louisiana, criminals are kept on a short leash with voice biometrics. This inexpensive approach allows law enforcement to check in with offenders at © SANS Institute 2004, Author retains full rights. Key fingerprint = AF19 FA27 2F94 998D FDB5 DE3D F8B5 06E4 A169 4E46 © SANS Institute 2004, As part of the Information Security Reading Room Author retains full rights. Lisa Myers Page 12 7/24/2004 random times of the day. The offender must answer the phone and speak a phrase that is used for authentication. This system guarantees that they are where they are supposed to be! Voice authentication has also been used in criminal cases, such as rape and murder cases, to verify the identity of an individual in a recorded conversation. There is a terrorism application also. Voice authentication is frequently used to validate the identity of terrorists such as Osama Bin Laden on recorded conversations. Hopefully these clues will one day assist in his capture
4.	**AHM (Australia Health Management**): Since 2007, Australia private health insurer AHM has successfully managed one of the largest public-facing deployments of speaker verification. With more than 400,000 yearly calls into its main contact center, ahm has implemented an automated voice verification system to provide quick, accurate authentication of callers enhancing member security and improving the customer experience
5.	**VoiceCash:**Based in Germany, VoiceCash an enabler of mobile payment solutions is targeting consumers interested in cross-border money transfers offering pre-paid payment cards that can be managed online or via SMS communications. The transfers can be authenticated utilizing voice verification technology supplied by VoiceTrust.
6.	**SIMAH:** The Saudi Arabia Credit Bureau is deploying a voice biometric solution provided by Agnitio and IST, a contact center system integrator. The technology is part of IST's iSecure product and will be deployed through SIMAH's new Cisco contact center.

7.	**Vodafone Turkey:** Vodafone Turkey has integrated PerSay VocalPassword with Avaya Voice Portal Platform to enable secure self-service applications such as GSM Personal Unlocking Key reset and access to Vodafone Call Centers

5. Leading Vendors of Speaker Recognition Systems

Table 3. List of vendors

S.No	*Vendor*	*Website*
1.	Persay (NY, USA)	www.persay.com
2.	Agnito (Spain)	www.agnitio.es
3.	TAB Systems Inc. (Slovenia, Europe)	www.tab-systems.com
4.	DAON (Washington DC)	www.daon.com
5.	Smartmatic (USA)	www.smartmatic.com
6.	Speech Technology Center (Russia)	www.speechpro.com
7.	Loquendo (Italy)	www.loquendo.com/en/
8.	SeMarket (Barcelona)	www.semarket.com
9.	RecognitionTechnologiesLtd.(NY)	www.speakeridentification.com

6. Speech database

Table 4. Publicly available speech databases

Database	*Website*
TIMIT	http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93S1
NIST	http://itl.nist.gov/iad/mig/tests/sre
NOIZEUS	http://www.utdallas.edu/~loizou/speech/noizeus/
NTIMIT	http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93S2
YOHO	http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC9z4S16

7. Current Research in Speaker Recognition Systems

Indian Institute of Technology, Guwahati: Study of Source Features for Speech Synthesis and Speaker Recognition & Development of Person Authentication System based on Speaker Verification in Uncontrolled Environment.

Indian Institute of Technology, Kharaghpur: Development of speaker verification software for single to three registered user(s) & Development of Speaker Recognition Software for Telephone Speech

Speech Technology and Research Laboratory, SRI International, CA: Speaker Recognition and Talk Printing

Speech and Speaker Modeling Group, University Of Texas at Dallas: Dialect / Accent Classification & In-Set Speaker Recognition & Speaker Normalization.

The Centre for Speech Technology Research, University of Edinburgh, United Kingdom: Voice transformation

Human Language Technology Group, Lincoln Laboratory, Massachusetts Institute of Technology: Forensic Speaker Recognition Project

CFSL Chandigarh: CFSL is the first Forensic Laboratory in the Country to develop text independent speaker identification system indigenously. A number of important cases related to corruption, threatening calls and identification of individuals through their voice have been solved by CFSL, Chandigarh. CFSL Chandigarh has the technique to match voices irrespective of the language used by the person.

8. Performance Metrics

Biometric systems are not perfect. There are two important types of errors associated with biometric system, namely a false accept rate (FAR) and a false reject rate (FRR). The FAR is the probability of wrongfully accepting an impostor user, while the FRR is the probability of wrongfully rejecting a genuine user. System decisions (i.e. accept/reject) is based on so called thresholds. By changing the threshold value, one can produce various pairs of (FAR, FRR). For reporting performance of biometric system in verification mode, researchers often use a decision error trade-off (DET) curve. The DET curve is a plot of FAR versus FRR and shows the performance of the system under different decision thresholds [43], see Figure 6(a). A modified version of the DET curve is a ROC (Receiver Operating Characteristic) curve, which is widely used in the machine learning community. The difference between DET and ROC curves is in ordinate axis. In the DET curve the ordinate axis is FRR, while in the ROC curve it is 1-FRR (i.e. probability of correct verification). Usually, to indicate the performance of biometric system by a single value in verification mode, an equal error rate (EER) is used. The EER is the point on the DET curve, where FAR=FRR, see Figure 6(a)

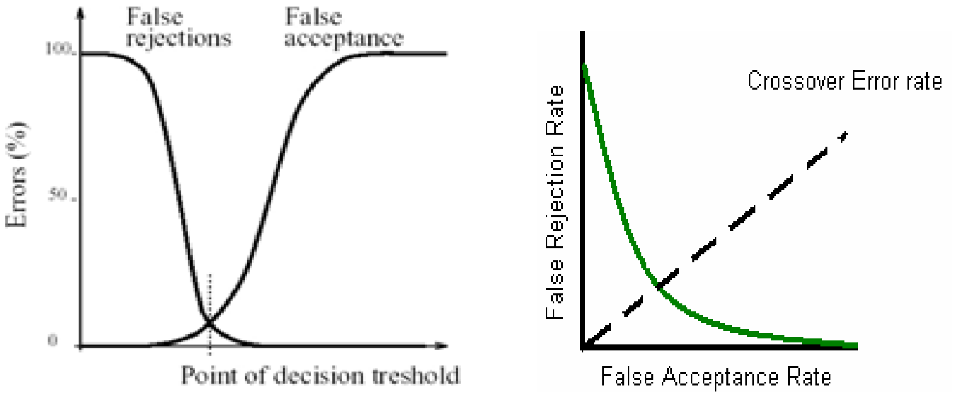

Figure. 6. (a) Point of Decision Threshold ; (b) CER Curve

To evaluate the performance of a biometric system in identification mode, a cumulative match characteristics (CMC) curve can be used. The CMC curve is a plot of rank versus identification probability and shows the probability of a sample being in the top closest matches [43], see Figure 6(b). In identification mode, to indicate performance of the system by a single number, the recognition rate (i.e. identification probability at rank 1) is used. In the next sections, when performance of the method is referred to the recognition rate the system is evaluated in the identification mode, and when it is referred to the EER the system is evaluated in the verification mode.

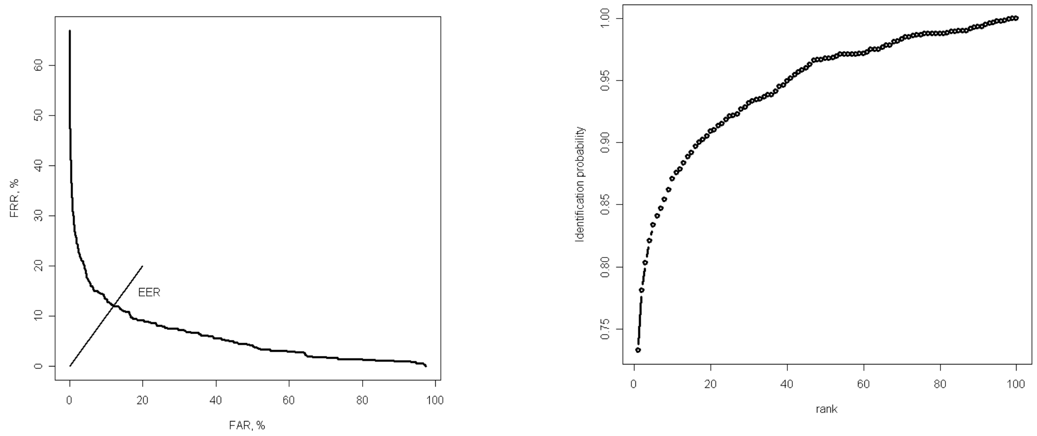

Figure.7. (a) DET Curve (b) CMC Curve

If you plot FAR and FRR against each other, the point at which they intersect is called the crossover error rate (CER). The lower the CER, the better the system is per-forming.

9. Issues Pertaining to SRS

Speaker recognition and verification has been an area of research for more than four decades and thus have many challenges that are needed to overcome.

(a) Hackers might attempt to gain unauthorized access to a voice authenticated system by playing back a pre-recorded voice sample from an authorized user..
(b) A major issue facing all biometric technologies that store data is maintaining the privacy of that data. As soon as a user registers with a voice biometric system, that voiceprint is stored somewhere just like an address or a phone number. What if companies decide to sell voiceprints like addresses?
(c) If the data is encrypted in storage and in transport, there is always the possibility of cracking the encryption and stealing the data.
(d) Designing long-range features (which by definition occur less frequently than very short-range features) that provide robust additional information even for short (e.g., 30 seconds) training and test spurts of speech.
(e) Develop methods for feature selection and model combination at the feature level, that can cope with large numbers of interrelated features, odd feature space distributions, inherent missing features (such as pitch when a person is not voicing), and heterogeneous feature types.

10. SRS Modules

10.1. Preprocessing

The captured voice may contain unwanted background noise, unvoiced sound, and there can be a device mismatch, environmental mismatch between training and testing voice data which subsequently leads to degradation in the performance of Speaker Recognition System. The process of removal of this unwanted noise, dividing sounds into voiced and unvoiced sounds and channel compensation etc for the enhancement of speech/voice is called pre-processing.

10.1.1. Speech Enhancement (Denoising): Numerous schemes have been proposed and implemented that perform speech enhancement under various constraints/assumptions and deal with different issues and applications.

Table 5. Various approaches for speech enhancement

S.No.	*Approach*	**Characteristic**
1.	Chowdhary et al. 2008][3]	Improvement over the conventional power spectral subtraction method.

2.	Jun et al. [2009][4]	Based on fast noise estimation.
3.	Hansen et al. [2006][5]	A Generalized MMSE estimator (GMMSE) is formulated after study of different methods of MMSE family.
4.	Hasan et al. [2010][6]	Considers the constructive and destructive interference of noise in the speech signal.
5.	Lev-Ari et al. [2003][7]	This algorithm is an extension to signal subspace approach for speech enhancement to colored-noise processes.
6.	Li, C.W. et al. [2007][8]	This approach is based on Signal Subspace Approach combined with RL noise estimation for non-stationary noise.
7.	Tinston M et al. [2009][9]	A subspace speech enhancement approach for estimating a signal which has been degraded by additive uncorrelated noise.
8.	Jia et al. [2009][10]	This method tracks real time noise Eigen value matrix in the subspace domain by applying statistical information in the whole time, and corrects speech Eigen value matrix making use of the principle of winner filtering.

10.1.2. Channel Compensation: Channel effects, are major causes of errors in speaker recognition and verification systems. The main measures to improving channel robustness of speaker recognition system are channel compensation and channel robust features.

Table 6. Various approaches for channel compensation

S.No	*Approach*	*Characteristics*
1.	Wu et al. [2006][8]	Utilizes channel-dependent UBMs as a priori knowledge of channels for speaker model synthesis.
2.	Han et al. [2010] [9]	Applies MAP channel compensation, pitch dependent feature and Speaker model.
3.	Calvo et al. [2007][10]	This paper examines the application of Shifted Delta Cepstral (SDC) features in biometric speaker verification and evaluates its robustness to channel/handset mismatch due by telephone handset variability.
4.	Zhang et al. [2008][30]	GMM super vectors generated by stacking means of speaker models can be seen as combination of two parts: universal background model (UBM) super vector and maximum a posteriori (MAP) adaptation part
5.	Neville et al. [2005][31]	Blind equalization techniques with QPSK modulation

5.	Deshpande et al. [2010][20]	AWP for Speaker Identification and multiresolution capabilities of wavelet packet transform are used to derive the new features.
6.	Barbu et al. [2007][21]	A text-independent voice recognition system representation of the vocal feature vectors as truncated acoustic matrices with DDMFCC coefficients.
7.	W¨olfel et al. [2009][22]	Replaces the widely used Mel-frequency cepstral coefficients by warped minimum variance distor-

10.2. Feature Extraction

The purpose of this module is to convert the speech waveform to some type of parametric representation (at a considerably lower information rate).The heart of any speaker recognition system is to extract speaker dependent features from the speech.They are basically categorized into two types: low level and high level features.

10.2.1. Low Level Features: Low level features are short range features.

Table 7. Various approaches for extraction of low level features

S.No	*Approach*	**Characteristic**
1.	Prahallad et al. [2007][16]	Auto-associative neural network (AANN) and formant features
2.	Chakroborty et al. [2009][17]	Gaussian filter based MFCC and IMFCC scaled filter bank is proposed in this paper
3.	Revathi et al. [2009][18]	Perceptual Features & Iterative clustering approach for both speech and speaker recognition
4.	Huang et al. [2008][19]	Fusion of pitch and MFCC GMM Supervectors Systems on score level

		tion less response cepstral coefficients for speaker Identification.
8.	Guo et al. [2006][23]	After MFCC extraction, both Cepstral Mean Subtraction (CMS)and RASTA filtering are used to remove linear channel convolutional effect on the cepstral features.

10.2.2. High Level Features: Higher level features are long range features of voice that have attracted attention in automatic speaker recognition in recent years.

Table 8. Various approaches for extraction of high level features

S.NO	*Approach*	**Characteristic**
1.	Campbell et al. [2007][24]	SVM and new kernel based upon linearizing a log likelihood ratio scoring system
2.	Baker et al. [2005][25]	Presegmentation of utterances at word level using ASR system, HMM & GMM used
3.	Mary et al. [2008][26]	Syllable-like unit is chosen as the basic unit for representing the prosodic characteristics.
5.	Campbell et al. [2004][28]	Proposes the use of support vector machines and term frequency analysis of phone sequences to model a given speaker.

10.3. Modeling

Speaker Model Generation: The feature vectors of speech are used to create a speaker's model/template. The recognition decision depends upon the computed distance between the reference template and the template devised from the input utterance.

Table 9. Various approaches for speaker model generation

S.No	*Approach*	*Characteristic*
1.	Aronowitz et al. [2005][29]	Used GMM simulation & Compression algorithm
2.	Zamalloayz et al. [2008][30]	GA(Genetic Algorithm) & Comparison with LDA and PCA
3.	Aronowitz et al. [2007][31]	Based on approximating GMM likelihood scoring using ACE ,GMM compression algorithm
4.	Apsingekar et al. [2009][32]	GMM-based speaker models are clustered using a simple k-means algorithm

5.	Chakroborty et al. [2009][33]	Fusion of two GMM for each speaker one for MFCC and other for IMFCC feature sets.

10.4. Matching /Decision Logic

Score Normalization: Score normalization has become a crucial step in the biometric fusion process. It makes uniform input data that comes from different sources or processes, hence reduces the biased information created by the differences between the different pre-processors.

Table 10. Various approaches for score normalization

S.No	*Approach*	*Characteristic*
1.	Puente et al. [2010][34]	New normalization algorithm DLin is proposed.
2.	Guo et al. [2008][35]	An unsupervised score normalization is proposed.
3.	Castro et al. [2007][36]	Score normalization technique based on test-normalization method (Tnorm) is presented
4.	Zajic et al. [2007] [37]	Unconstraint cohort extrapolated normalization, is introduced.
5.	Sturim et al. [2005][38]	A new method of speaker Adaptive-Tnorm that offers advantages over the standard Tnorm by adjusting the speaker set to the target model is presented.
6.	Mariéthoz et al. [2005][39]	New framework is proposed in which Z- and Tnormalization techniques can be easily interpreted as different ways to estimate score distributions.
7.	Barras et al. [2003][40]	This paper presents some experiments with feature and score normalization for text-independent speaker verification of cellular data

11. Conclusions

In this paper, we have presented an extensive survey of automatic speaker recognition systems. We have categorized the modules in speaker recognition and discussed different approaches for each module. In addition to this, we have presented a study of the various typical applications of Speaker Recognition Systems, list of vendors worldwide and the current research being carried out in the field of speaker recognition. We have also discussed issues and challenges pertaining to the Speaker Recognition Systems.

References

[1.] SANS Information Security Reading Room, http://www.sans.org

[2.] Biometrics.gov, http://www.biometrics.gov/Documents/SpeakerRec.pdf

[3.] Chowdhury, M.F.A., Alam, M. J., Alam, M.F.A, and O'Shaughnessy, D.: Perceptually weighted multi-band spectral subtraction speech enhancement technique. In: ICECE 2008. International Conference on Electrical and Computer Engineering, Page(s): 395 – 399 (2008)

[4.] Jun, L., He, Z.: Spectral Subtraction Speech Enhancement Technology Based on Fast Noise Estimation. In: ICIECS (2009)

[5.] Hansen, J.H.L., Radhakrishnan, V., Arehart, K.H.: Speech Enhancement Based on Generalized Minimum Mean Square Error Estimators and Masking Properties of the Auditory System. In: IEEE Transactions on Audio, Speech, and Language Processing, Volume: 14, Issue: 6, Page(s): 2049 – 2063 (2006)

[6.] Hasan, T., Hasan, M.K.: MMSE estimator for speech enhancement considering the constructive and destructive interference of noise. In: Signal Processing, IET Volume: 4, Issue: 1, Page(s): 1 – 11 (2010)

[7.] Lev-Ari, H., Ephraim, Y.: Extension of the signal subspace speech enhancement approach to colored noise. In: Signal Processing Letters, IEEE Vol.: 10, Issue: 4, pp.: 104 – 106 (2003)

[8.] Li, C.W., Lei, S.F.: Signal subspace approach for speech enhancement in nonstationary noises. In: ISCIT '07. International Symposium on Communications and Information Technologies, Page(s): 1580 – 1585(2007)

[9.] Tinston M, Ephraim Y.: Speech enhancement using the multistage Wiener filter. In: CISS 2009 43rd Annual Conference on Information Sciences and Systems, Page(s): 55 – 60 (2009)

[10.] Jia, H., Zhang, X., Ji, C.: A Modified Speech Enhancement Algorithm Based on the Sub space. In: Knowledge Acquisition and Modeling, KAM '09, Page(s): 344 – 347 (2009)

[11.] Wu, W., Zheng, T.F., Xu, M.: Cohort-Based Speaker Model Synthesis for Channel Robust Speaker Recognition. In: ICASSP (2006)

[12.] Han, J., Gao, R.: Text-independent Speaker Identification Based on MAP Channel Compensation and Pitch-dependent Features. In: IJECSE (2010)

[13.] Calvo, J.R., Fernandez, R., Hernandez, G.: Channel / Handset Mismatch Evaluation in a Biometric Speaker Verification Using Shifted Delta Cepstral Features. In: CIARP (2007)

[14.] He, L., Zhang, W., Shan, Y., Liu, J.: Channel Compensation Technology in Differential GSV–SVM Speaker Verification System. In: APCCAS (2008)

[15.] Neville, K., Jusak, J., Hussain, Z.M. and Lech, M.: Performance of a Text-Independent Remote Speaker Recognition Algorithm over Communication Channels with Blind Equali sation. In: Proceedings of TENCON (2005)

[16.] Prahallad, K., Varanasi, S., Veluru, R., Bharat Krishna, M., Roy, D.S.: Significance of Formants from Difference Spectrum for Speaker Identification. In: INTERSPEECH-2006

[17.] Chakroborty, S., and Saha, G.: Improved Text-Independent Speaker Identification using Fused MFCC & IMFCC Feature Sets based on Gaussian Filter. In: IJSP (2009)

[18.] Revathi, A., Ganapathy, R. and Venkataramani, Y.: Text Independent Speaker Recognition and Speaker Independent Speech Recognition Using Iterative Clustering Approach. In: (IJCSIT), Vol 1, No 2 (2009)

[19.] Huang, W., Chao, J., Zhang, Y.: Combination of Pitch and MFCC GMM Supervectors for Speaker. In: ICALIP (2008)

[20.] Deshpande, M.S., and Holambe, R.S.: Speaker Identification Using Admissible Wavelet Packet Based Decomposition. In: International Journal of Signal Processing 6:1 (2010)

[21.] Barbu, T.: A Supervised Text-Independent Speaker Recognition Approach. In: World Academy of Science, Engineering and Technology (2007)

[22.] W¨olfel, M., Yang, Q., Jin, Q., Schultz, T.: Speaker identification using Warped MVDR Cepstral Features. In: International Symposium on Computer Architecture (2009)

[23.] Guo, W., Wang, R., and Dai, L.: Feature Extraction and Test Algorithm for Speaker Verification. In: International Symposium on Chinese Spoken Language Processing (2006)

[24.] Campbell, W.M., Campbell, J.P., Gleason, T.P., Reynolds, D.A., Shen, W.: Speaker Verification using Support Vector Machines and High-Level Feature.: In: IEEE Transactions on Audio, Speech, And Language Processing, Vol. 15, No. 7 (2007)

[25.] Baker, B., Vogt, R., and Sridharan, S.: Gaussian Mixture Modeling of Broad Phonetic and Syllabic Events for Text-Independent Speaker Verification. In: Euro speech, (2005)

[26.] Mary, L., Yegnanarayana, B.: Extraction and representation of prosodic features for language and speaker recognition. In: ELSEVIER Speech Communication 50 782–796 (2008)

[27.] Dehak, N., Dumouchel, P. and Kenny, P.: Modeling Prosodic Features with Joint Factor Analysis for Speaker Verification. In: IEEE Transactions on Audio, Speech and Language Processing 15(7), 2095-2103 (2007)

[28.] Campbell, W. M., Campbell, J. P., Reynolds, D. A., Jones, D. A., Leek, T. R., Phonetic Speaker Recognition with Support Vector Machines, In Proc. NIPS (2004)

[29.] Aronowitz, H., Burshtein, D.: Efficient Speaker Identification and Retrieval. In: Proc. Interspeech 2005, pp. 2433–2436 (2005)

[30.] Zamalloayz, M., Rodriguez-Fuentesy, L.J., Penagarikanoy, M., Bordely, G., Uribez, J.P.: Feature Dimensionality Reduction Through Genetic Algorithms For Faster Speaker Recognition. In: EUSIPCO 2008 16th European Signal Processing Conference (2008)

[31.] Aronowitz, H., Burshtein, D.: Efficient Speaker Recognition Using Approximated Cross Entropy (ACE). In: IEEE Transactions on Audio, Speech, and Language Processing, Vol. 15, No. 7 (2007)

[32.] Apsingekar, V.R. and De Leon, P.L.: Speaker Model Clustering for Efficient Speaker Identification in Large Population Applications. In: IEEE Transactions on Audio, Speech, and Language Processing, Vol. 17, No. 4 (2009)

[33.] Chakroborty, S., and Saha, G.: Improved Text-Independent Speaker Identification using Fused MFCC & IMFCC Feature Sets based on Gaussian Filter. In: International Journal of Signal Processing 5;1 (2009)

[34.] Puente, L., Poza, M., Ruiz, B. and García-Crespo, A.: Score Normalization for Multimodal Recognition Systems. In: JIAS (2010)

[35.] Guo, W., Dai, L., Wang, R.: Double Gauss Based Unsupervised Score Normalization in Speaker Verification.: In: ISCSLP 2008, pp. 165-168 (2008)

[36.] Castro, D.R., Fierrez-Aguilar, J., Gonzalez-Rodriguez, J., Ortega-Garcia, J.: Speaker Verification using Speaker and Test Dependent Fast Score Normalization. In: Pattern Recognition Letters, vol. 28, pp. 90-98 (2007)

[37.] Zajíc, Z., Vaněk, J., Machlica, L., Padrta, A.: A Cohort Method for Score Normalization in Speaker Verification System, Acceleration of On-line Cohort Methods. In: SPECOM (2007)

[38.] Sturim, D.E. and Reynolds, D.A.: Speaker Adaptive Cohort Selection for Tnorm in Text-independent Speaker Verification. In: Proceedings of ICASSP, 2005

[39.] Mariéthoz, J. and Bengio, S.: A Unified Framework for Score Normalization Techniques Applied to Text-Independent Speaker Verification. In: IEEE Signal Processing Letters, Vol. 12, No. 7 (2005)

[40.] Barras, C. and Gauvain, J.: Feature And Score Normalization For Speaker Verification Of Cellular Data. In: Proceedings of ICASSP 2003, pp. 49-52 (2003)

[41.] Gupta, C.S.: Significance of Source Feature for Speaker Recognition. In: A M.S Thesis IIIT Madras (2003)

[42.] Markowitz, J.A.: VoiceBiometrics. In: Vol. 43, No. 9 Communications Of The ACM (2000)

[43.] Martin, A., Doddington, G., Kamm, T., Ordowski, M. and Przybocki, M.: The DET curve in assessment of detection task performance. In: Eurospeech, pages 1895– 898 (1997)

Authors

Shri Zia Saquib is the Executive Director of CDAC-Mumbai and Electronic City Bangalore since 2006. His research interests are in areas of coding theory, applied cryptography, network security and Biometrics.

Ms. Nirmala Salam joined CDAC Mumbai in 2001 and currently working as a Sr. Staff Scientist. Her research interests include Remote speaker recognition, Multilingual speech recognition; Image processing, Digital Signal Processing and Biometrics.

Ms. Rekha Nair joined CDAC Mumbai in 2000 and currently-working as a Sr. Staff Scientist. Her research interests include Remote speaker recognition, Multilingual speech recognition, Image processing, Digital Signal Processing and Biometrics.

Mr. Nipun Pandey joined CDAC Mumbai in 2006 and currently working as a Staff Scientist. His research interests include Re mote speaker recognition, Multilingual speech recognition, Image processing, Digital Signal Processing and Biometrics.