Biometrics

A Survey on Biometrics
- Dr. Rajarshi Pal, Assistant Professor .................................................. 01

Voice based Authentication
- Dr. V. N. Sastry, Professor ................................................................. 54

Biometric Template Protection for Banking
- Dr. M. V. N. K. Prasad, Associate Professor ........................................ 102

Biometrics and Its Impact in India
- Dr. S. Ananth, Adjunct Faculty ............................................................. 119
We are happy to present the next edition of our Staff Paper Series. The series has been helping the Institute in presenting the academic work carried out by our faculty in various areas related to banking technology. The theme of the current edition is Biometrics.

The world in general, and banking sector in particular, has been working on usage of biometric-based identification and authentication techniques. In the process of adoption of such techniques, they have been encountering certain difficulties. However, there has been considerable research in progress to address most of the difficulties. The current IDRBT Staff Paper Series is an attempt to present the issues and the approaches to address them.

The first paper is a Survey on Biometrics by Dr. Rajarshi Pal. Dr Sastry presents various methods used in Voice Based Authentication. Dr. Prasad dwells on Biometric Template Protection, a crucial aspect for Banking. Dr Anant discusses the Impact of Biometrics on Banking in India.

We trust that the papers would provide necessary inputs and impetus to regulators, academic institutions, banks and industry in defining, designing, developing and deploying suitable biometrics based systems in India.

Feedback is welcome at publisher@idrbt.ac.in.

Date: October 2017  
Place: Hyderabad
A Survey on Biometrics

- Dr. Rajarshi Pal,
  Assistant Professor, IDRBT
1. Introduction

The word ‘biometric’ refers to a few unique characteristics of a person’s physiology or behavior which do not usually change with time. Examples of such physiological characteristics include fingerprint, iris, palmprint, face, etc. Examples of behavioral biometric includes hand-written signature, voice, gait and typing style on a keyboard.

Biometric can play a major role to verify or to identify an individual. Verification is the process to confirm the identity of a claimant. In this process, one or more biometric features of the claimant are validated against the known biometric profile of the individual. Therefore, the verification process requires a one-to-one match. In the case of identifying an individual, the biometric identity of the unknown individual is matched with the biometric of several others in an existing database. Hence, identification involves a one-to-many comparison. Uniqueness of biometric characteristics stops an imposter against making a false verification and identification attempts.

The following example can be considered to explain the concept of the verification process. To avail an Aadhaar-enabled service, the user types her Aadhaar number to specify her identity. Then, the system compares the biometric of the user with that of the enrolled user. Here, one-to-one matching takes place. In this case, an acceptable similarity between the captured biometric of the claimant and the biometric of the enrolled user establishes the genuineness of the claim. Otherwise, the claimant is considered as an impostor and her claim is rejected. Hence, the verification process attempts to establish the following: “you are who you say you are”.

On the contrary, in an identification process, an individual claims that she is one of the registered members as per the record. As a part of this process, the individual’s biometric is matched with that
of every member in the record to identify her as one with whom the highest similarity score has been found. But, if the highest similarity score is less than a threshold, then it can be concluded that there is no similarity between the input and the registered members. This establishes the claimant as an impostor. Therefore, an identification process requires 1-to-N comparisons. N is the number of registered members in the above discussion. The identification process establishes an individual as “someone who is already enrolled”.

This article revisits various biometric traits, the steps involved in recognition, multimodal biometric systems and the security issues in biometric recognition systems. The organization of the rest of the article is as follows the steps in a biometric recognition system have been depicted in Section 2, various categories of biometric traits are presented in Section 3 and Section 4 discusses about multimodal systems. Fusion in multimodal biometric system can take place at various levels as discussed in Section 5. Section 6 discusses the security concerns supposed to be addressed by a biometric recognition system. Amidst all these theoretical discussions and practical challenges, banks across the globe have embraced biometrics as a factor of authentication. Section 7 provides a glimpse of such adoptions of biometric by banks and the article concludes with Section 8.

2. Steps in Biometric System

Registration (or enrollment) and recognition are two phases of a biometric based recognition system. The block diagram is presented in Figure 1. The registration phase includes pre-processing, region-of-interest detection and feature extraction steps. The extracted features are then stored in the database. The recognition phase includes pre-processing, region-of-interest detection, feature extraction, matching and decision making steps. Matching module compares the extracted features with the stored features in the database for either identification or verification task.
2.1. Pre-processing

The pre-processing step is primarily used to improve the acquired image (or signal) in order to obtain an accurate extraction of region of interest and the biometric features. Generally, rescaling, for example mean subtraction and feature standardisation are commonly used as pre-processing for several of the biometric recognition systems. The approach of biometric modality-specific pre-processing is abundant in the literature. Few examples of these are presented below.

An experimental study of several illumination pre-processing methods for face recognition is reported in [Han et al, 2013]. These
methods have been divided into three main categories as – gray-level transformation, gradient or edge extraction and reflectance field estimation. [Jahanbin et al, 2011] has used following four steps as pre-processing for a face recognition system: gamma correction, Difference of Gaussian (DoG) filtering, masking and equalization of variation. The axis-symmetric nature of a face is considered to generate an approximately symmetrical face image in [Xu et al, 2016] for face recognition. This increases the accuracy of face recognition methods.

Filtering, equally-spacing, location, size and time normalization are key pre-processing steps for an online signature verification in [López-García et al, 2014]. To avoid the acquisition device dependency, the acquired data is also normalized in a fixed range in an online signature verification system [Tolosana et al, 2015].

2.2. Finding Region of Interest

Locating the Region of Interest (ROI) for a biometric trait is essential precursor for feature extraction step. This step identifies the main or interesting portion of the image (or signal) from where the biometric traits are extracted. For example, several studies on palmprint recognition consider the size of the palm to determine the ROI. Similarly, iris localization is integral to iris recognition [Lili and Mei, 2005].

The methods to extract ROI certainly depends on the modality of the biometric system. The techniques to identify the region of interest can be grouped into three major divisions, namely:

(i) Bottom-Up Feature Based Approach: This approach does not assume any apriori information about the region of interest. Hence, these approaches are purely driven by detection of key points in a purely bottom-up approach. For example, face localization can be carried out using the Scale Invariant Feature Transform (SIFT). A scale invariant region detector and a descriptor based on the
gradient distribution in the detected regions play major role in this approach.

(ii) Top-Down Knowledge Based Approach: This approach is influenced by additional relevant information about the physiological characteristics. For example, [Jones and Viola, 2006] has considered individual’s motion and appearance in determining the region of interest.

(iii) Appearance Based Approach: This approach considers the inherent physiological appearance of a biometric trait. As an example, it can be studied how region of interest of a palm is extracted in [Saliha et al, 2014]. A key point localization is developed to spot the crucial points of a palm. It helps in the proper alignment of the hand image. This approach is further based on a projection of the X-axis and the projection of the upper and lower edge. Hence, it extracts the horizontal limits of the hand contour. In [Belahcene et al, 2014], a 3D face recognition is proposed by finding regions of interest in a face, which include mouth, nose, pair of eyes, etc.

2.3. Feature Extraction

In the feature extraction step, the properties or inherent patterns in a biometric trait is extracted from the input (and possibly pre-processed image/signal). Thus, the derived properties or patterns are a better representation of the unique elements of an individual’s biometric trait. Definitely, the type of biometric decides the feature extraction step. The following paragraph highlights few examples in support of this dependency.

Two different feature extraction approaches are present for hand-written signatures as they aim to capture the static or the dynamic features. Geometrical features of the signature are considered for the static approach. Dynamic features of hand-written signature includes the speed and the acceleration of the pen movement, pen-up and pen-down times, etc. Ear curves are extracted for an ear
recognition system in [Ghoualmi et al, 2015]. The face recognition system in [Kumar and Kanhangad, 2015] has used the techniques like wavelet transform, spatial differentiation and twin pose testing scheme for feature extraction from faces. According to [Ukpai et al, 2015], principal texture pattern and dual tree complex wavelet transform produce iris-specific features from an iris image. The next section on various biometric traits will lead to a better understanding of this through narration of different biometric traits.

2.4. Matching and Decision

In this step, the extracted features are compared with the enrolled features to obtain a matching score. The subsequent decision making step either accepts or rejects an individual using this matching score.

3. Types of Biometrics

3.1 Fingerprint

Uniqueness and consistency in performance have established the fingerprint as the most widely used biometric trait. Usage of fingerprint can be traced back to previous centuries. Ease in acquisition, availability of 10 different fingers and its acceptance for law enforcement and immigration purposes have established it as a very popular form of biometric.

A fingerprint is obtained from the friction ridges of the finger. High and peaking part of the skin causes the dark lines in the fingerprint as it is shown in Figure 2. White spaces in between dark lines are due to the shallow parts of the skin, which are also called the valleys. The ridges and furrows (as appearing in Figure 2) enable us to firmly hold objects. Their presence causes a friction, which is needed to grab any object. But uniqueness of fingerprint is not due to these ridges and furrows. Uniqueness is achieved due to minutiae points. The minutiae points are defined as the points where the ridges end, split
and join, or appear as a simple dot. The patterns of placement of these minutiae points lead to uniqueness. The minutiae consists of bifurcations, ridge dots, ridge endings and enclosures. The minutiae points are further broken down into sub minutiae such as pores, crossovers and deltas to ensure further uniqueness. Tiny depressions within the ridge are called the pores in a fingerprint. An ‘X’ pattern in the ridge is called crossover. A triangle-shaped pattern in the ridge is called delta.

The widespread adoption of fingerprint for biometric recognition is due to several factors like its reasonably good accuracy, ease of use and the small amount of memory space to store biometric template. With the emergence of mobile based applications, the above strengths of fingerprint biometric have led to the use of it for mobile authentication. But the performance of fingerprint recognition drops due to scaly or dirty skin of finger and changes with age.

The combination of minutiae points of two different fingers of an individual enables privacy protection in [Li and Kot, 2013].

![Fingerprint](image)

**Figure 2: Fingerprint**
3.2 Iris

Iris is considered to be another reliable biometric trait. The iris is the muscle in the eye which controls the pupil size in order to regulate the amount of light rays entering the eye. The iris can be identified as an annular region in between the sclera (white portion of the eye) and the pupil (Figure 3). The pattern of iris of an individual is also unique [Jain et al, 2004]. A set of twins also possess distinguishing iris patterns. The speed and accuracy of iris recognition have also caused widespread adoption of iris biometric.

Registration of iris takes relatively longer time as it needs several iris images. A test template is generated upon scanning of an individual’s iris. Subsequently, the produced template is matched with the existing templates which were produced at the time of registration. Zero crossing representation of the one-dimensional wavelet transform has been proposed in [Radu et al, 2012] to encode the texture in iris. In [Sun et al, 2005], an integration of Local Feature Based Classifier (LFC) and an iris blob matcher increases the accuracy of iris recognition. The noisy images are detected by the iris blob matcher. Therefore, it helps in the situations where the LFC does not guarantee the performance.

![Figure 3: Iris as part of human eye](image)

In [Dong et al, 2011], a set of training images is used to learn class-specific weight map for iris matching technique. Experiments have
revealed the effectiveness of the technique. Moreover, a pre-processing method has been suggested to enhance the performance of iris biometric on mobile phones which usually are constrained by computing power. The system has demonstrated its capability to decide whether the person is dead or alive.

3.3 Face

The structure of human face is characterized by peaks and valleys at different altitudes and features which are present at different specific latitudes and longitudes as demonstrated in Figure 4. This distinguishes one individual from another.

![Figure 4: Extraction of Key Features from Human Face](image)

Earlier attempts of face recognition used simple geometric models. But sophisticated mathematical representation of features has led to better models of face recognition [Jain et al, 2004]. Combination of certain features with AdaBoost leads to a face and eye detection method in [Parris et al, 2011]. Results are encouraging enough to adopt face based authentication in mobile phones. The face recognition method by [Lai et al, 2014] is assisted with motion sensors. Apple’s iDevice has a face recognition system to lock and unlock it [Gao et al, 2014]. According to [Srinivasan and Balamurugan, 2014], Pictet and Banquiers (one of the leading banks in Switzerland) has deployed an efficient 3D face recognition system for providing access to its staff within the bank’s environment. A
graph based model for face recognition has been proposed in [Cao et al, 2012]. Based on 3D features, a face recognition system has been proposed to improve the performance of the system. For detecting facial features, the active appearance model has been used in [Drosou et al, 2012]. A support vector machine based face recognition system has been proposed in [Hayat et al, 2012]. In this technique, an elastic graph matching has been utilized to locate the feature points of the facial image.

A face recognition system fails when the face is partly covered. In this case of occlusion, the important characteristics of the face cannot be captured.

### 3.4 Ear

The appearance and shape of the human ear is also found to be unique. It changes little during an individual’s lifetime. Three main steps of an ear biometric system are – (a) imaging of the ear, (b) image segmentation, and (c) recognition. A camera is obviously used for image acquisition. Segmentation is carried out to isolate the ear from the background in the image. A convex curved boundary is identified to locate the ear as in [Maity and Abdel-Mottaleb, 2015]. But experiments have revealed a high false positive due to occlusion. Recognition is performed by comparing the ear biometric traits with stored templates in the database. Local surface patch representation at 3D space leads to a 3D ear recognition system in [Abate et al, 2006].

The ear biometric is captured by Nippon Electric Company (NEC) as the vibration of sound as influenced by the shape of an individual’s ear. It is claimed to be unique for every person. According to this system, an earphone with a built-in microphone captures the sounds as they vibrate within the ear.
3.5 Hand Geometry

Measurements of a human hand are used to recognize an individual in the case of hand geometry as biometric trait [Jain et al, 2004]. Shape and size of the palm along with shape, width and length of each finger are considered as important measurements in this context. Edge detectors like Sobel or Canny operators can be used to detect palm lines. Ease of use of this biometric trait leads to wide acceptance of this biometric even in mobile devices [Chen et al, 2007]. But lack of uniqueness of this trait is a major drawback. Hence, its usage is confined only to one-to-one matching. For example, this can be used for access control, where the concern is about an individual’s attempt to gain access through someone else’s access card or personal identification number. The individual’s physical presence is ensured through the presentation of her hand to the hand reader. Though, it can be combined with other biometric traits [Javidnia et al, 2016].

3.6 Palm Vein

Vein pattern in the palm is considered to be another unique trait to recognize an individual. The presence of blood vessels underneath the skin causes this pattern. It is less susceptible to external distortion. Forgery is also difficult for this biometric trait. Moreover, the vein pattern is said to remain static during the lifetime of an individual. The acquisition device throws an infra-red beam on the palm as it is put on the sensor. The veins in the palm are identified as black lines in the captured image. They are matched with an existing vein pattern to recognize an individual [Tome and Marcel, 2015]. Haemoglobin, which is a key ingredient of blood, absorbs the near infra-red (NIR) light. Hence, [Sugandhi et al, 2014] has suggested usage of near infrared (NIR) light to acquire the vein image of fingers. As a result, the vein pattern in fingers appears as shadows.

Inspired by fingerprint recognition, palm vein recognition system in [Vaid and Mishra, 2015] also extracts vein minutiae.
3.7 Palmprint

A comparatively new biometric trait has been discovered in terms of the palmprint (Figure 5). Reliable and unique characteristics of palmprint justify its high usability. Similar to fingerprint, the palmprint also has unique features, namely, principal lines, minutiae features, delta points, wrinkles, and ridges. Additionally, a wider surface area of the palm (as against the surface area being captured for the fingerprint) leads to more number of unique traits. Hence, palmprint biometric is believed to mature quickly as a reliable recognition system.

![Figure 5: Palmprint](image)

But deformation of images due to challenges of acquisition pulls down the accuracy of a palmprint recognition system. A contact based acquisition device which can pose constraints on acquisition environment is used to solve this problem. Research is still required to tackle the issues arising out of positioning, rotating, and stretching the palm. Moreover, the bigger size of the acquisition device does not allow its usage over mobile phones. Contact based acquisition may also be considered unhygienic. Hence, contactless palmprint acquisition has also been introduced in [Wu and Zhao, 2015]. The users need not touch the acquisition device.
3.8 Retina

Each individual possesses unique retina vasculature. Replication of it is not easy. The acquisition environment demands an individual to focus her eye on a scanner. Therefore, the system may cause some medical complications like hypertension. This is one reason why this biometric system has not received a full acceptance by the public.

3.9 Radio Biometric

Certain physical characteristics (such as height and mass), the condition of the skin, the volume of total body water, and nature of other biological tissues influence the wireless propagation around the human body. Radio biometrics is defined as the identity information as specified by the human-affected wireless signal under alterations and attenuations. The variability of these physical characteristics and biological features among different individuals ensures that two humans are less likely to demonstrate the identical radio biometric. As the chance of two persons having exactly same physical and biological characteristics is very little, the multi-path profiles of the electromagnetic waves after interference from human body vary for each individual. Consequently, human radio biometric, which records how the wireless signal interacts with a human body, are altered according to individuals’ biological and physical characteristics and can be viewed as unique among different individuals.

Radio biometric captures the response of radio waves from the entire body including the face of an individual. Hence, it shows more uniqueness than a face. The human identification system in [Xu et al, 2017] uses the entire profile of physical characteristic of an individual.

3.10 Signature

Signature defines the way in which one individual writes a specific word (mostly her name or a symbol). It is one of the oldest forms of
biometric and has gained acceptance for several application scenarios. With the progress of technology, two different kinds of signature biometric emerged. Offline signature considers geometrical features of the signature biometric. Online signature provides a cue about dynamic features of hand-written signature. These dynamic features include the speed and the acceleration of the pen movement, pen-up and pen-down times, etc.

3.11 Gait

The posture and the way a person walks maintain uniqueness about the person. It is non-invasive and hard to conceal. It can be easily captured even at public places at a low resolution. Unlike other biometric systems, the individual is not required to pay any attention while her gait is being captured. As the gait can be captured from a distance, it becomes very useful for security. It requires detection of the subject, silhouette extraction, extraction of features, selection of features and classification.

A lot of research has been carried out in gait recognition system. Body part (specially, feet and head) trajectories were introduced for extracting gait features. A gait energy image will lead to a set of view intensive features [Liu and Sarkar, 2006]. Recognition is carried out by a match of similar patterns in the gait energy image. Silhouette quality quantification has a key role in the method in [Han and Bhanu, 2006]. A one-dimensional foreground sum signal modeling is used in [Han and Bhanu, 2006] to analyze the silhouettes. Segmenting the human into components and subsequent integration of the results to derive a common distance metric has lead to an improved performance for gait recognition [Vera-Rodriguez et al, 2013]. Using a population based generic walking model, a gait recognition system attempts to solve the challenges which the system encounters due to surface time, movement speed, and carrying condition [Liu and Sarkar, 2005].
The concept of point cloud registration has been proposed in [Lee et al, 2009] in order to avoid the problem of occlusion during gait recognition. Moreover, face and gait characteristics have been extracted using principal component analysis from a side face image and gait energy image, respectively. A set of synthetic features have been attained by integration of these features and application of multiple discriminant analysis. Performance improvement is achieved in comparison to individual biometric features [Tan et al, 2006].

3.12 Voice

The voice recognition system extracts several characteristics of voice to identify an individual. Enrollment phase of voice biometric records the voice sample of an individual, extracts a template from it, and uses it for verification of the individual at later phase.

Apple’s Siri is a question-answering system based on a voice recognition technology. Mel-frequency cepstral coefficients (MFCC) and support vector machine is used to recognize an individual speaker. One of the drawbacks of this biometric is that a pre-recorded voice can easily be played back by an imposter for unauthorized identification. Moreover, a few specific kind of illness (e.g., catching cold) affects the voice and thus, causes hurdle for the voice biometric.

3.13 Key Stroke

It is believed that there is a pattern about how a person types on a keyboard. This behavioral biometric, which is referred as key stroke dynamics, shows traits to identify or verify a person. But it is not considered to be as unique as several other biometric traits.

4. Multimodal Biometric System

A unimodal biometric system identifies or verifies an individual based on a single biometric trait. Reliability and accuracy of
unimodal biometric systems have improved over time. But they always do not demonstrate desired performance in real world applications because of lack of accuracy in the presence of noisy data, non-universal nature of some biometric characteristics, and spoofing. The problems associated with unimodal biometric systems are discussed below.

4.1 Noisy Data

Lack of maintenance of sensors may introduce noise within biometric data. For example, typical presence of dirt is common in a fingerprint sensor. It generates a noisy fingerprint. Inability to present the original voice generates a noisy data too. Moreover, iris and face image may not appear as clear without an accurate focus of the camera.

4.2 Non-universality

In a universal biometric system, every individual must be capable of producing a biometric trait for recognition. But biometric traits are not always universal. An estimation reveals that about 2% of a population may not be able to produce a good quality fingerprint. Disability of individuals may cause problem for a smooth registration process. Successful enrollment is not possible for such individuals.

4.3 Lack of Individuality

Sometimes similar traits are extracted from a biometric system. For example, faces may appear quite similar for father and son, and even more for identical twins. As a consequence of lack of uniqueness, the false acceptance rate increases.

4.4 Susceptibility to Circumvention

Sometimes biometric traits are spoofed by an impostor. It is established how fake fingers can be generated by using fingerprints. These can be used to illicitly gain access to a biometric system.
Because of these problems, the error rates are, at times, high for unimodal biometric systems. Hence, they are not always acceptable for security applications. Multimodal biometric systems are conceived to tackle the above mentioned issues. In a multimodal biometric system, multiple biometric features are considered for recognizing an individual. In general, the usage of multiple biometric features contributes in an improved biometric recognition system. For example, a typical error can be caused by worn fingerprints. Presence of other biometric modalities may save the system from failure in the case of multimodal biometric. Thus, multimodal biometric system has less failure to enroll rate. It is considered to be main advantage of multimodal biometric.

Multimodal biometric system can be of three types based on how information is fused from various sources of information: (a) fusion of multiple representations of single biometric, (b) fusion of multiple classifiers of single biometric, and (c) fusion of multiple biometrics. Good recognition rate is achieved in a multimodal biometric system involving multiple evidences of a single biometric through fusion of multiple representations or multiple classifiers. But, to a true sense of multimodal biometric system, the use of multiple biometric traits is beneficial than usage of multiple forms of a single biometric in the terms of performance issues, including resistance to low quality samples, lack of individuality, user acceptance, etc. A detailed review of multimodal biometric system can be found in [Oloyede and Hancke, 2016].

Multimodal biometric systems are of three types: 1) multi-physiological, 2) multi-behavioral, and 3) hybrid multimodal systems. In multi-physiological category, only physiological characteristics (for example, fingerprint, retina, face, etc.) are fused. As an example, a multimodal biometric system in [Chang et al, 2003] combines face and ear biometrics. Most of the initial researches in multimodal biometrics belong to this category. Over the past few years, the rapid developments in human-machine interface have triggered an
evolution in behavioral biometric recognition. Hence, the field of behavior based multimodal biometric system has drawn attention of many researchers. A multi-behavioral biometric system in [Fridman et al, 2013] considers inputs from mouse, keyboard, writing sample, and history of web browsing. In [Bailey et al, 2014], another multi-behavioral biometric system considers inputs from graphical user interface interactions alongside mouse and keyboard inputs. Moreover, a hybrid multimodal biometric system combines physiological and behavioral features. Notable works on hybrid multimodal biometric include fusion of face, audio and speech using multiple classifiers by [Fox et al, 2007], fusion of face and gait by [Tan et al, 2006], and signature, face, and ear biometric fusion at the score level by [Monwar and Gavrilova, 2009]. Another good hybrid multimodal biometric system in [Paul et al, 2014] combines signature, face and ear biometric alongside social network analysis. It has been shown here that inputs from social network analysis further strengthen the biometric recognition system.

Contextual information (such as spatiotemporal information, appearance, and background) has a key role alongside soft biometrics (for example, height, weight, facial marks, and ethnicity) in identifying a person [Park and Jain, 2010]. In [Bharadwaj et al, 2014], face biometric and social contextual information have shown a significant improvement over performance in a challenging environment. It is to be noted that neither of extraction of appropriate contextual information or acquisition of soft biometric are easy tasks. These tasks may even require image processing. Moreover, social behavioral information is a common contributor in the normal recognition process in the human brain. [Sultana et al, 2017] administers a reinforcing stimulus in the form of social behavioral information to the matching decisions of traditional face and ear based biometric recognition system.
5. Levels of Fusion in Multimodal Biometric System

A detailed classification of various fusion techniques for multimodal biometric can be found in [Dinca and Hancke, 2017]. Fusion in multimodal biometric can occur at various levels – such as sensor, feature, matching score, rank and decision level fusion. Each of these is explained in this section.

5.1 Sensor Level Fusion

This fusion strategy directly mixes raw data from various sensors (Figure 6) – for example, from the iris and fingerprint sensors. Raw information is captured at several sensors to fuse at the very first level to generate raw fused information. Sensor level fusion strategies can be put into following three groups: (i) single sensor multiple instances, (ii) intra-class multiple sensors, and (iii) inter-class multiple sensors. In the case of single sensor multiple instances, a single sensor captures multiple instances of the same biometric trait. For example, a fingerprint sensor may capture multiple images of the same finger to reduce the effect of noise. Simple or weighted averaging, and mosaic construction are some of the common fusion methods in this case [Yang et al, 2005]. Multiple sensors are used to acquire multiple instances of the same biometric trait in the intra-class multiple sensors category [Yang et al, 2005]. For example, a 3D face image is obtained by using multiple face images taken from various cameras. In the case of inter-class multiple sensors, two or more different biometric traits are used together. For example, images of palmprint and palm vein can be fused together for biometric recognition.

Mosaicing is a nice application of sensor level fusion. Several researchers [Ratha et al, 1998; Jain and Ross, 2002; Ross et al, 2005] have proposed fingerprint mosaicing. It provides a good recognition accuracy as it combines multiple images of the same fingerprint. Therefore, it can handle the difficulty in recognition due to data quality. The fingerprint mosaicing technique uses a modified
Iterative Closest Point (ICP) [Jain and Ross, 2002] algorithm to generate 2D or 3D surfaces by considering the inputs from multiple instances. In [Fatehpuria et al, 2006], a touchless fingerprint system is developed using a 3D touchless setting with multiple cameras and structured light illumination (SLI) to generate 2D fingerprint images and 3D fingerprint shape. This kind of set up is expensive due to deployment of multiple cameras. Alternatively, usage of a single camera and two mirrors are suggested in [Choi et al, 2010]. Two mirrors have been used to obtain finger side views.

Sensor level fusion is generally applied for the same trait. There are also instances of applying sensor level fusion for different traits. Few of these are mentioned here. Face and palmprint images are combined in [Jing et al, 2007]. Pixel level fusion is preceded by Gabor transform of the images. Infrared images of palmprint and palm vein are fused in [Wang et al, 2008]. At first, image, registration is carried out on these images. Subsequently, a pixel level fusion takes place.

5.2 Feature Level Fusion

Features extracted from several biometric traits are integrated into a single vector. According to this fusion strategy, biometric sensor
signals (from camera or microphone) are preprocessed and feature vectors are derived from them independently. Then a composite feature vector is generated by combining these individual feature vectors (Figure 7). The feature level fusion exhibits a better performance than score level and decision level fusion techniques as feature level fusion techniques directly deals with the unique biometric features.

Normalization and selection of features are two important processes in feature level fusion. Min-max technique and media scheming based normalization is carried out to change the scale and location of feature values. Scale invariant feature transform is also carried out from the normalized images.

Dimensionality reduction through appropriate feature selection also enhances the accuracy of the techniques. Sequential forward selection, sequential backward selection, and partition about medoids are standard feature selection techniques. Particle Swarm Optimization (PSO) is applied on the feature vector for dimensionality reduction. The multimodal biometric techniques in [Raghavendra et al, 2009; 2011] uses this concept while combining face and palmprint features.

Figure 7: Feature Level Fusion
Incompatibility of the feature sets among different biometric traits and non-linearity of the joint feature set of different biometric traits poses challenges for feature level fusion. A feature vector can be generated using weighted average of multiple feature vectors if those vectors correspond to same biometric. For example, this becomes possible if all of these vectors are obtained from fingerprint images of an individual. If these vectors correspond to different biometrics, then they are concatenated to obtain a single vector.

Another example of feature level fusion can be found in [Kim et al, 2011]. Simultaneous use of time-of-flight (ToF) depth camera and near infrared (NIR) camera acquires face and hand vein images in a touchless acquisition set up.

Several multimodal system also combines face and ear, as ear is considered to be one of the most unchangeable feature of the human traits. Unlike face, human ear is not generally affected by age. PCA based feature extraction and a sparse representation method for feature level fusion is proposed in [Huang et al, 2013]. Experimental results reveal that this technique performs better than their unimodal components. Experiments also show that the performance is similar to that of the unimodal systems even if one of the modality is corrupted. Local 3D features (L3DF) are generated from ear and frontal face images in [Islam et al, 2013]. Feature level fusion is applied in these cases.

A matrix interleaved concatenation based new approach is presented in [Ahmad et al, 2016] for face and palmprint biometrics. Discrete Cosine Transform (DCT) is used here to extract the features. Then, these features are concatenated in an interleaved matrix which estimates the parameters of the feature concatenation and exhibits their statistical distribution.

A fingerprint and iris based multimodal biometric recognition technique has been proposed in [Gawande et al, 2013]. Minutiaei and wavelet features are extracted from fingerprint images. Haar
wavelet and block sum techniques produce features from iris images. A feature level fusion of these four feature vectors exhibit better performance than a unimodal fingerprint or iris biometric.

A feature level fusion of fingerprint and palm biometric traits has been proposed in [Mohi-ud-Din et al, 2011].

Another example of feature level fusion can be found for hand geometry recognition in the contactless multi-sensor system in [Svoboda et al, 2015] using an Intel RealSense 3D camera. This technique carries out foreground segmentation of the acquired hand image to determine the hand silhouette and contour. Then, the fingertips and the valleys are located alongside determination of the wrist line from the identified contour. Subsequently, two features vectors have been formed as follows: (i) comprising finger length, width, and wrist valley distances and (ii) finger widths as computed using a traversal of the overall hand surface and median axis to surface distances.

A finger and finger-vein based system has been proposed in [Yang and Zhang, 2012]. Gabor features are extracted and the feature fusion strategy is based on a Supervised Local-Preserving Canonical Correlation Analysis (SLPCCAM). In [Yan et al, 2015], feature level fusion has also been used for a contactless multi-sample palm vein recognition technique.

Automated access control systems in buildings and other secure premises are based on the capability of identifying an individual from a distance. Because of its importance in securing important establishments, it is emerging as an area of interest to the research community. Improper lighting condition or not-so-high resolution surveillance cameras pose a constraint for recognizing an individual based on her face. Use of multimodal biometric involving face and gait exhibits better performance [Ben et al, 2012; Huang et al, 2012]. Unlike traditional methods of face and gait multimodal biometric

A Robust Linear Programming (RLP) method [Miao et al, 2014] for multi-biometric recognition exhibits good results in noisy environments in spite of using less training data. It uses uncertain constraints and concatenates heterogeneous features from different biometric traits. Each biometric modality has been assigned a weight to specify its degree of contribution in the fusion. More weight is a greater relevance of the corresponding biometric trait.

In the multimodal biometric recognition system by [Chang et al, 2003], features are extracted from face and ear using Principle Component Analysis (PCA). Subsequently, fusion takes place at the feature level. A method of liveliness detection to prevent spoofing attack in [Chetty and Wagner, 2005] also uses a feature level fusion. In a recent development, sparse-based feature-fusion [Huang et al, 2015] of physiological traits has drawn sufficient interest of researchers due to robust performance.

### 5.3 Matching Score Level Fusion

In the case of matching score level fusion, matching scores are separately obtained for each biometric trait and subsequently fused to arrive at an overall matching score. The block diagram is presented in Figure 8. Matching score level fusion is also referred as measurement level fusion.
Figure 8: Matching Score Level Fusion

There exists three different approaches for matching score based fusion – density based, classification based, and transformation based. The density based scheme is on the basis of distribution of scores and its application to popular models like Naive Bayesian and Gaussian Mixture Model [Murakami and Takahashi, 2015]. In the classification based approach, the matching scores of individual matching modules are concatenated to obtain a single feature vector. The decision to either accept or reject an individual is based on the classification of this feature vector. According to the transformation based approach, the scores of individual matching module are, at first, transformed (normalized) into a pre-decided range. This transformation changes the position and scale parameters of the matching score distribution so that these normalized scores can be combined to obtain a single scalar score [Murakami and Takahashi, 2015]. These normalization techniques to handle the dissimilarities in matching score also draw attention of researchers.

In another noted development of recent time, an order-preserving score fusion method has been proposed in [Liang et al, 2016].
5.4 Rank Level Fusion

Ranking the potential matches between the query template and the templates in the database generates an ordered list of all templates in the database. The first choice is the match. These ranked are obtained for every biometric trait. In a multimodal biometric recognition system, these rank orders are fused to generate a final ranking of each template in the database. Unlike score level fusion, normalization is not required for rank level fusion.

Rank level fusion is applied in [Kumar and Shekhar, 2011] for combining various methods for palmprint identification. A Nonlinear Weighted Ranks (NWR) method aggregates the ranks as obtained from individual matching modules.

Rank level fusion may not always perform well in noisy conditions having low quality data. Though it has been applied on low quality fingerprints [Abaza and Ross, 2009]. It applies a derivation of the Borda count method, which involves image quality. This approach has a similarity with logical regression. But unlike logical regression, this approach does not need a training phase. Image quality is considered in this approach instead of weights.

As ranks are assigned to only few of the stored templates with possible match, the rank level fusion may create a challenge for large databases. Ranks do not cover every template in the database. In this context, a Markov chain based method has been proposed in [Monwar and Gavrilova, 2011] for rank level fusion. Markov chain is used to represent a stochastic series of events, where the present or the preceding states determine the next state. A graph is used to formally model the Markov chain. A vertex in the graph represents a state or an event. An edge in the graph denotes the transition from one state to another state. At first, ranks are generated for each biometric trait. If the matching module creates partial ranking (for example, the first three ranking results), elements are inserted randomly to complete the list. The state transition probabilities are
computed and the stationary distribution of the Markov chain is obtained. The templates in the database are ranked based on a decreasing order of the scores of the stationary distribution starting from the highest score. This fusion strategy is applied on a multimodal biometric recognition system involving iris, ear, and face. [Monwar and Gavrilova, 2009] proposes another method to solving this discussed problem of rank level fusion. The multimodal biometric recognition system has three matchers for each of signature, face and ear. However, fusion is carried out between the identities put out by at least two matchers.

5.5 Decision Level Fusion

In the case of a decision level fusion, each individual matcher, at first, takes its own decision. Subsequently, the fusion of various biometric modalities takes place by combining the decisions of these individual matchers. Hence, each biometric trait is independently pre-classified and the final classification is based on the fusion of the outputs of the various modalities (Figure 9). The simplest forms of decision level fusion uses logical operations such as ‘AND’ or ‘OR’. Some advanced fusion strategies at this level also use behavior knowledge space, the Dempster-Shafer theory of evidence, and Bayesian decision fusion.

Figure 9: Decision Level Fusion
When every individual decision module supplies positive outcome, then the ‘AND’ rule positively recognizes a query template. Otherwise, the ‘AND’ rule rejects the query. Hence, ‘AND’ rule is generally reliable with extremely low false acceptance rate (FAR). But false rejection rate (FRR) is higher than that of individual trait. On the contrary, the ‘OR’ rule provides positive output about a query template when at least one decision module gives positive response about it. As a result, FRR is extremely low and FAR is higher than individual trait. In [Tao and Veldhuis, 2009], an optimized threshold method has been proposed using the ‘AND’ and ‘OR’ rule. The thresholds of the classifiers are optimized during the training phase. Majority voting is another common approach for decision fusion. If the majority of the individual traits decide positively about the query template, then final decision is positive. Majority voting method gives equal importance to each individual decision modules. Otherwise, a weighted majority voting can be applied. In this method, higher weights are assigned to decision modules which perform better.

A multi-algorithm decision level fusion is used in [Prabhakar and Jain, 2002] for fingerprints. This method considers four distinct fingerprint matching algorithms. These are based on Hough transform, string distance, 2D dynamic programming, and texture. This method selects appropriate classifiers prior to applying decision fusion.

A threshold for each individual classifier influences the outcome of a decision level fusion. Here, threshold specifies a minimum score to decide whether the sample is genuine or an impostor. If the matching score of the sample is higher than the threshold, then the sample is considered as genuine. On the contrary, if the matching score of the sample is less than the threshold, then it belongs to an impostor. The classifiers are assumed to be independent of one another in some biometric systems. However, other works have assumed the dependency among the classifiers. A verification system has been introduced in [Veeramachaneni et al, 2008] based on two
fusion strategies for correlated threshold classifiers. Between these two strategies, Likelihood Ratio Test (LRT) still depends on the threshold of each individual classifier. The Particle Swarm Optimisation (PSO) based decision strategy is considered effective in comparison. Even the PSO strategy performs better than some of the score level fusion methods. A real time sensor management using PSO is suggested in [Veeramachaneni et al, 2005] for a multimodal biometric management. This method performs a real time search of the optimal sensor configuration and optimal decision rule. A similar concept is proposed in [Kumar et al, 2010] which uses Ant Colony Optimization (ACO) technique, for a multimodal biometric system involving palmprint and hand vein. [Kumar and Kumar, 2015] extends this experiment on multiple multimodal databases involving palmprint and iris, fingerprint and face, and face and speech.

Another decision level fusion for multimodal biometric recognition is proposed in [Paul et al, 2014]. In this multimodal system, signature, face and ear biometric features are combined with social network analysis. The Fisher image feature extraction, which is a combination of PCA and Linear Discriminant Analysis (LDA).

5.6 Hybrid Fusion Model

A hybrid fusion model which uses both pixel level fusion and score level fusion demonstrates good performance in [Kusuma and Chua, 2011]. This multi-sample face recognition (both in 2D and 3D) in [Kusuma and Chua, 2011] recombines images using principal component analysis (PCA). Two recombined images are fused using a pixel level fusion scheme. Additionally, score level fusion is also applied to produce good result.

6. Security and Privacy Issues in Biometric

There are several security and privacy concerns associated with usage of biometric. These are listed below:
• Biometric is not secret. Biometric data can be captured by a third party very easily. At times, even the original user may not be aware of the spying of her biometric data. For example, voice recording during a telephonic conversation or by a rogue mobile app can disclose the user’s voice biometric. Similarly, a video recording in the guise of surveillance cameras captures the user’s face or gait biometric.

• Biometric cannot be changed or revoked. Unlike password or pin, it is impossible to issue a new biometric trait, e.g., fingerprint. It is permanent. One-time compromise makes it unusable. Moreover, the user can never be dissociated with the compromised data.

• Biometric can be used to track an individual forever.

There exists eight different ways to attack a biometric system [Ratha et al, 2001]. These are demonstrated in Figure 10 with numbers which are explained below:

1. **Fake biometric**: Fake biometric is presented to the sensor with an intention of fooling the system. There have been several successful demonstrations of this kind of attack.

2. **Resubmitting stored signals**: The biometric signal can be pre-recorded and can be presented to the system at later times.

3. **Overriding feature extraction process**: The feature extraction module is compromised. The actual feature set is replaced with the desired one by an attacker.

4. **Tampering with biometric representation**: The templates representing the actual biometric trait are replaced with a desired one by the attacker.

5. **Corrupting the matcher**: The matching module is compromised and the attacker generates a matching score as desired by herself.
6. **Compromising stored templates**: An attacker may illicitly gain access to the stored templates. She can steal those templates to spoof the users’ identities.

7. **Communication interception**: The information being passed between the matcher and the database can be altered if an attacker intercepts the communication between these two modules.

8. **Overriding the final decision**: An attacker may override the decision being taken by the matcher.

![Figure 10: Biometric Attack Places](image)

All the above points indicate how a biometric recognition system can be compromised. But most of the fraudulent attempts take place in the context of faking a biometric data or tampering with the stored template. Fake biometrics can be recreated from the stored templates. Even it can be acquired directly from the sensor without the knowledge of the user. Moreover, compromising the database may lead to editing or deleting of the templates. These two topics are discussed here at length.

### 6.1 Faking Biometric Data

Numerous researches reveal how biometrics can be faked. Following are the three easy steps to spoof fingerprint data: (i) The residual fingerprint of a user can be obtained from a mobile phone or any other surface; (ii) Lifted fingerprint impression can be used to create a dummy finger; (iii) The dummy finger can be put on the fingerprint sensor to claim the identity. Demonstrations of this kind of attack
Biometrics against the fingerprint sensors of popular smartphone brands are publicly available [web1].

Biometric data is often captured in Internet of Things (IoT) devices with or without user’s knowledge. Hence, these devices represent the most danger to biometric. Biometric identity can be spoofed using the captured data. Recently, 5.6 million fingerprint records have been stolen from Office of Personnel Management (OPM) of US military. Such a large scale breach reveals devastating consequences of poor data security practices. Moreover, it strengthens the concern over storage of biometric data, as the OPM unit of US military is considered to enforce stringent security practices while storing biometric data in comparison to several private companies. Apple’s Siri, Google Now, and Microsoft Cortana record every uttering of a user and sends the data back to the servers of their organisations through Internet [Sparkes, 2015]. Samsung TVs automatically record conversations of the users to use these for automatic speech recognition [Matyszczysz, 2015]. Behavioral biometric is captured by many of the wearable devices.

Several forms of biometric data can also be captured through a smartphone. Hence, smartphones pose a risk to privacy. Though misuse of biometric data by big corporations is debatable, numerous third party applications installed in smartphones may appear as security and privacy risk. These third party applications often ask for more permission in the device than what is actually needed for them to complete their tasks. Permissions for accessing the camera and the microphone of a smartphone are the most misused ones [Felt et al, 2011]. These permissions enable the application to capture face, retina, and voice samples of the user. Moreover, many applications request for root permission to access every device sensor such as fingerprint, gait, heart monitoring and key logging for getting behavioral information about keystrokes and screen shots. These are even very handy in spying for user credentials. It is debatable whether these third party applications use the data with malicious
intensions. But their business models, in several cases, allow them to sell the users’ data to advertisers. With the increasing demand of users’ data in underground market, it may be more lucrative to sell users’ data than to make money through in-app advertising. As per a 2012 report [Labs, 2012] from Zcaler’s labs, over 40% of mobile applications communicate data to third parties. During installation of a new application, majority of the smartphone users do not check the permission requests. Even many of them may not be aware of the implications. Additionally, the entire permission system can be bypassed using root exploits [Zhou and Jiang, 2012]. Moreover, storage of data at third party servers poses a risk. Past cases show that there have been breaches even in military servers. Then, breaching the security of a small smartphone application vendor may not be challenging for the attackers.

A comprehensive review of biometric authentication in a smartphone is presented by [Meng et al, 2015]. Twelve types of biometric authentication can be carried out in a smartphone. Among those, following six are physiological: fingerprint, iris, retina, face, hand and palmprint. Remaining six types are behavioral in nature. These are signature, gait, voice, keystroke dynamics, touch dynamics, and behavioral profiling. A survey of successful attacks on smartphones is presented in this report [Meng et al, 2015]. Another successful attack to guess passwords using touch devices is also reported in [Zhang et al, 2012].

The proposals for securing smartphone authentication schemes, and authentication in general, are use of multimodal biometrics, check for liveness, combining with other authentication techniques (dual factor authentication) and use cancelable biometrics to store the templates. These proposals are useless if the user installs malware or a free application including a root kit that bypasses the permission system and captures biometric data from the user as he uses the smartphone.
Several problems exist with smartphone permission across various operating systems. Experts have also suggested improvements in this regard. But all of these efforts in security may fail at a single point, i.e., the user. This is because the user, sometimes, installs rogue applications without proper checking of the source or vendor. There exist numerous fake applications which pretend to be popular ones. Even the users, at several times, do not read the permissions while installing the applications. Hence, all good security practices fail. Certain permissions to an application provide the application an access to several device features and implicitly to data – biometric or otherwise. Even if the application owner is not misusing the data, there may be a breach by a malicious attacker to steal the data for illicit gains.

An artificial finger can spoof a fingerprint on many fingerprint sensors. Several researchers have demonstrated it in time and again [Cappelli et al, 2007; Galbally et al, 2010; Espinoza et al, 2011]. Research also suggests how this spoofing attack on fingerprints can be prevented [Marcialis et al, 2010]. Similarly, iris biometric is also vulnerable to spoofing through fake iris scans. Several techniques [Wei et al, 2008; Rigas and Komogortsev, 2015] suggests how to detect a fake iris. Another biometric very susceptible to spoofing attacks is face authentication by using pictures. There are a lot of techniques proposed to detect this issue [Maata et al, 2011; Komulainen et al, 2013; Pereira et al, 2014]. Hands geometry can also be spoofed by creating fake hands. [Chen et al, 2005] who proposed a practical model using plaster to create fake hands. The authors demonstrate that the fake hands can be created without the user knowledge from hand templates stored into the database. Other soft biometrics can be easily spoofed – voice can be easily recorded or spoofed artificially [Alegre et al, 2012], gait can be spoofed using a video camera from a distance to capture the user motion [Gafurov et al, 2007].
6.2 Template Security

Until one decade ago, it was believed that a stored template cannot recreate the original biometric data. Several researchers proved this wrong [Ross et al, 2007; Jain et al, 2008]. Encryption cannot be used to prevent a template compromise, as it is not possible to carry out recognition in the encrypted domain [Jain et al, 2008]. Tamper resistant storage in a smart card seems feasible for a single template for verification. Otherwise, it cannot be applied to large biometric databases. Solutions exist in the form of private templates [Davida et al, 1998] and cancelable biometrics [Ratha et al, 2007]. Still, several biometric recognition systems have not adopted these solutions to secure the templates in the database.

The concept of cancelable biometric to tackle the above stated problem was first proposed in [Ratha et al, 2007]. Several other template protection schemes have been developed subsequently. These schemes can be grouped into following two categories – cancelable transformations and biometric cryptosystems.

The characteristics of a template protection scheme are mentioned here:

- **Diversity**: The template of a biometric of same individual has to be distinct in different databases. It will prevent an attacker from gaining access to multiple systems through a compromise in one database
- **Revocability**: In the case of a compromise in an individual’s biometric template, it will be possible to issue a new template to her from the same biometric data
- **Security**: It will not be possible to recreate the original biometric data from a template. It is a one-way transformation
- **Performance**: There should not be any impact on the performance of the biometric system in terms of false acceptance rate (FAR) and false rejection rate (FRR).
[Jain et al, 2008] describes the advantages and disadvantages of each template protection type.

There is extensive literature on cancelable uni-biometric schemes and cryptosystems thoroughly surveyed by [Rathgeb and Uhl, 2011]. Even a subsequent chapter in this Staff Series discusses the issue of cancelable biometric at depth.

### 6.3 Other Types of Attacks on Biometric Systems

Similar to a brute force approach of password guessing, a brute force attack with a large number of input fingerprints can be used. The difficulty of such an attack is that the search space for guessing the fingerprint is prohibitively large. But for fingerprint authentication on several mobile devices, only a part of the full fingerprint is utilized. This provides the attacker with a much smaller search space.

On the contrary, a dictionary attack tries only those possibilities which are deemed most likely to succeed. Although dictionary attacks have been extensively studied and analyzed for traditional password-based authentication systems, they have not been systematically considered by the research community in the context of fingerprint verification. To perform a guessing attack with fingerprints, the question arises as to whether there are some fingerprints that are more likely to match a target than the others? It has been observed in the previous literature, that different users have different performance characteristics based on their fingerprint. [Yager and Dunstone, 2010] has introduced a menagerie consisting of dove (users with high genuine scores and low imposter scores), chameleons (high genuine scores and high imposter scores, thus are easy to match with everyone, including themselves), phantom (hard to match with most of the users), and worm (hard to match with themselves but easy to match with others). [Yager and Dunstone, 2010] have identified the existence of chameleons in datasets of full fingerprints.
A metric to estimate the strength of a biometric recognition system against impersonation attacks, namely Wolf Attack Probability (WAP), is proposed in [Une et al, 2007]. In this context, a wolf indicates an input sample which wrongly matches with multiple biometric templates. [Roy et al, 2017] shows how a master print can be located or generated. Such a master print can be used to match with multiple biometric templates. In case of partial fingerprints, the probability of detecting a master print and the attack accuracy increase. This finding reveals the risks of usage of partial fingerprints for authentication. In [Nagar et al, 2012], an evidential value analysis is carried out for latent fingerprints. It has been observed that less number of minutiae points or a small surface area of the latent fingerprint leads to a low evidential value of the fingerprint. Hence, probability of matching error increases in these cases. Based on this finding, an estimate in [Roy et al, 2017] shows that the probability of finding masterprints that incorrectly match with a large number of templates is high for partial fingerprints.

7. Usage of Biometric in Banks and Financial Institutions

Amidst all these practical challenges and concerns over security and privacy issues in biometric, banks and financial institutions have taken significant steps in embracing it. Banks in India have also embraced biometric as one of the authentication factor. Here is a small illustrative list of initiatives by banks in India:

1. **DCB Bank**: Fingerprint based cash withdrawal is possible from the ATMs of DCB Bank. These ATMs connects with Aadhaar database to authenticate a customer using her fingerprint as enrolled with Aadhaar.

2. **Federal Bank**: A zero balance selfie account opening is possible with Federal Bank using its banking app. For an instantaneous account opening, a user scans her Aadhaar and PAN cards and then, clicks her selfie photo.

3. **HDFC Bank**: As part of financial inclusion initiatives by HDFC Bank, the bank has introduced fingerprint verification using a
hand-held device or a micro-ATM. Fingerprint based verification with the Aadhaar database enables the bank for instant KYC (know your customer) check for its users.

4. **ICICI Bank**: Voice recognition enables the customers of ICICI Bank to interact smoothly with the bank’s call center. During such an interaction, authentication credentials are not being asked to the customers, as their voice can authenticate themselves.

5. **State Bank of India**: A fingerprint based authentication is carried out in order to provide the bank’s employees an access to the core banking system.

There are several use cases of biometric in banking and financial service industry around the globe too. Fingerprint is commonly being used by mobile banking applications to authenticate a user for last few years. For example, Bank of America, Chase and PNC are some of the institutions which have adopted fingerprint based user authentication for their mobile applications. Master Card has launched a ‘selfie-pay’ to authenticate online purchases through either face or fingerprint recognition. Citi has registered its customers’ voice samples.

USAA, which serves to members of the military and their families in United States, has rolled out three different biometrics – fingerprint, face and voice recognition – for customer authentication. Pictet and Banquiers (one of the leading banks in Switzerland) has deployed an efficient 3D face recognition system for providing access to its staff within the bank’s environment.

8. **Conclusion**

Biometric recognition is gaining popularity for identification and verification of individuals through their specific physiological and behavioral traits. In certain scenarios, its importance is perceived in the form of a second factor of authentication, in addition to
knowledge based or possession based authentication requiring security for transactions. It enables the government as well as private and public businesses to reduce identity theft and related crimes.

In this article, the strengths and weaknesses of several biometric recognition systems have been discussed through a comprehensive review of the developments in this field. Furthermore, both unimodal and multimodal biometric systems have been discussed. Various types of fusion strategies have also been explained in the context of multimodal biometric systems. This article also discusses various security and privacy concerns which are associated with usage of biometrics. A compilation of the progress in this field helps the readers get an overall grasp.

References


Y. Huang, D. Xu, and F. Nie, “Patch Distribution Compatible Semi-Supervised Dimension Reduction for Face and Human Gait Recognition,” IEEE


Y. Li, K. Xu, Q. Yan, Y. Li, and R. H. Deng, “Understanding OSN-based Facial Disclosure Against Face Authentication Systems,” in Proc. ACM Symposium
on Information, Computer and Communications Security, New York, USA, pp. 413-424, 2014


G. Marcialis, F. Roli, and A. Tidu, “Analysis of Fingerprint Pores for Vitality Detection,” in Proc. 20th International Conference Pattern Recognition (ICPR), pp. 1289-1292, August 2010


D. Tan, K. Huang, S. Yu, and T. Tan, “Efficient Night Gait Recognition Based On Template Matching,” in Proc. 18th International Conference on Pattern Recognition (ICPR), August 2006


Voice Based Authentication

- Dr. V. N. Sastry,
  Professor, IDRBT
Executive Summary

Scientific and technological developments in sound representation, audio media recording, storage, transmission and reproduction have helped people to avail various voice based services such as voice communication, listening to speech and music, dictation by converting speech into text in a language, giving oral instructions, production of speech from text in a language, converting speech from one language to another, etc. Tele and Interactive Voice Response System (IVRS) based services have helped organisations and Government to provide multiple services remotely to customers and citizens, for example, Telebanking and Railway Enquiry. As mobile phone has become a smart companion of an individual globally with strong capabilities of voice processing, it has become necessary to investigate whether a mobile phone user can be authenticated by his/her voice to get various mobile services such as m-Governance and m-Banking. Research in voice based speaker authentication needs to address – (i) can voice be treated as a unique factor of biometric authentication or as an additional factor of authentication, (ii) can voice authentication be guaranteed within human perception limits of tolerance during real time dialogue, (iii) can voice authentication be provided as a centralized service for large population, like in UIDAI or separately only by organisations, (iv) can voice authentication process confirm whether it is the same or is it mimicry or a device. Here, we present the importance of voice-based services and multilingual services, elements of voice transmission, speech processing, voice based speaker authentication, and some results of voice based experiments carried out at CMB. We also present some proposed architectures that would be mainly useful to providers of mobile services including Government and Banks to offer efficient and secure voiced based mobile services such as Mobile Governance and Mobile Payment Services.
1. Introduction

The scientific and technological developments in voice representation, reception, production, transmission, storage and replay have played significant role in rendering voice based services to people across the globe. Various voice based products such as Telephone, Mobile Phone, Voice Recorders, Music Players, Amplifiers, Multi-media players, etc., have helped a common person to get multiple types of services such as communication, learning, entertainment, business, etc. Organisations and Government offer various voice based business and public services to their customers and citizens through Tele Services, Interactive Voice Response System (IVRS) Services and Mobile Services, for example, tele-banking, voice guidance for booking, searching for an audio clip by voice phrase, rendering online music, mobile governance services, audio and video conferencing service, etc.

People find it more convenient to give voice instructions and verbal commands to navigate, control applications or request for services compared to typing text because typing binds ones hands and visual attention due to physical or virtual keyboard, lighting and posture requirements. Converting speech into text by mobile phone is very useful for hearing impaired, children, senior citizens and illiterates. Listening is more convenient than reading long text in a mobile phone. Hence, converting text into its audio form by mobile phone helps one to listen while mobile, in darkness and particularly useful for visually impaired/blind. Voice based services help people to directly enquire, avail services round the clock and get faster responses, particularly on mobile phones. Voice based services help organizations to reduce paper work, reduction in staffing costs and provide multilingual support to customers.

Mobile phone has become a personal assistant of an individual to get various mobile services. To offer personalized and sensitive services, authenticating the person becomes essential. User can be
authenticated by three factors, namely, knowledge, possession and biometric factor. Voice is a crucial biometric of an individual. Authenticating a person based on his/her voice within minimum acoustic tolerance period is a challenge. Here, we focus on voice based authentication and related voice based mobile services.

In Section 2, we present the elements of voice communication, signals, various devices used in voice communication and their functions. In Section 3, speech recognition aspects of converting speech to text and speech synthesis aspects of converting from text to speech are discussed. In Section 4, authentication process and four associated problems namely, speaker verification, identification, clustering and classification are analyzed. The use and design of IVRS for mobile services is given in Section 5. Development of various mobile services as required in multiple languages of India and some experimental results of voice related works carried out at CMB are mentioned in Sections 6 and 7 respectively. In Section 8, conclusions are presented. In the Annex, various figures are presented for clear understanding of various processes of voice communication, voice authentication and proposed solutions of voice based mobile services. Mathematical Equations, Transforms, Algorithms and Data Analytics part are excluded for general reader’s sake.

2. Voice Based Services

**Tele Banking** enables a customer to call to the Customer Care Centre of the bank using telephone, mobile phone or a computer supporting Voice Over Internet Protocol (VoIP) and get various banking services. These services are provided by the bank through an automated system or by assisted customer service representatives. Voice call is relatively easy as there is voice guidance for options to choose from menu options provided. It is beneficial to banks as multiple customers can be supported concurrently. Customers find it convenient as they can get help round the clock and from anywhere
without getting restricted to personally visiting the bank during only office hours.

**Interactive Voice Response System (IVRS)** is a communication system which provides automated telephone call access to specified computer database information. It allows computer to detect DTMF (Dual-tone Multi-frequency) keypad inputs and generate voice response according to input. The most common way for a phone to communicate with a computer is through the tones generated by each key on the telephone keypad. These are known as **Dual-Tone Multi-Frequency (DTMF)** signals. Computer needs a hardware called **Telephony Card**. Speech recognition is included in the system to understand names, phrases and numbers ([Figure 19](#)). Text-To-Speech (TTS) software can also be employed to fully automate outgoing messages. Instead of recording the voice prompts, TTS can be used to convert text to speech. IVRS interacts with callers, gathers information and routes calls to the appropriate recipient. IVRS accepts a combination of voice telephone input and touch-tone keypad selection given by the users and provides appropriate responses in the form of voice, fax, callback, e-mail and perhaps other media links. These attributes can be best utilized for mobile banking applications like providing financial services through the banks to their customers. IVRS installed in a server or host computer can cater to all the services, for different channels like USSD, SMS, etc. Mobile banking plays a predominant role in further improving the services with customer convenience. Using Computer Telephony Integration (CTI), IVR applications can hand off a call to a human being who can view data related to the caller at a display. It can also provide easy access to powerful branching logic, online access to third-party applications and database, access to real-time information on the contact centre, Text-To-Speech, CTI integration, and massive personalisation.
Bell laboratories introduced the first telephone that could dial using DTMF technology at Seattle World Fair in 1962. DTMF telephones enabled the use of in-band signaling, i.e., they transmit audible tones in the same 300 Hz to 3.4 kHz range occupied by the human voice. Despite the increase in deployment of IVR technology in the 1970s to automate tasks in call centres, the technology was still complex and expensive. Early voice response systems were Digital Signal Processing (DSP) technology based. The mature technology allowed clusters of 96 channels of high-density digital phone interface gear (terminating four T1 lines of 24 channels each) to each being controlled by one application processor, running individual applications, one per channel, accepting DTMF touch tone inputs, accessing a large stored vocabulary for output, recording and playing back user speech, when necessary.

In 1990s, call centres began to migrate to multimedia. Companies started to invest in Computer Telephony Integration (CTI) with IVR systems. IVR became vital for call centres deploying solutions for universal queuing and routing and acted as an agent which collected customer data to enable intelligent routing decisions. With improvements in technology, systems could use speaker-independent voice recognition of a limited vocabulary instead of requiring the person to use DTMF signaling. In the subsequent decade, voice response started to become more common and cheaper to deploy. Presently, voice based search and voice based authentication services are becoming more prominent.

**Mobile Services** such as Mobile Governance services to a citizen and mobile banking services to a customer can be provided through voice in multiple languages to suit Indian requirements of linguistic diversity. Developing Mobile Applications supporting voice is recommended for effective usability. A mobile app can provide responses as tutor to the user’s questions and execute the commands of the user. It can recognize the verbal inputs using Speech-to-Text conversion and respond back using Text-to-
Speech conversion. Besides the Voice Actions, the respective Application Programming Interfaces (API) can serve the purpose of user defined voice commands to offer several mobile services. Such developments will not only enhance the overall voice integrated mobile services domain, but help the differently-abled to carry out the required transactions in an effective way. Similar to Aadhaar number linked fingerprint and iris based citizen authentication service by UIDAI, voice based centralized authentication services can also be provided to citizens to get various mobile services and through IVRS in multiple languages.

3. Voice Representation, Transmission and Reproduction

3.1 Sound, Voice and Speech

**Sound:** Sound is produced by a physical entity due to vibration, speedy movement, sudden expansion or impulsive contraction, for example, playing any string instrument, fountain, explosion of crackers, clapping of hands or hammering. Variety of sounds, which are produced through different sources such as living creatures, instruments (wind or string), thunders, crackers, whistle, horns, sirens, machines, vehicles, snoring, laughing, electronic speakers, etc., can be realized through sensory receptors. The frequency range of sound produced from the mouth and the frequency range of acoustic reception of the ear of a person are very small compared to the overall frequency spectrum of the electromagnetic waves due to limitation in the physical capacity of vibrators in these organs. Sound originating from a source propagates in the form of waves in all radial directions and its intensity gets damped down in the medium as its distance from the source increases.

**Voice:** Voice is modulated sound of creatures embedded with emotions. Voice of an individual is the output of a complex mechanism useful for expression and communication (Figure 1).
Voice is an intrinsic and unique biometric of an individual. Instrumental Music is rhythmic form of voice, which is soothing to ears.

**Speech**: Speech of a person is modulated voice uttered verbally in a language. Natural speech is spoken expression in a language consisting of meaningful sentences and embodiment of emotions. A sentence is an ordered sequence of words. Words are meaningful units of a language forming parts of speech. These are pronounced through syllables consisting of vowel and consonant letter sounds. The accent and intonation of speech, spoken output of a person, depends upon perception of listening of pronunciation, structural characteristics of production mechanism of voice, understanding of language, vocabulary, translation rate, context, training and age. Vocal songs are rhythmic forms of speech (Figure 1).

Natural voice communication is constrained by distance, physical medium and interference, for example, generally two individuals can directly speak to each other clearly in aerial medium only if they are within 100 feet apart. Since direct verbal conversation is of broadcasting nature, it cannot be confidential. To overcome these barriers, researchers worldwide have explored, developed technologies and facilitated everyone with gadgets of modern voice communication and mobile phone devices. The way human deciphers speech is by processing and then storing whereas machines first store and then process. Hence, they differ in storing and retrieving capabilities. Sound is of wave form, which is represented as signals.

### 3.2 Signals

Sound propagates as waves. A signal is the abstract description of unseen physical waves that vary with time, space or any other independent variable. It is represented as a mathematical function of one or more independent variables in continuous time or in discrete time domain. Natural Voice is basically analog signal with
continuous amplitude and continuous time (Figures 4,5,6). For machine processing, it is discretized through sampling techniques and then digitized through coding techniques. Analog signals generate more data, they require more effort for accuracy, and system hardware is relatively expensive. Digital signals can be handled conveniently with flexible operations, as they are more accurate with lesser effort, easily portable and consume less storage, and system hardware is cheaper and much of the work is implemented through software. Digital signals are significant but have the following issues – System complexity is increased as other functionalities are introduced; bandwidth is limited by sampling rate since systems having wide bandwidths, require fast sampling rate, and components dissipate more power as compared to analog counterpart. Its simplistic forms are Sinusoidal signal, Square wave, Triangular wave, Rectangular wave, Saw tooth wave, etc. Types of standard input signals are Unit Step Signal, Unit Ramp Signal, Exponential Signal, Impulse or Delta Signal.

Continuously oscillating and periodic signals are used to send data. Voice communication should not have delay and data communication should not have errors. A carrier signal or carrier, is a waveform (usually sinusoidal) that is modulated (modified) with an input signal for the purpose of conveying data. This carrier is usually a much higher frequency signal than the input signal. The purpose of the carrier is either to transmit the data through space as an electromagnetic wave (as in radio communication) or to allow several carriers at different frequencies to share a common physical transmission medium. Long range (Cellular, WiMAX, VSAT, Microwave, etc.) and short-range (Bluetooth, Wi-Fi, Zig-Bee, Infrared, NFC, etc.) wireless technology has become common utility of signalling and digital communication embedding both voice and data. Cellular telephony is a radio-based electromagnetic signal propagation for long distances in the 850 MHz/900 MHz and 1800 MHz/1900 MHz frequency bands in India. Signal processing operations are mainly time and amplitude scaling, multiplication,
addition, shifting, folding, convolution, filtering, interpolation and decimation (Figure 2). There are four basic scenarios of voice communication between persons and devices as in the table below:

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Voice Production (Source)</th>
<th>Audio Reception (Destination)</th>
<th>Example of Audio Communication</th>
<th>Interaction Type (Person P, Device D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>Yes</td>
<td>Direct person to Person Dialogue</td>
<td>P2P</td>
</tr>
<tr>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>Person Speaking and Device Recording</td>
<td>P2D</td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>Playing Electronic Device (Keyboard) and Person Listening</td>
<td>D2P</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
<td>Device Playing (Transmitting) and Device Recording (Receiving)</td>
<td>D2D</td>
</tr>
</tbody>
</table>

3.3 Devices and Functions

Voice communication and services use various transformations, functions and devices. Voice based technology devices such as Mike, Microphone, Antenna, Modulator, Synthesizer, Amplifier, Speaker, Sound box, Earphone, Headphone, Telephone, Mobile Phone, Voice Recorder, etc., have specific purpose. For example, **antenna**, a metallic device for radiating and receiving radio waves. It is a transceiver as it transmits and receives a signal. It is also called as a transducer as it converts electrical signal into electromagnetic waves at transmitting side and vice versa at receiving side. Antennas are of various types such as wire antennas, aperture antennas, micro-strip antennas, array antennas, reflector antennas, lens antennas. They may be (i) **Directional Antenna** – which radiates in any one direction. It has single peak in radiation pattern and doesn’t have symmetry.
Ex: Dish antennas, slotted waveguides; (ii) Isotropic Antenna – which radiates equally in all directions, although practically difficult; (iii) Omni directional Antenna – which radiates in more than one direction. Different communication equipment when assembled together form a communication system as given in (Figure 2). It consists of source, transmitter, channel, receiver and destination as given in (Figure 3). Presently, voice communication is inter-continental and extra-terrestrial. Functions of these corresponding blocks are briefly given below.

**Information Source:** A communication system serves to communicate a message or information. This information originates in the information source. In general, there can be various messages in the form of words, group of words, code, symbols, sound signal, etc. However, out of these messages, only the desired message is selected and communicated. Therefore, we can say that the function of information source is to produce required message, which has to be transmitted.

**Source Encoder:** A source encoder converts client input to a binary stream. This is typically done by sampling an analog input, digitizing the samples and coding them into a binary stream. A common component that does all this is called a Codec (Figure 9). A number of diverse voice-encoding algorithms ( codecs) are being used and they have been standardized by the ITU-T as a series of recommendations known as the G-series. These codecs differ in the algorithms they use for sampling the analog voice signal and in the sophistication of the compression used. Speech coders utilize the properties of human speech production and perception systems to achieve the bit-rate reduction. Wideband speech codecs work with 16-kHz sampling rate and above for achieving good quality speech with bit rates comparable to narrow band codecs. The following Narrowband – G.711, Narrowband- G.729AB and Wideband- G.722 are popularly used speech codecs in VoIP, standardized by ITU-T.
**Channel Encoder**: A channel encoder tries to correct errors within transmitted message over the channel. It basically refers to adding some extra bits in the form of parity bits so that one can protect the data from becoming corrupt. A simple example can be that of a checksum being sent across the message. The checksum allows the channel to correct some of the errors in the message.

**Modulator**: A modulator is a device that performs modulation. Modulation is the process of changing the characteristics of the carrier wave in accordance to the characteristics of the actual message signal that typically contains the actual information that has to be sent. A modulator essentially adds the information of the message signal into the carrier signal and transmits the carrier signal. Modulation is generally done at the transmitter side. Modulation is the process of conveying a message signal, for example a digital bit stream or an analog audio signal, inside another signal that can be physically transmitted.

The aim of **analog modulation** is to transfer an analog baseband (or lowpass) signal, over an analog bandpass channel at a different frequency. The aim of **digital modulation** is to transfer a digital bit stream over an analog bandpass channel or over a limited radio frequency band. Analog and digital modulation facilitate frequency division multiplexing (FDM), where several low pass information signals are transferred simultaneously over the same shared physical medium, using separate passband channels (several different carrier frequencies), ([Figures 10 & 11]).

**Channel**: The term channel means the medium through which the message travels from the transmitter to the receiver. In other words, we can say that the function of the channel is to provide a connection between the transmitter and the receiver. It may be guided (physical link is present as wire) or unguided (no physical link is present as wireless). A base station or access point need to share its limited channel capacity to serve bandwidth to multiple mobile
terminals at the same time (both downlink and uplink), which is done by geographically splitting the coverage area as cells and multiplexing various signals of user’s devices. Frequency Division Multiple-Access (FDMA), Time Division Multiple-Access (TDMA), Code Division Multiple-Access (CDMA) and Space Division Multiple-Access (SDMA) are the major multiple access techniques that are used to share the available bandwidth in a wireless communication system. Depending on how the available bandwidth is allocated to the users, these techniques can be classified as narrowband and wideband systems.

**Demodulator:** A demodulator (detector or demod) is a device that performs demodulation, the inverse of modulation. A demodulator is a circuit that is used in receiving side to separate the information that was modulated onto the carrier from the carrier itself. A demodulator is the analog part of the modulator. A modulator puts the information onto a carrier wave at the transmitter end and then a demodulator pulls it so that it can be processed and used at the receivers end. A **modem** (from modulator–demodulator) is a device that can perform both operations of modulation and demodulation.

**Channel Decoder:** The channel decoder recovers the information bearing bits from the coded binary stream. The channel decoder also performs error detection and possible correction.

**Source Decoder:** The source decoder converts the binary output of the channel decoder into a symbol sequence. The decoder for a system using fixed length code words is quite simple, but the decoder for a system using variable length code words is complex. The aim of the source coding is to remove the redundancy in the transmitting information, so that bandwidth required for transmission is minimized.
Destination: Finally, the receiver at the destination receives the message. Speaker System is an audio output device, which converts digital sound to analog sound and produces audio output. Stereo phone speakers have better sound clarity.

3.4 Characteristics of Voice and Speech

Voice is an analog signal. Figures 4, 5, 6 and 13 represent a typical voice signal. It has different amplitudes at different instances of time. It can be characterized by various parameters such as amplitude, time, phase, wavelength, frequency, angular velocity, noise, energy, pitch, intensity, duration, etc. Speech signal is assumed to have prominent energy than noise in the signal. In case of extremely noisy environment where energy of noise is sufficiently higher than the speech signal itself, filtering is necessary. The voice of an individual depends on many components like pitch, sound intensity, duration, etc. Speech is embedded with information, context, perception, emotions and expressions, so the effect on the listener is strong. Intonation and Resonance can be characteristics or artifacts of speech signature of a person. Speaker specific voice features and recorded speech features need not exactly be the same. Voice has mainly three categories of features.

(A) Spectral Features

Liner Prediction Analysis: Under the linear prediction analysis voice is represented by a simple model which is based on speech production. Here, vocal tract is modelled as a tube which is same in length and has constant diameter. It is assumed that the glottis produces buzzing sound or noise. Up to a certain condition, there is no energy loss and nonlinear effect inside the vocal tract. Hence, vocal tract transformation can be represented by the Z-transform.

Mel-Frequency Cepstral Coefficient (MFCC): According to the Filter Bank model, due to the ability of human ear, it can resolve the
frequencies nonlinearly across the audio spectrum and it decreases the higher frequencies into lower frequencies. Filter Bank is a set of band pass filters. A band pass filter allows only limited set of frequencies by eliminating all the lower and higher frequencies components, for example, triangular filters (Figures 12 & 14).

(B) Prosodic Features

In the formal linguistic model of a speech communication, the timing and rhythms of a speech play an important role. Prosodic features are generally based on tone and rhythms in the speech. These features depend on the shape of the vocal tract shaping factor. Human upper articulator’s moments are related to vocal-tract shaping factor. The source factors are changed in speech, breathing muscles and vocal folds.

Pitch is generated by the vibrations of vocal folds. There are two common features of the pitch signal – one is the pitch frequency, and other one is glottal air velocity. Rate of vibration of vocal folds is equal to the fundamental frequency of phonation, which is also called pitch frequency. During the vocal fold vibration, the velocity of air is termed as the glottal volume velocity. The pitch signal is commonly estimated by an algorithm based on auto-correlation. This algorithm works by low filtering of the signal (at 900 Hz) and then breaking down the signal into shorter-time frames of speech, after which a nonlinear clipping method is used in each frame for checking the first formant interference with the pitch.

Energy: Energy is calculated as the mean of signal energy. The short term energy of speech frame is duration.

Duration: Features based on the duration is dependent on the temporal lengthening of words. Along with total duration of the
word, number of syllables in the speech, positions on the time axis of energy and pitch graphs could also be used as duration features.

(C) Cepstral Features

A speech signal is often considered as an output of a Linear Time Invariant (LTI) system – the source (vocal cord), the tract filter and the radiation parameters. These filters can be modeled by deconvolution of the speech signal using cepstral analysis.

Mel-Frequency analysis is based on human perception of speech. Human ear concentrates only on certain frequency components and the filters are placed logarithmically along the frequency axis i.e. more filters are in the low frequency regions and less in high frequency regions. After performing cepstral analysis over Mel Spectrum, we get Mel-Frequency Cepstral Coefficients (MFCC). Thus, a speech can be represented as a sequence of cepstral vectors. MFCCs contains the information about the dominant frequency and its formants which in turn contains identity of voice and thus, the cepstral vectors can directly model the voice-prints of the user (Figures 14) for MFCC computation steps as briefed below:

Frame Blocking: The input signal is split into frames of length 20-30 ms (milli seconds) with a 33% overlap. Overlapping is mainly done to produce continuity between various frames.

Pre-emphasis: The speech signal is sent to a high pass filter, which ranges from 0.9 to 1.0. The Z-transform can be used in pre-emphasis filter to emphasize high frequency components, which is normally suppressed during speech production.

Hamming Window: Each of the frames is then multiplied with a window to keep the continuity between consecutive frames.
**Fast Fourier Transform**: The Fast Fourier Transform is performed to obtain magnitude of each frequency component in the signal.

**Mel Filter Bank**: After performing Fast Fourier Transform, the signal is then passed through a bank of filter called Mel Filter Bank. From center frequency, 133.33Hz to 1kHz, there are 13 filters with overlap and from 1kHz to 8kHz, there are 27 filters placed logarithmically.

**Logarithm**: Logarithm of the output signal from Mel Filter Bank is taken. This is mainly done to incorporate human perception of sound.

**Discrete Cosine Transform (DCT)**: A discrete cosine transform is applied to triangular band-pass filters to get Mel-scale cepstral coefficients and respective DCT coefficients.

**Liftering**: It is the weighted function of Mel-scale coefficients used to increase the accuracy of recognition.

**Deltas**: It represents the change in MFCCs with respect to time and it represents the trajectory of MFCCs. When MFCCs up to 13 coefficients are considered, then the key feature vector is 13 dimensional.

**Double Deltas**: This represents change in deltas and thus computed by taking difference between consecutive deltas. Deltas and double deltas tend to improve the accuracy of the model. It represents the rate of change of various vocal tract filters, which also contributes to identity of the person.

4. **Speech Processing**

Speech is mainly composed of: *(a) Voiced speech* – It is produced by actual vibration of vocal chords as well as constriction of articulatory
organs, (b) Unvoiced speech – This is mainly produced by constriction of articulatory organs, (c) Silenced portion – Here neither voiced speech nor unvoiced speech is present and the energy of signal is very low, typically less than 30 dB (Decibel). A given speech signal contains data rate of typically 64kbits/s which is relatively very high as compared to the original message, which is typically of order 100bit/s implying it contains redundancy thus a pre-processing is needed before it is used for any application. Normally, human speech is low pass in nature and contains frequency from 20Hz to 20kHz, but the power contained beyond 4kHz is significantly less and contains very less information. Thus, according to Nyquist rate, a sample rate of 8kHz would be sufficient for a normal speech signal. Also, each of the samples are then quantized to 8-bit or 16-bit data, so that it can be processed digitally. Raw audio signals contain redundant data which needs to be processed first so that it can be used for any application. Voice Activity Detection (VAD) is performed to remove silenced portion from the sample and process only those portions, which contain speech. To achieve this, one may use energy based approach, typically in a silent environment the noise is less than 30 dB and normal conversation are at 60 dB, this very fact is exploited to classify speech and speechless portion.

Initially, the user is requested to remain silent for a duration of three seconds where the background noises are recorded and a probabilistic model is created. During processing of speech, the same model is used to classify a frame as a speech or silenced frame and MFCCs are computed only on frames that contain speech. The threshold criteria for removing silenced portions from the signal is based on the noise mean in various frames and the noise variance. Some transformations which form an integral part of every preprocessing in case of speech signals are:

**Fourier Transform:** A Fourier Transform represents a time varying signal with a linear combination of sinusoids. For a time varying
signal, Fourier Transform and the inverse Fourier Transform are used. Fourier Transform has a limitation, it has a very poor time resolution, thus it fails with non-stationary or quasi-stationary signals and it has been shown that speech signals are quasi-stationary signals therefore short term Fourier Transform is used in case of speech signals. A window of fixed width is taken and then the signal is decomposed into its sinusoids, since the statistical parameters remain constant for that window. Thus, it can be used for various applications (Figure 8).

**Discrete-time Fourier Transform (DFT):** Here, the actual voice signal is first sampled at some frequency greater than twice the bandwidth of the signal and a frequency analysis is performed, which gives information about various frequencies present in the signal.

**Short-time Fourier Transform:** Since voice signals are quasi-stationary in nature, that is, its statistical parameters varies over time, but they remain constant for a small period. Thus, the signal is first windowed and only then Fourier Transform is applied over it. This is practically more useful in speech processing due to its quasi-stationary nature and various encoding schemes use this technique to process or store speech signals.

Speech is considered as one of the most prominent biometric factor. It is the ability of a machine to respond by either identification or verification procedures. This kind of procedure provides a smarter and reliable way of authentication for banking applications. Voice recognition has become so advanced and mainstream that it is applicable even in sensitive applications as financial services. Technological changes have made voice recognition software and hardware equipment much user friendly. Speech processing can be classified into (A) **Speech Recognition** – It is speech-to-text (STT) conversion and (B) **Speech Synthesis**- It is text-to-speech (TTS) conversion. Language and Speech Processing gives rise to various scenarios given in the table below.
<table>
<thead>
<tr>
<th>Input</th>
<th>Output</th>
<th>Scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Text</td>
<td>Speech</td>
</tr>
<tr>
<td></td>
<td>Language -X</td>
<td>Language -Y</td>
</tr>
<tr>
<td>Text in Language -X</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Language Translation</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Speech Synthesis</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Speech Synthesis with Translation</td>
</tr>
<tr>
<td>Speech in Language -X</td>
<td>Yes</td>
<td>Speech Recognition</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Speech Recognition with Translation</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Speech Modulation</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Speech Translation</td>
</tr>
</tbody>
</table>

**Speech Representation**

Generally, speech can be represented using three different approaches, namely, (a) Frame based, (b) Segment based and (c) Landmark-based representation, which are explained below:

(A) **Frame-based representation**: In this representation, the given speech signal is divided into frames of equal length. Frames are obtained by considering the speech signal within a window. The position of the window is decided based on the starting point and shift in window. Proper choice of frame size and shift along with the shape of the window is crucial for the analysis of the speech signal. Due to variabilities in speech, proper choice of these parameters is difficult. Generally, Hamming window is...
used with a size of 20-30 msec and a shift of 5-10 msec. Features are extracted in each frame and are used for further processing of the speech signal. In frame-based representation, fixed set of feature attributes are extracted in all the frames, and the features are not changed from frame to frame.

(B) **Segment-based representation:** In this representation, segmentation is performed by finding boundaries in the speech signal. These boundaries delimit unequal length, semi-steady state, abutting regions, with each region corresponding to a phone or a sub-phone unit. Segmentation does not involve the subsequent process of associating phonetic significance to the obtained segments. Post segmentation, analysis is performed either within each segment or at the boundaries. Unlike frame-based representation, the features extracted in segment-based representation can differ across segments depending on the characteristics of the segment. Hence, segment-based analysis is more structured than frame-based analysis. Also, segment-based approach performs better or is comparable to frame-based approach, while reducing the computation load in training and testing by a significant amount. The problem arises when segmentation is to be performed on speech signals with no sharp boundaries like those corresponding to diphthongs and semivowel, vowel boundaries. Attempt to accommodate such boundaries may lead to over segmentation, which is an obvious problem in segment-based representation.

(C) **Landmark-based representation:** In landmark-based representation, speech signal is represented as a sequence of landmarks, where the landmarks are defined as time instants of the speech signal where the acoustic manifestations of the linguistically motivated distinctive features are most salient. Landmarks are foci, so speech processing is done around a landmark rather than between two landmarks, and hence reducing the computations required. Generally, many of the boundaries obtained in segmentation can be considered as
landmarks, but not all boundaries are landmarks and not all landmarks occur at the boundaries.

**Speech Generation**

The methods of conversion of phone sequence to speech waveform could be categorized into – (a) parametric, (b) concatenative and (c) statistical parametric synthesis.

**Parametric Synthesis:** Parameters such as formants, linear prediction coefficients are extracted from the speech signal of each phone unit. These parameters are modified during synthesis time to incorporate co-articulation and prosody of a natural speech signal. The required modifications are specified in terms of rules, which are derived manually from the observations of speech data. These rules include duration, intonation, co-articulation and excitation function. Derivation of rules in parametric synthesis is a time taking task.

**Concatenative Synthesis:** Here the samples of speech units are stored and used during synthesis. The speech units used in concatenative synthesis are typically at diaphone level so that the natural co-articulation is retained. An index of the units in the speech database is then created based on the segmentation and acoustic parameters like the fundamental frequency (pitch), duration, position in the syllable, and neighboring phones.

**Statistical Parametric Synthesis:** It is one of the latest trends in TTS by generating the average of some set of similarly sounding speech segments. This contrasts directly with the desire in unit selection to keep the natural unmodified speech units, but using parametric models offers other benefits. The SPS methods are found to produce intelligible and consistent speech as compared to natural and often inconsistent speech by unit selection techniques. Also, the process of reconstructing speech from parameters is still not ideal. Although modeling the spectral and prosodic features is relatively well-
defined, models of residual/excitation have yet to be fully developed even though composite models such as STRAIGHT are proving to be useful.

4.1 Speech Recognition

Voice/Speech recognition systems can be classified into speaker independent and speaker dependent recognition systems. The former system doesn’t require training and the later system involves signal modeling and pattern recognition. In speaker dependent model, initially the features of the speaker’s voice print are extracted and stored in the database. Then, the features of the test input are extracted. Both are compared to find the degree of matching.

This model comprises of enrollment and verification parts. In enrollment part, after acquiring the voice imprint of an individual, its features are extracted where speech samples are converted into speech feature vectors. Speaker model characterizes a speaker from his/her speech using feature vector using models like GMM (Gaussian Mixture Model). The results of which are stored in the model database. In verification part, the same procedure is followed and features are extracted. Score normalization is the crucial component which scales the matching score against the scores given by database, to make a calibrated decision. Finally, the resulting decision is compared with the threshold in the decision making block. The Figures 14,16,19 in the annex. exemplify the procedure of recognition.

Simulation for speech recognition can be done using Matlab tool. The decision of recognition system is uncertain, as the voice print is severely distorted by environmental disturbances and presence of silence regions. Data Bank of Vocabulary of Words or Query List, which may be generic or specific to application domain such as Banking and medicine, is used in Voice Services. Speech may be continuous or discrete depending upon whether there is pause between successive words spoken by a person or not.
Automatic Speech Recognition (ASR) Technology helps in precise understanding of voice commands with automatic and instantaneous transcription of one’s speech and their execution. Speech engine has a set of algorithms that transcribe sound into keystrokes or into text that is correctly spelt.

**Speech to Text (STT) Conversion**

1. When someone speaks then there is vibration in air (wave) and audio signals are produced. Any natural signal generated in the real world is an analog signal. A mike or microphone in the mobile phone captures these audio waves continuously and converts these vibrations(speech) into digital form (text). Speech to text conversion is a process of analog to digital conversion. Our speech is in analog form and it is converted into digital form (text) by the sampling of sound or audio signal. To convert our speech into text form, a computer or mobile phone executes many complex steps and transformations.

2. **Filter:** A filter is used to reduce noise and interference in a desired signal.

3. **Analog to Digital Converter (ADC):** The analog signal is converted to digital signal by feeding it to ADC.

4. **Digital to Analog Converter (DAC):** The digital signal is processed by implementing specific algorithms. The digital signal is again converted back to analog signal by feeding it into DAC. The DAC connect the dots in digital system by performing interpolation. Other approaches include zero order hold or stair case approximation.

5. **Sampling:** It is done by using a switch to convert analog signal to digital signal. The time interval \( T \) between successive samples is called sampling interval. Its reciprocal is called sampling frequency. The sampling of an analog signal is based on sampling theorem. According to sampling theorem, sampling is the process of converting a continuous time signal \( x(t) \) into a discrete time signals \( x(n) \) by measuring
the amplitude of the x(t) at integer multiples of a sampling intervals T_s. Sampling Theorem for Low Pass Filter: Sampling theorem states that a band-limited signal of finite energy which has no frequency component higher than W_m rad/sec (or F_m Hz) may be completely recovered from a knowledge of its samples if the sample frequency W_s > 2W_m samples rad/sec. Considering the audio frequency range (20Hz to 20KHz) of the electromagnetic spectrum, the speech signal is sampled at 8 kHz frequency for narrowband and it is sampled at 16kHz for wideband (according to the sampling theorem) as shown in Figure 9 in the annex.

6. **Quantizer**: The value of each sample is represented by the finite value.

7. **Coder**: The finite values are represented by a binary code.

A person who speaks several languages, uses mixed words, abbreviations and tones, which are challenging for machines to handle.

4.2 **Speech Synthesis – Text to Speech (TTS) Conversion**

Text-to-speech synthesis (TTS) is the artificial production of human speech by a computing device called speech synthesizer, which is embedded nowadays in software or hardware products. A TTS system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech. The quality of a speech synthesizer is judged by its similarity to the human voice and by its ability to be understood clearly.

In the real world, the typical input to a text-to-speech (TTS) system is text from console or as available in electronic documents, newspapers, blogs, emails etc. The text available in real world is anything but a sequence of words available in standard dictionary and several non-standard words (NSW) such as numbers, abbreviations, homographs and symbols. The goal of text processing
module is to: (i) process the input text, (ii) normalize the non-standard words, (iii) predict the prosodic pauses, duration and intonation and (iv) generate the appropriate phone sequences for each of the words using standard pronunciation dictionary.

The process of text to speech conversion is performed in two stages: (a) **Front end process** does two major tasks:

(i) **Pre-processing**: Here, the raw text containing symbols like numbers and abbreviations are converted into equivalent written words.

(ii) **Linguistic Representation**: Phonetic transcription is assigned to each word. Then it divides and marks the text into sentences, clauses, etc. Phonetic transcriptions and prosodic unit’s information - both make up the symbolic linguistic representation.

(b) **Back end process** works as synthesizer, which converts the linguistic representation into sound (*Figure 15*).

In order to evaluate the quality of text-to-speech systems, subjective and objective evaluations are used. In subjective evaluation, the synthetic speech is played to the native speakers and their view on the quality of speech is sought. In objective evaluations, the synthetic speech is compared with the natural speech utterances and metrics such as spectral distortion are computed.

### 4.3 Multilingual Speech

Providing multilingual services for an application is generally done by first designing it in one language and then expanding this design for other languages. Sometimes different designs are needed for individual languages or a group of languages. Typically, this process includes typesetting the texts into other languages. Each language’s translation is brought into the design and then adjusted and reformatted in terms of text sizes, line spacing, etc. to ensure that the new fonts are visually appealing, fit into the text boxes and are
readable. While developing voice based services, developers should look into National Language Support (NLS) aspects such as National Character Handling (Character Set, Keyboard Arrangement, Display, Printing and Fronts), Country Support (Currency, Date and Time Format, Data Manipulation Standard), Translation (Coding Standard, Documentation) and Linguistic Support (Jurisdiction and Format).

In a spoken language, one finds same word having different meanings based on context, several different words having similar pronunciation or meaning and a combination of such words in active and passive sentences leads to linguistic uncertainty. It is a big challenge in automation as digital machines can handle information only if it is 100% precise. Machine Intelligence techniques such as fuzzy logic, neural networks and rule-based techniques are useful to resolve such ambiguity and bring required level of precision at the cost of higher computation. Centralized multilingual mobile services may be provided to citizens in India.

5. Voice Authentication

5.1 Authentication: It is a process involving interaction between at least two entities X and Y. “X authenticates Y” means the execution of the process involving the following four phases: (i) enrollment (registration) phase, where entity X collects the profile of Y; (ii) submission (presentation) phase, where entity Y submits its credentials of the profile required by X to get the access of required resource; (iii) verification phase, where entity X verifies whether the presented credentials match with the registered profile of Y or not; (iv) decision phase, where decision to grant permission to allow or deny the access of desired resource is taken based upon the level (full, partial, nil) of confirmation of verification of credentials matching.
The focus of **Voice Authentication or Speaker authentication** is **Speaker Recognition** which is identification of a speaker, based upon voice of the speaker. A human speaker is identified by his/her voice in the form of speech. Speech is useful to authenticate a user so as to provide specific services to legitimate user.Humans use several levels of information to recognize speakers from speech alone, but automatic systems are still dependent on the low-level acoustic information. The challenges in this area are to find, reliably extract, and effectively use these higher levels of information from the speech signal.

Block diagram of Speaker authentication model is given in **Figure 16**. It addresses the following problems:

**5.2 Speaker Verification** – Using utterances from a speaker, determine whether the caller is who he/she claims to be (requires an identity claim).

*Problem:* Given a voiceprint (voice sample or a segment of speech) $y$, determine whether it corresponds to a specific user, whose identity is known. It is verifying through one-to-one matching.

**5.3 Speaker Identification** – Using utterances from a speaker, determine who the caller is, out of a set of known speakers.

*Problem:* Given a voiceprint (voice sample or a segment of speech) $y$, determine who is the user $s$ to whom it corresponds to. Assuming (i) $y$ was spoken by $s$, (ii) $s$ belongs to a Set $S$ of $n > 1$ users, (iii) Set $Y$ of voiceprints contains at least one voice print of each speaker $s$ in $S$. It is verifying through one to many matching, which is repeated verification with every entity in $S$.

**5.4 Speaker Clustering** – Given a Set $S$ of $n$ Users and the Set $Y$ of their voiceprints. Analyze $Y$ based on similarity and form subgroups by partitioning of $S$, such that the voiceprints within any subgroup or cluster are similar. It requires **Speaker De-duplication:** Ensuring that
data set Y of all users S is unique. It means confirming the fact that no two voiceprints in Y of any user or across any two users are identical. Clustering techniques are used to form different groups of voice samples from a population of unknown voice samples based purely on their similarity in characteristics.

5.5 Speaker Classification: Given (i) a Set S of n Users, (ii) the Set Y of their voiceprints, (iii) the number of classes (clusters or subgroups) of S, (iv) each user s belongs to which particular class. If a new voiceprint y* is provided, determine it belongs to which class. Design of classifiers which can do this task automatically and accurately is an important area of voice based research. It involves training and testing in various scenarios and application of deterministic and heuristic optimization techniques. Classification techniques are used to classify the voice samples linked with known and unknown persons. It helps to identify whether a given voice sample file corresponds to a male or female voice, whether it corresponds to an infant, teenager, adult, middle aged or old person, whether the person was in normal, happy, stressed, sorrow state.

5.6 Voice Recording & Recognition

Voice of a user is recorded in the form of speech in two stages (i) Enrollment and (ii) Presentation. For an efficient functioning, the user has to go through a process of enrollment where the user is asked to speak a given text, the recorded voice is then processed and the voice-prints are stored. During the time of presentation, the user has to again speak to record the voice. The recorded voice sample is matched with the stored voiceprints and if the score is above a threshold value, the user is authenticated successfully. There are two different approaches to this task, one is text dependent and the other text independent. The first approach is to have an individual repeat the same text during enrollment as well as presentation. The second approach is to have text independent presentation process, whereas user is not bound to repeat the text displayed. This
approach captures the generic behavior of MFCCs and thus, it has a disadvantage, a user voice might be recorded and replayed during authentication to break into the system.

Various recording standards exist today and different schemes are used in different scenarios. A high quality microphone is useful in enrollment process. Mobile phones support wireless microphones. Digital Telephony uses sample rate of 8kHz and 8 bits per sample thus, the bit rate is 64 kbps. Due to advancement in technology, many computers and even small electronic devices such as smartphones, tablets, etc. use a sampling rate of 16 kHz or even 32 kHz with 16-bit precision, 16 kHz with 16-bit precision being most popular. In professional arena and music industry, sampling rate as high as 96 kHz is used and even bits per sample is as high as 24 bits.

The role of the front-end processing is to extract from the speech signal features that convey speaker-dependent information. In addition, techniques to minimize confounding effects from these features, such as linear filtering or noise reduction are employed in the front-end processing. The output of this stage is typically a sequence of feature vectors representing the test segment, \( X = \{x_1, \ldots, x_T\} \), where \( x_t \) is a feature vector indexed at discrete time \( t \in [1, 2, \ldots, T] \). There is no inherent constraint that features extracted at synchronous time instants be used; as an example, the overall speaking rate of an utterance could be invoked as a feature. These feature vectors are then used to compute the likelihoods. It is assumed that a Gaussian distribution best represents the distribution of feature vectors.

In general, it has been found that to obtain the best performance with this approach requires the use of speaker-specific background speaker sets. This can be a drawback in applications using a large number of hypothesized speakers, each requiring their own background speaker set. The second major approach to alternative hypothesis modeling is to pool speech from several speakers and train a single model. The main advantage of this approach is that a
single speaker-independent model can be trained once for a particular task and then used for all hypothesized speakers in that task. It is also possible to use multiple background models tailored to specific sets of speakers.

The most important part of voice recognition is feature extraction and recognition method. Feature Extraction is done to extract important and key features from the voice. In our case, we have used Mel Frequency Cepstral Coefficients (MFCCs) and the deltas. Various transforms such as Fourier, Laplace, Z-Transform, Henkel and Melian Transforms help in (i) the extraction of spectral properties of signal in a transform domain; (ii) The transform function directly indicates the locations of poles and zeros that help in analyzing the stability of communication system, (iii) compression of information for energy compaction.

5.7 Decision Approach

The single-speaker detection task can be posed as a basic hypothesis test between:

\[ H_0: \text{y is from the hypothesized speaker s} \]
\[ H_1: \text{y is not from the hypothesized speaker s}. \]

The optimum test to decide to accept or reject between these two hypotheses is a likelihood ratio test. The basic goal of a speaker detection system is to determine techniques to compute values for the two likelihoods, \( p(y|H_0) \) and \( p(y|H_1) \).

The recognition method is another important aspect for which Gaussian mixture model seems to be appropriate.

**Gaussian Mixture Model**

This is perhaps one of the most important element in the entire classification process. It is a probabilistic model that is used widely in
many applications as many of the random processes found in nature tend to be Gaussian. A one dimensional Gaussian pdf, parameterized by mean and variance is used. Similarly, a multivariate Gaussian is parameterized by mean and co-variance matrix. So, a Gaussian mixture is linear combination of two or more Gaussians. Since the maximum likelihood does not work here as there is no closed form solution, therefore the parameters are estimated using Expectation Maximization (EM) technique. EM algorithm has two main steps are:

1. Estimation Step: For a given parameter values, we can compute the expected values of the latent variable.
2. Maximization Step: Updates the parameters of our model based on the latent variable calculated using ML method. The steps are repeated iteratively until convergence is achieved.

**Score:** It is the average log-likelihood over a data with n-samples. Higher score represents the more confident decision. Thus, once a Gaussian model is fitted over the MFCCs vectors, it can be then used to compute the posterior probability to authenticate the person.

**Threshold Estimation:** A different approach was taken while deciding the threshold. Setting an arbitrary value would have been inefficient as the model is unique for every person and a constant value would not have taken each person into consideration. Thus, a unique threshold was determined for every individual. From MFCCs, data feature vectors were computed. The entire data was then divided into two parts, 3/4\(^{th}\) of data was reserved for training and 1/4\(^{th}\) for threshold estimation. Gaussian parameter were obtained on training data. Data reserved for threshold estimation was further partitioned into subsets of fixed size of 150 samples. Score for each of subsets was obtained, mean and standard deviation was computed. Threshold was set as a difference of mean and standard deviation. The above observation was made while testing on data set
from CMU sphinx. Error rate was minimum when threshold was set to difference of mean and standard deviation.

Authentication requires a pre-trained model; in our case, we have used Gaussian. The recorded sample is first processed and the feature vectors are extracted. The parameters that are stored locally are then used to compute the posterior probability and log-likelihood over data is computed, the process is very similar to threshold estimation. Entire data is divided into subsets of fixed size of 150 samples. Score over each of the subset is computed. Mean of scores is computed. The computed mean is then compared with the threshold value and the person is authenticated only if computed score is higher than threshold.

Justification for Partitioning of Data Set: For a data set having N samples, the parameters estimated represents the probability density function. This removes the uncertainty of the score diverging to a large negative value, thus concept threshold can be easily employed to reduce the false rejection rate.

6. Experimental Results

At the Centre for Mobile Banking, we have designed and implemented IVRS system for Mobile Banking Awareness. Work towards voice based mobile services for Banking and Mobile Governance with multilingual support is in progress. In the prototype design, we have followed:

- **Recording**: We have used in-built library provided with Linux distribution, `rec`, to record the audio file. The recording was done with a sampling rate of 8000Hz, mono mode and each sample were 8-bit. While recording, it was ensured that the length of recording was sufficient i.e. the number of MFCC samples were sufficiently greater than number of mixtures employed.
For a sampling rate of 8kHz, higher frequencies are missed. Thus, another prototype was also built with a sampling rate of 8kHz and precision up to 16-bits.

- **Random Sentence Generation**: Python script, sentence generator was built to produce a sequence of random sentence for the user, since Mel-frequency based authentication does not depend upon text, therefore the user is not bound to repeat the words produced by the program and can produce any arbitrary speech.

- **Speech Processing**: Essentia, an open source sound processing library was used to perform various transformation like framing, spectrum analysis, energy computation, MFCC computation, etc.

- **Noise Modeling**: Noise was assumed to be from Gaussian distribution and first few seconds of the recording was used to model noise only where the user was asked to stay silent. The same model was used later to classify a frame as to speech or speechless.

- **Making Feature Vector**: After processing of audio signal MFCCs up to 13th coefficients were used and a python script was used to compute deltas. The delta and double deltas were appended to MFCCs vector, thus extending the feature vector to 39-dimensional.

- **Enrollment or Training the Model**: A Gaussian mixture of 128 Gaussians was used. Python library Sci-kit contains optimized implementation of Gaussian mixture model which was used for parameter estimation. Once convergence was achieved the means, weights and co-variance matrices were stored locally to the disk for future use.

- **Testing or Authentication**: During authentication a similar process to enrollment is done for the audio processing. The Gaussian parameters are fetched from the disk and score is computed, if the score is above the threshold value, the user is successfully authenticated.
• **Speech to Text in Android:** The speech to text library in Android API is a package android.speech and specific class is android.speech.RecognizerIntent. We trigger an Intent (android.speech.RecognizerIntent) which shows dialog box to recognize speech input. This activity then converts the speech into text and sends back the result to our calling activity. When we invoke android.speech.RecognizerIntent intent, we must use startActivityForResult() as we must listen back for resultant text.

• **Text to Speech in Android:** Android provides TextToSpeech class for this purpose. In order to use this class, one need to instantiate an object of this class and also specify the initListener. Language can be set by calling setLanguage() method. Its syntax is given below – ttobj.setLanguage(Locale.UK). List of some other useful methods given below can be used:

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>addSpeech(String text, String filename)</td>
<td>Adds a mapping between a string of text and a sound file</td>
</tr>
<tr>
<td>getLanguage()</td>
<td>Returns a Locale instance describing the language</td>
</tr>
<tr>
<td>isSpeaking()</td>
<td>Checks whether the TextToSpeech engine is busy speaking</td>
</tr>
<tr>
<td>setPitch(float pitch)</td>
<td>Sets the speech pitch for the TextToSpeech engine</td>
</tr>
<tr>
<td>setSpeechRate(float speechRate)</td>
<td>Sets the speech rate</td>
</tr>
<tr>
<td>shutdown()</td>
<td>Releases the resources used by the TextToSpeech engine</td>
</tr>
<tr>
<td>stop()</td>
<td>Stop the speak</td>
</tr>
</tbody>
</table>
• A mobile phone provides voice interface to the user. A mobile application acts as a voice assistant to get users voice inputs through microphone. It forwards the inputs for necessary processing and presents the received response to user in visual form on screen or audio form through speaker or both as the user wishes. The voice assistant uses various Application Programming Interfaces (API). To recognize the verbal inputs, it uses Speech-to-Text Conversion API, for example, Google Speech to Text API and SIRI. To synthesize processed information into speech output, it uses Text to Speech Conversion API, for example and Google Text to Speech API *(Figures17 & 20)*.

7. Conclusion

Voice would dominate natural conversation and would become primary interface for multilingual communication and mobile services in future except when silent communication through text and gestures are desired. Real-time voice authentication as a primary factor may emerge in future if the accuracy and response time are made within human acoustic tolerance limits. Till then, it can be used as an additional factor for user authentication. Mimicry of a speaker’s voice by another person although is not an easy task, and can be misused to spoof a person. Here, we have presented the elements of voice communication, speech analysis, speaker authentication, multilingual speech services and some challenges. For deeper understanding and implementation, one may refer to the standard books and recent research works for signal processing, transforms, speech recognition, multilingual processing, etc. The proposed architecture for mobile governance and mobile payment services in *Figure 18*, mobile application architecture in *Figure 17* and voice application architecture in *Figure 20* may prove to be useful to the developers and implementers.
Acknowledgement: I thank my students for their interest to work in the Voice Services area at Centre for Mobile Banking (CMB) and to Prof. B. Yegnanarayana for his insightful thoughts during his recent talk at CMB. I also thank various other researchers whose works reported in the published literature helped me to prepare this article.

References

1. Tele banking Services: Voice Based Authentication, Rahul Kumar (IIT, Guwahati), Internship Project Report-IDRBT, Guide: Dr.V.N.Sastry, July 2017
2. Voice Based Mobile Banking services in Hindi and English, Deepa Yadav (NIT, Silcher), Internship Project Report-IDRBT, Guide : Dr.V.N.Sastry, July 2017
6. Some Experiments on Speech Processing using Matlab, V.Sudheer, Research Associate, CMB, Guide: Dr.V.N.Sastry, 2014
7. Customer Education on Mobile Banking (Using IVRS & USSD), Purushottam Vishnu Bhandare (Pondichery University), Internship Project Report-IDRBT, Guide: Dr.V.N.Sastry, 2012
Annexures

Figure 1: Natural Organs of Voice Production and Reception

<table>
<thead>
<tr>
<th>Ear: Natural Architecture of Voice Reception Mechanism (Destination)</th>
<th>Mouth: Natural Architecture of Voice Production Mechanism (Source)</th>
</tr>
</thead>
</table>

Figure 2: Automated Voice Communication through Signal Processing Devices
Figure 3: Basic Units of Voice Transmission (Source, Transmitter, Channel, Receiver, Destination)

![Basic Units of Voice Transmission Diagram]

Figure 4: Periodic Analog Signal (Sinusoidal)

![Periodic Analog Signal Diagram]

Figure 5: Voice Signal Representation in 2D (x-axis: Time (ms), y-axis: Amplitude (mm))

![Voice Signal Representation Diagram]
Figure 6: Analog Speech Signal in Continuous Time Domain \( (t, f(t)) \) in 2D

![Analog plot of the utterance](image)

Figure 7: Speech Signal Representation in Discrete Time Domain \( (t_k, f(t_k)) \) in 2D

![DT plot of the utterance](image)
Figure 8: Fourier Transform Representation of Voice Signal

Sampling Effects: Frequency Domain

Figure 9: Analog to Digital Conversion (ADC)
Figure 10: Modulation (Amplitude and Frequency Modulation)

Figure 11: Digital Modulated Signal
Figure 12: Mel Frequency Spectrum

![Mel Filter Bank]

Figure 13: Multivariate Gaussian Model

![Multivariate Normal Distribution]
Figure 14: Block Diagram Representing the Computation of MFCCs

Figure 15: Speech Synthesis: Process from Text to Speech (TTS)
Figure 16: Speaker Authentication Model

Figure 17: Voice Command based Mobile Application Service Workflow Diagram
Figure 18: Proposed Architecture of Integrated Voice based Mobile Services

Figure 19: Automatic Speech Recognition System Components
Figure 20: Voice Application Architecture
Biometric Template Protection for Banking

- Dr. M. V. N. K. Prasad, Associate Professor, IDRBT
1. Introduction

The digital needs of the world are increasing day-by-day. Therefore, managing numerous passwords for different applications is a concern. Hence, financial institutions inspected acceptable biometric alternatives for authenticating different users. Biometrics balance both security and simplicity. In the recent years, we have seen increased use of biometric authentication in banking applications. Biometrics has brought a paradigm shift in the way banking provides authentication. Financial institutions invest more in security while minimizing their own financial loss – for e.g., when a credit card data is stolen, other personal data of the customer can be compromised which results in identity theft. If the biometric data is compromised, the identity theft is unresolvable. Therefore, banks have to design technologies to secure biometrics. If the biometric data of a person is exposed in various applications, the privacy concern arises due to storage and misuse of biometric data. Conventional authentication systems like token or password could not meet these security requirements because they can be lost, forgotten or misplaced. Hence, recent years have seen an epidemic growth in identity theft [1].

Biometrics can be used for automated person recognition. An individual can be recognized using her anatomical and/or behavioral characteristics. Anatomical biometric traits include fingerprint, face, iris, palmprint, hand geometry, hand vein, finger vein, voice, etc. Behavioral biometric characteristics include keystroke dynamics, gait, etc. [2, 3]. The techniques to system biometric trends are improved in the recent years. Hence, biometrics is used for authenticating users in programs that include laptop login or building access. The traditional authentication systems based on passwords (something you realize) or tokens e.g. Smartcards (something you’ve got) are inconvenient because the passwords can be forgotten or guessed, and the tokens may be lost or stolen. Biometrics are being utilized in various programs spanning forensic, governmental,
industrial, and safety sectors. The authorities of India capture biometric traits (face, fingerprints and iris) from the populace i.e. Aadhaar Card. The biometric characteristics of a person are obtained by using sensors. The first time a user uses the biometric recognition system is called enrollment [3]. During enrollment, the features are extracted from the biometric image and stored as templates. During verification/authentication, the features are extracted from the query biometric image, the enrolled template and query template are compared which yields in an acceptance or rejection of an individual. If the stored templates are attacked by the adversary, he may change and use them to act as a genuine user. Hence, template protection for biometrics is needed to secure biometric templates from many threats.

2. Biometric Recognition System

The block diagram of a biometric recognition system is shown in Figure 1. The sensors are used to obtain the biometric characteristics [4]. While obtaining the data of a biometric, there may be a possibility of inclusion of noise or unwanted background information or foreign objects, etc. Hence, preprocessing is needed to remove them. Filters are used to remove the noise. Segmentation is used to remove the background information and foreign objects [5].

![Figure 1: Block Diagram of Biometric Recognition System](image)
The feature extractor module is an important step in the biometric recognition system. The type of features to be used may vary depending on the application [6]. During enrollment stage, the feature extractor and template generator run simultaneously to extract features from biometric data and store them as templates in the database [7]. During Identification/Verification stage, the query biometric is processed in the same way. Now, the query template is compared with the enrolled template that yields in a match (accept) or non-match (reject) [4].

Figure 2 shows the attacks may be on user interfaces, on modules, on channels between modulus, and on template databases.

2.1. Attacks on User Interfaces

The attack at the user interface is due to the presentation of a fake biometric trait [1]. In this mode of attack, a possible reproduction of the biometric being used will be presented to the sensor. Examples include a face mask, fake finger, or a copy of a signature. If the sensor does not differentiate the spoof and genuine biometric traits, the adversary enters into the system under a false identity [9].
2.2. Attacks on Modules

The attacks on the sensor module can be a coercive attack, spoofing attack or device substitution or denial of service attacks [10]. Coercive attack refers to presenting a true biometric to the sensor in an unauthorized manner, for e.g. the adversary forces the genuine user to grant him access to the particular system [11]. Spoofing attack refers to copying the biometric of the legitimate user and transfers it to the adversary, thereby fooling the system. A sensor device may be simulated and replaced with the genuine capture device [12]. The attacks on feature extractor module are forcing it to produce pre-selected features of the adversaries. Matchers are attacked to override the match scores and produce fake scores [13].

2.3. Attacks on Channels between Modules

The adversary may intercept the channels between the modules like, the channel between sensor and feature extractor, feature extractor and matcher, template database and matcher, matcher and application device [14]. Possible attacks here may be replay attack, the synthesized feature vector, overriding the decision [15]. Replay attack allows an old recorded signal to be replayed into the system bypassing the sensor [8]. A synthesized feature vector may be produced after the features have been extracted from the input signals. If the final result can be overridden with the choice of result from the hacker, the final outcome is very dangerous [10].

2.4. Attacks on Template Databases

The templates in the database are attacked by the adversaries, reading them, modifying and replacing the templates, changing the IDs and biometric associated with it. These types of attacks are considered to be the most threatening attacks [10]. The attacker tries to modify one or more templates in the database
which could result in authorization for a fraudulent individual, or at least denial of service for the person associated with the corrupted template [8].

In this study, we discussed various biometric template protection approaches that can be adopted to the Indian Banking Sector. These approaches are classified as cancelable biometrics, biometric cryptosystems, hybrid methods, and Homomorphic Encryption based methods [16]. A valid template protection approach should satisfy the following four requirements [1]:

- **Diversity**: The secured templates must not be cross-matched. This ensures user’s privacy
- **Revocability**: From same biometric data, the protection method should be able to revoke a compromised template
- **Irreversibility**: The original biometric cannot be obtained by secured (or transformed) template. The computational cost should be infeasible for obtaining the original biometric data, while it should be easy to generate the secured template
- **Performance**: The template protection technique developed should not degrade the accuracy of the recognition system.

3. **Categories of Biometric Template Protection Schemes**

The major schemes of “Biometric Template Protection” can be categorized into cancelable biometrics, biometric cryptosystems, hybrid methods, and homomorphic encryption-based methods as shown in Figure 3. Further, the cancelable biometric systems are divided into salting methods and non-invertible transforms, and biometric cryptosystems are divided into the key binding and key generating systems. Each of these schemes is discussed in detail in the following sections.
4. Cancelable Biometrics

In a cancelable biometric system, a transformation function is used to transform the template extracted from the biometric images. Now, the transformed template is stored in the database. During authentication, the query image is applied with the same transformation function used during enrollment. The matching is conducted in the transformed domain. Figure 4 shows the operation of a cancelable biometric system. A cancelable biometric system can be implemented using salting and non-invertible transforms.
Figure 3: A Hierarchical Taxonomy of Biometric Template Protection Schemes
4.1. Salting

In salting, biometric features are transformed using an invertible function. These systems use a key for transformation. Hence, the key should be secured or recalled by the user for authentication. A compromised template can be easily revoked by changing user-specific keys.

4.2. Non-invertible Transforms

The non-invertible transforms apply a one-way transformation. The parameters of the transformation function should be presented in the authentication stages. Since the transformation function is non-invertible, the raw biometric image cannot be recovered, thereby increasing security [17]. In designing a non-invertible transform, the trade-off between discriminability and non-invertibility should be maintained.
5. Biometric Cryptosystems

A biometric cryptosystem binds a key to the biometric features or generates a key from the biometric features [18]. Helper data is used to bind/generate keys [1] from the biometric data. The biometric cryptosystems are classified into key binding and key generation systems based on how the helper data is derived.

5.1. Key Binding Biometric Cryptosystems

In a key binding cryptosystem, a user-specific chosen key is bound to a biometric template to obtain helper data. The combination of key and the biometric template bound is stored as a secure template, called as helper data. The keys are obtained from helper data by a decoding attempt [12]. The framework of a key binding cryptosystem is shown in Figure 5.
5.1.1. Fuzzy Commitment Schemes

A fuzzy commitment scheme combines cryptography and error correcting codes. During enrollment, a random key is chosen and is encoded using error correcting codes. From this, a code-word ‘c’ is generated. Now, the biometric feature vector is XORd with this codeword and stored as helper data. The hash value of ‘c’ (say hash(c)) is stored along with the helper data. During authentication, the feature vector of the query image is presented to the decoder and is XORd with the stored helper data to get the codeword c'. If hash(c) = hash(c'), then c = c', so the user is accepted else the user is rejected. Figure 6 shows the framework of Fuzzy Commitment Scheme.

![Figure 6: Fuzzy Commitment Scheme](image)

5.1.2. Fuzzy Vault Schemes

A Fuzzy Vault Schemes is shown in Figure 7 built by Juels and Sudan [19]. A key k is locked by an unordered set A, which results in a vault V_A. During enrollment, A is projected onto a polynomial p. During authentication, if another set B overlaps A, key k is reconstructed.
5.2. Key Generating Biometric Cryptosystems

In key generating biometric cryptosystems, the keys are generated from biometric templates as shown in Figure 8. Quantization scheme is an example of key generating biometric cryptosystem. The helper data is quantized to obtain intervals of the feature elements using keys. During authentication, biometric features are calculated and mapped to the determined intervals.

6. Hybrid Methods

Hybrid methods are developed by combining Cancelable Biometrics and Cryptosystems. For e.g., a method that secures a cancelable template using a biometric cryptosystem may have the advantages of both cancelable biometrics (which provides high diversity and revocability) and biometric cryptosystem (which provides high security) approaches.
7. Homomorphic Encryption

Homomorphic Encryption was embodied into biometrics by Ye et al. [20]. These schemes allow a limited subset of computation on the encrypted biometric data. Combining Homomorphic Encryption with biometric recognition systems would give better performance without degrading the accuracy.

8. Information Fusion in Biometrics

An individual can be recognized by presenting multiple biometric characteristics, both behavioral and physical. Such biometric systems are called as multi-biometric systems. Multi-biometric systems are more accurate because multiple pieces of evidence can be presented to the system [21].

Fusion uses multiple inputs or methods of processing of biometric samples [2]. Fusion improves accuracy, efficiency, applicability and robustness of the biometric systems. Biometric fusion can be done by using a single trait or multiple traits as shown in Figure 9. Multi-
sensor, Multi-sample, Multi-instance, Multi-algorithm systems use single trait for fusing biometric data [22]. Multi-modal systems use multiple biometric traits of a person for fusion.

In multi-sensor systems, a single biometric trait is sensed through multiple sensors. In multi-sample systems, multiple samples are collected by using the same sensor at different times. Multi-algorithmic systems use multiple feature sets extracted from the same biometric data or multiple matching schemes operating on a single feature set. Multi-instance systems use multiple instances of same biometric data. Multi-modal systems combine different biometric traits for establishing identity [16]. Fusion can be done at the sensor level, feature level, decision level, and score level [23] as shown in Figure 10. Sensor level and feature level fusion are considered as fusion prior to matching. Decision level and score level fusion are considered as fusion after matching [24].

![Figure 9: Multiple Sources of Evidence](image-url)
9. Conclusion

The popularity of biometrics in the financial sector is increasing worldwide. With the widespread use of biometric technology, users need not rely on passwords or carry a token. In the recent years, financial institutions are investing in biometric innovation. There may be many obstacles using biometric technology such as compromised templates. Hence, biometric template protection techniques need to be developed to play a key role in banking and business in the near future.

References


Biometrics and Its Impact in India

– Dr. S. Ananth,
  Adjunct Faculty, IDRBT
1. Introduction

Biometrics and biometric authentication has gained a centrality that was unimaginable at the turn of this decade. Any discussion of the use, consequences and impact of biometrics in India necessitates an elaborate discussion on Aadhaar. The importance of biometrics has increased with the growing centrality of Aadhaar. The rapid pace at which it is transforming governance, banking and segments of the economy is astounding. The adoption of biometrics in India has been mired in controversy with the matter finally heading to the courts with vociferous arguments put forth by proponents and opponents. The increased adaption of biometrics in a country like India has forcefully drawn attention to larger questions about the role and impact of technology on society, economy and governance.

This paper attempts to trace the adoption of biometrics in India by the banks and the Unique Identification Authority of India (UIDAI). It is divided into two sections. The first section provides a broad overview and survey of opinions on Aadhaar as well as its role in the last mile access in connection with the banking system and financial inclusion. The second part attempts to understand the cost-benefits of Aadhaar.

The UIDAI was established through a notification issued by the Planning Commission on 28 January 2009 in pursuance of a decision to this affect by the Empowered Group of Ministers in November 2008. It was constituted with an initial core team of 115 officials and staff under the Planning Commission. The UIDAI envisions the provision of a ‘unique identity and a digital platform to authenticate anytime, anywhere’. Aadhaar finally gained legal sanctity with the passage of the bill formally known as Aadhaar (Targeted Delivery of Financial and Other Subsidies, Benefits and Services) Bill 2016. It

promises to be a cornerstone that will change the administrative and economic landscape of India in the next decade.

At the outset, it is important to note that Aadhaar has been a stupendous success only because of efforts of the Union and State Governments and institutions like the Reserve Bank of India. As on 30 June 2017\(^3\), 115.79 crore or a little more than 86.6 per cent\(^4\) of Indians have been provided with the Aadhaar number. The numbers surrounding Aadhaar are staggering by any standards and its importance increases when we consider the fact that 93% of adults have an Aadhaar number while 25.48 crore bank accounts, 9.5 crore Jan Dhan Accounts, 12.28 crore LPG consumers, 11.39 ration card holders and 5.9 crores MGNREGS card holders have their Aadhaar numbers seeded (linked). Invariably, it is pertinent to note that any enrolment or programme would have had a similar success if it had received such unstinted support ever since its inception by successive governments irrespective of political affiliations. The sheer size of the enrolment makes Aadhaar the most extensive in the series of technology programmes that various State and Central government have taken up over a period of time.

An overview of the nature of Aadhaar indicates the diverse opinion about it. This diversity of opinion exists not just about the technology platform but also its functionality and importance that is given to it. In a perceptive analysis, Ashish Rajyadhiksha’s observation that ‘unlike purpose-driven identity definitions such as social security numbers, voting registrations or driving licenses, Aadhaar is in itself purposeless, an empty vessel: purpose would be needed to be added to it by the services that need it and it would then define its own purpose to suit theirs’, is largely responsible for its extensive reach. Since it can be used for almost anything, ‘it has become something of a chimera, attributed with whatever downstream purpose to

\(^3\) Latest official figures are as on 01 July 2017 on the UIDAI homepage.
\(^4\) https://uidai.gov.in/images/state_wise_aadhaar_saturation_as_on_22052017.pdf
which it will be used\(^5\). He opines that UIDAI has willingly taken on, having assiduously advocated and marketed the varied uses to which both the State and the market intends to put it, but it is important to note that these are autonomous domains of functioning, often defining policies and procedures that are not common to each other\(^6\). This unique flexibility does not exist with any other programmes and may be responsible for the increased adoption of Aadhaar by the government agencies and private businesses. In other words, the focus of Aadhaar is to add value and become the means to an end for a particular activity rather than limit itself.

2. **Contextualising Technology and Governance in India**

The UIDAI is not the first agency to have collected biometrics. Aadhaar needs to be contextualised in the increased tendency of policy makers to use technology as the panacea for the ills facing India. This tendency was clearly discernible and became pronounced since the turn of the century. The increased recourse to technology has come at a time when there is increased pressure and contest among stakeholders and the increased need to balance contesting demand for scarce resources, including benefits from the State. The underlying assumption being that technology can overcome the existing obstacles to better delivery of services, especially the problem of weeding out fake beneficiaries who were believed to be the primary cause of problems related to delivery of government services. This use of technology required greater investments in various technologies which led to a problem of multiplicity of databases.


The case of undivided Andhra Pradesh is illustrative of this trend. Technology adoption in Andhra Pradesh is interesting for a number of reasons. The first, it was one of the first states to embark on a huge technology intensive ‘modernisation’ of government administrative apparatus. Second, these technology-intensive efforts led to the government encouraging the creation of a number of databases, through which it has exhaustive information at its disposal and mapping of almost all parts of the state under various schemes. It included even an early attempt in 2002 to generate and provide citizens with a unique number, which was disrupted with the onset of elections in 2004.

After the 2004 elections, there were attempts to collect biometrics, including retina scans for the ration cards, but it faced issues related to de-duplication and was subsequently shelved. A cursory glance at some of the databases in place in undivided Andhra Pradesh indicates that a large part of the identity of the citizens is digitized and accessible to various institutions of the State in varied forms. The emphasis was more on creating the technology infrastructure and the database of users with little focus on making the databases interoperable by various other government programmes even when needed. The greater the number of welfare services accessed, the greater the amount of information at disposal of the government, although it was not de-duplicated and was not available as a single centralised database.

The ration cards are often the connecting thread to various databases. A number of these databases, like Aryogyasri and Community Managed Sustainable Agriculture (CMSA) have a reputation of efficiency. But, the practical issue that affects all these databases is that they are not interoperable. Hence, the attraction of administrators is that Aadhaar would fill the gap thereby leading to interoperability of databases. The table below offers an overview of the particulars of the database and the number of citizens mapped in these databases.
## Already Mapped: Government Databases in AP before 2014

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Particulars of Database</th>
<th>Citizens/Households Mapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Self Help Groups</td>
<td>1.14 crore Individuals</td>
</tr>
<tr>
<td>2.</td>
<td>NREGA</td>
<td>1.2 crore job cards</td>
</tr>
<tr>
<td>3.</td>
<td>Pensions</td>
<td>68 lakh</td>
</tr>
<tr>
<td>4.</td>
<td>MFI Borrowers</td>
<td>All the 1.1 crore borrowers</td>
</tr>
<tr>
<td>5.</td>
<td>Community Based Insurance(^7)</td>
<td>90 lakh</td>
</tr>
<tr>
<td>6.</td>
<td>Student scholarships</td>
<td>Nearly 4 lakh</td>
</tr>
<tr>
<td>7.</td>
<td>Health Insurance (Aryogasri)</td>
<td>52 lakh patients</td>
</tr>
<tr>
<td>8.</td>
<td>Co-contributory Pension Scheme</td>
<td>48.93 lakh</td>
</tr>
<tr>
<td>9.</td>
<td>Disability(^8)</td>
<td>9.75 lakh</td>
</tr>
<tr>
<td>10.</td>
<td>Community Managed Sustainable Agriculture</td>
<td>11.89 lakh small farmers</td>
</tr>
<tr>
<td>11.</td>
<td>Employment Guarantee and Marketing Mission</td>
<td>4 lakh</td>
</tr>
</tbody>
</table>

*Source: Compiled from AP Government Reports and Website 2013-14*

Interoperability is only one of the issues that policymakers face. Fictitious or non-existent beneficiaries accessing the benefit sometimes through impersonation is another reason why biometrics is increasingly considered to be a solution for authentication. There are a number of areas where the use of technology has helped expand governance and increased penetration of the economy.

The promise of Aadhaar is that it is supposed to facilitate anywhere access to the applications by the users once biometric authentication is completed. Further, the short time of a few seconds is supposed to make this access quicker. One such place, if it lives up to the promise is in the sphere of access to rations. If this becomes possible, migrants would find it easy to access their rations not just in their

---

\(^7\) Including *Aam Admi Bhima Yojana* and *Janashree Bhima Yojana*

\(^8\) A specialized technology platform has been created by TCS for the Department of Rural Development, called *Software for the Assessment of Disability for Access Rehabilitation and Empowerment (SADAREM)*
villages where they are registered with the civil supplies department but from anywhere within the state – assuming that the civil supplies departments can seamlessly handle the logistics. It is this promise that states like Andhra Pradesh are hoping to implement through the “ePoS” scheme. Under this programme, a below poverty line beneficiary can access their rations from any shop in the village. This has been facilitated by the state government not only linking the civil supplies database with Aadhaar but also providing fingerprinting and iris scan hardware to dealers. Further, the biometrics of all the family members has been mapped and the state offers rations to any member of the family if the biometrics of any one finger matches those in the database. Though biometrics is the pivot, its success is dependent on other factors including the ability of the “fair price shop dealers” to manage their logistics.

Another area where consumers can benefit includes those accessing banking services in the last mile. This in turn has helped an increasingly mobile population across the length and breadth of the country. One segment that has benefitted from this spread of technology includes those who migrate for their livelihoods or those who travel for their business purpose. In the case of the banking sector, it is common for those travelling as part of their work to deposit money in the non-home bank branches.

The use of biometrics for authentication is seen as a solution not just in India but in other countries too. A recent report quoted China as a country that is increasingly adopting biometrics and facial recognition for not only government programmes but also by the corporate sector. The Chinese government has announced plans to install iris scanners at security check posts in troubled regions concurrent to data mining various social media platforms. They aim to create a “national social credit system” by 2020, in which the
government will assign every citizen a rating based on how they behave at work, in public venues and in their financial dealings. In India, biometrics is under various stages of implementation for airport check-in. Gradually, Aadhaar is being made mandatory for various family welfare programmes. Hence, in the next few years it is likely to be an integral, pivotal component for the delivery of various government and other welfare programmes. Aadhaar is increasingly seen as the solution for its de-duplication prowess and its ability to quickly authenticate large number of requests – technology that is difficult for others to match. There are an estimated two crore Aadhaar authentications per day.

Aadhaar is now central to public policy and an increasing number of programmes are linked to it. The increased use of Direct Benefits Transfer (DBT) as well as the distribution of rations by the state government civil supplies department is dependent on the use of biometric authentication. As the scope of using biometric authentication increases, it becomes more ubiquitous and seemingly indispensable. The scope of Aadhaar has now extended to include in its scope various payments linked through Aadhaar Enabled Payment System (AEPS). UIDAI is structured in a manner that is completely different from anything that India has seen. Two salient features of UIDAI is that it (a) would like to encourage innovation and provide a platform for public and private agencies to develop Aadhaar linked applications and (b) make it attractive for global expertise to collaborate and provide insights to UIDAI organisation. These have

served to accelerate the fears and suspicions about Aadhaar. The last on Aadhaar is yet to be said and in all probability will be decided by the Supreme Court of India.

In the last few years, Aadhaar has been linked to almost all the existing identification cards issued by various government agencies at the Central and State level. The first was the linking of Aadhaar with the civil supplies data as part of the initial enrollment. In most of the states, Aadhaar enrollment was pre-populated with the state civil supplies department’s database during the enrollment drive. The scope of Aadhaar has gradually increased with most government agencies mandatorily requiring its submission. These include registering for examination, filing of police cases, linking car and diving license, existing mobiles, purchasing new SIMs, etc. The most recent measure is the compulsory requirement for Goods and Service Tax and Permanent Account Number (PAN) registration. The Finance Bill, 2017-18 introduced an amendment in the Income Tax Act, sub-section (2) of Section 139AA, which makes it compulsory to link the PAN with the Aadhaar by July 1 2017.

The amended rules require all applicants for new PAN number submission of Aadhaar and the quoting of Aadhaar number for filing income tax returns. Failing this linkage, the PAN will continue to remain valid but tax payers cannot file their returns electronically. Till date, about 7.36 crore of the 30 crore PAN holders have completed this process. Sometimes, the Aadhaar linkage can border on the peculiar. Skype announced that Aadhaar has been integrated functionally and can be used for verification of users and

---

their identity online which is expected to help users communicating with others in a more secure manner and environment\textsuperscript{14}.

3. Biometrics and the Last Mile Banking Access

The UIDAI was not the only agency or the first to collect biometrics. Banks have done the same, often concurrent to the UIDAI’s efforts. The use of biometrics in the banking sector is that it rests on the fond hope that biometrics can help solve the problem of last mile access in the banking sector. It is with this hope and to overcome the problem of lack of education among rural citizens that biometric authentication is an integral part of financial inclusion. Financial Inclusion was introduced in 2005 by the RBI with the intent of expanding banking services in the rural areas. The banks are allowed to appoint banking correspondents as intermediaries who will facilitate the expansion of banking services. These banking correspondents are allowed to offer basic banking services in unbanked and/or under-banked areas. The banks allow the operation of these basic banking services using biometrics to overcome issues related to education and awareness in accessing the bank accounts. Gradually, biometric based Know Your Customer (KYC) norms have been introduced.

Thanks to this programme, banks were among the early institutions to collect biometrics for their financial inclusion customers. The fond hope is that digital technologies of various hue led by biometrics will enable disaggregation of the value, open the possibility of creating new platforms and application programming interface (APIs), facilitate client identification and credit risk assessment while enabling meeting due diligence requirements for customers, especially those with insufficient traditional forms of identification with better data collection and analytics leading to customer

\textsuperscript{14} http://www.financialexpress.com/industry/technology/microsoft-aadhaar-card-linking-skype-lite-needs-to-be-integrated-now-heres-how-to-do-it/750625/ (Website last visited on 05 July 2017)
segmentation and customisation of services\textsuperscript{15}. All these are fond hopes that have till date not materialised at the field level, especially in the case of financial inclusion.

The emphasis on collection as well as dependence on biometrics is a major departure from the past. M. S. Sriram traces the increasing role of biometric authentication and points out that biometrics and Aadhaar have led to a fundamental change in the manner in which the state approaches and dispenses benefits, subsidies as well as access to credit. He observes that the design of the welfare programmes have undergone a change from the past. In the past, though classification of the beneficiary was important, issues of identity or authentication of identity was not as critical to the design of the welfare scheme. Each was a standalone intervention. This has changed with Aadhaar, which now lays importance on “convergence”. Importantly, he points out that state intervention to expand financial inclusion in rural areas has ‘moved away from building state-led institutions to setting policy that pushed financial inclusion through non-state agencies as well’ and by allowing Aadhaar project to set the discourse on financial inclusion, the state has moved away from the RBI’s definition of comprehensive and meaningful financial inclusion towards a model that facilitates transactional aspects, with significant costs added at the intermediary levels\textsuperscript{16}.

This fundamentally changed approach leads us to the most important question of does biometric usage solve the banking sectors last mile problems, especially those related to access to the banking system by the poor? Facilitating the expansion of banking access using financial inclusion is one of the early stated goals of Aadhaar and was stressed since its inception. Considering this focus, one would have expected that the banking sector’s woes in the last


mile would have been overcome with the use of biometrics. Unfortunately, field visits suggest that the biometric solution has been long on promise and short on delivery.

The changed emphasis has the potential to alter the manner in which information is stored and accessed. The importance of these measures has to be seen in the context of Reserve Bank of India’s intention to make the bank account portable across banks and places. While bank account portability is unlikely to become a reality in the near term, financial inclusion efforts seem to be creating the necessary framework for this account portability along with a concurrent system based on identification and biometrics. The importance of these efforts need to be seen in the contextualized nature of the present banking system, which already has in place a system of information sharing on borrowers. This would mean that the accounts opened under financial inclusion, especially those opened by Business Correspondents (BCs) have the advantage of portability. Since the biometrics, credit and personal histories are available, a host of opportunities are likely to open up – faster than envisaged by Aadhaar and without raising questions like reliability of the process of verification, etc.

In other words, the success of the BCs of banks has the potential to not only to redefine the manner in which business services (especially banking services) are delivered, but also the manner in which the government delivers services to its citizens. The speed with which these accounts are opened (five days) and a more reliable process of identification means that after these accounts are opened, banks and a host of government departments may not actually need to be built around the UID. This is because the banking system’s Know Your Customer (KYC) norms are more exhaustive and reliable. Biometric authentication is an important precondition to access these accounts. Thus, these bank accounts have the potential to become the unique number on which a host of other services can be provided and accessed. The theoretical importance of these
banking innovations stems from their ability to provide targeted services more effectively since the required financial information may at least partly be available.

The use of biometrics is changing the banking system, though not necessarily for the better from a financial inclusion customers’ perspective. The growing emphasis and importance of Aadhaar as a compulsory KYC norm has led to it emerging as obstacle to banking access. This is especially visible in rural areas. The case of Kurnool district in Andhra Pradesh is illustrative of these issues that the poor face while accessing their bank accounts. Bankers and Business Correspondents point out that customers who have opened their bank accounts between 2010 -13 are facing problems in accessing their bank accounts because the banks have blocked access to those who have not submitted their Aadhaar numbers to the bank. This problem also encompasses those who have not been able to seed their Aadhaar to the bank accounts – either due to delay from their side or the delay from the branches’ side. This is especially important since one of the positive contribution of the effort to expand financial inclusion is the huge increase in banking access in the form of increased number of people who have opened their bank accounts.

Blocking the accounts due to non-seeding/submission of Aadhaar means that customers cannot access the funds in their accounts. This is aggravated as the customers may have submitted this account for various direct benefits transfers. Blocking of the account or non-submission of Aadhaar has even led to closure of accounts thereby resulting in return of money paid through DBT. The alternative for the customer is a circuitous route of opening another bank account in a different bank and run from pillar to post in each government office to change their bank account – a difficult alternative for customers with low levels of education and attendant ability to navigate the administrative maze.
In Kurnool Town, a customer at a BC outlet complained that they had saved nearly Rs. 8000 in their bank, mostly DBT monies transferred to pay for a recurring deposit. This money cannot be withdrawn though the recurring deposit had matured due to Aadhaar seeding and KYC compliance issues. The net result of these problems in rural banking is that they are either forcing customers to spend inordinate time visiting the branch or simply withdraw from the formal banking system. Problems with withdrawal of money at a time when it is needed the most either due to non-working channels or problems with KYC compliance has convinced people that the best place for their money is their pocket or at home under the mattress. Business correspondents attest to this trend and point out that increasingly customers are not keen on depositing money with the banks.

The collection and access to banking services has not led to any improvement in services. That is ironic considering that the promise of biometrics is that it guarantees the authenticity of the user and hence access from any point is not or should not be an issue. However, in reality, in most instances, the banks do not allow financial inclusion account holders to access their bank account even from the link branch. A few banks, usually a small minority, collected the biometrics and issued smart cards that allow customer to access the accounts from the business correspondents or the link branch.

Surprisingly, unlike the non-financial inclusion accounts, banks do not allow access to the accounts from non-home branches. In contrast, a non-financial inclusion account holder can access their account from any non-home branch of the bank across the country. This refusal to provide access often does not justify the reason for banks’ behaviour. In the past few months, banks started offering Aadhaar-enabled banking access from the business correspondent outlets, especially Aadhaar Enabled Payments (AEPS). The usage of AEPS has not gained traction, the causes for which need to analysed as there are few field studies.
Part II: Implications on Economy and Governance

The introduction and linking of Aadhaar to a number of government programmes has immense implications for governance and the economy. Though implementation of various programmes using Aadhaar as the basis may be a work in progress, there is a need to critically understand not only its impact on the lives of the various stakeholders, the manner in which it is perceived by the stakeholders, but also the issues and challenges that biometric authentication offers during the course of its usage – most of which have been overlooked.

The greatest benefit is for the government and business. where for the first time, it offers to undertake advanced analytics that can expand their own abilities and activities. Invariably, the one benefit of biometrics authentication and more importantly Aadhaar in the Indian context is something which nobody wants to publicly discuss – the advantage that it can offer to various businesses.

Over the past few decades, Indian business, especially the financial sector has been unable to take advantage of the growing consumption culture because they are unable to undertake advanced analytics that are possible in some of the other countries due to the lack of data on consumption patterns. This is natural in a country with wide geographic diversity and lack of connectivity. Various attempts to expand market access have been grounded by an inability to analyse the socio-economic background of consumers.

The most recent, high profile of such failures is the case of the microfinance sector in Andhra Pradesh in 2009-10, where an attempt to expand aggressively led to the use of coercive recovery practices by the lenders which led to the State Government to intervene decisively on behalf of customers. One of the reasons for this failure is because of multiple borrowing and lending – often from as many as 7-8 companies by daily wage labourers. In other words, the lenders who were only interested in expanding their portfolio had no
method (and interest) to cross check the ability of borrowers to repay their loans. Thus, any analysis of consumption and behavioural pattern becomes easy and possible only if the transactions generate data.

The larger the availability of transaction and other information, the more effective the analysis; hence, the only way that such transactional data can be generated is by linking all economic activity and benefits drawn to Aadhaar or to any number around which analysis can be undertaken. This need has now been made possible with the concurrent growth of telecommunications and computing power which helps an analysis of the transaction data. More importantly, as data itself becomes a commodity, it gains a unique value in the marketplace thanks to the valuation model and spread of the financial markets. Little wonder that the most important push of Aadhaar has been in areas of consumption and other forms of economic activity rather than in other spheres.

1. Cost-Benefit of Aadhaar

The cost-benefit of Aadhaar has been a matter of much and never-ending debate. Aadhaar has not come cheap. There are essentially two types of costs: (a) cost of the enrolment and its initial establishment and, (b) cost of integration – especially the challenge and cost of integrating existing databases and programmes with Aadhaar.

A study by National Institute of Public Finance and Policy (NIPFP) claims that substantial benefits would accrue to the government by integrating Aadhaar with schemes such as PDS, MNREGS, LPG and fertiliser subsidies as well as other programmes. The benefits accrue in the form of reduction in leakages that occur due to identification and authentication issues. According to NIPFP, even after taking all the costs of developing, maintaining and integrating Aadhaar with different schemes over the next few years, assuming that the
present system has leakages that amount to 7-12%, the internal rate of return in real terms from the implementation of Aadhaar could be as high as 52.85% over the long-term. The NIPFP estimates that it will cost Rs. 37,187 crore over a 10-year period (up to 2021) which includes the cost of developing, maintaining and integrating Aadhaar with existing databases17.

However, Reethika Khera thinks that the analysis is not persuasive. The analysis is based on assumptions, not estimates due to benefits of integration with Aadhaar, where estimates (not assumptions) of bogus beneficiaries are used, they are unreliable and out of date. Further, the report does not take into account alternative technologies that could achieve the same or similar savings, possibly at lower cost. She cites the case of Chhattisgarh where the leakages have reduced from 50 percent to 10 percent between 2004-05 and 2009-10 due to the use of local biometrics rather than Aadhaar18.

Rahul Lahoti opines that the tall claims of potential savings in LPG leaves many questions unanswered about the extent and mechanism of savings through Direct Benefits Transfer (DBT). According to him, even if the reduction in leakages in LPG subsidies turns out to be substantial, the government should be especially careful about extrapolating this impact on other subsidies such as PDS and MGNREGS, since they significantly differ in the way they are structured and used. He points out that while LPG is mostly urban and is centrally administered by a few companies through a fully computerised list of beneficiaries and does not need biometric verification, PDS and MGNREGA are more rural based, managed through multiple agencies with only partial computerisation of user lists and need repeated biometric verification thereby making it

more challenging\textsuperscript{19}. Extrapolating it further, he adds that it is clear that the context in which many of these schemes are very different and in the case of rural areas, the last mile issues are more important than in urban areas.

It is in these last mile areas that the role and contribution of Aadhaar has been far less beneficial than what it was claimed and is often celebrated. Kieran Clarke proffers that based on the information in the public domain in 2015-16, ‘instead of resulting in significant savings, the net fiscal impact of integrating Aadhaar into DBTL in the current financial year was significantly negative, and that expectations of substantial net savings in subsidy expenditure from the introduction of the Aadhaar scheme are likely to be misplaced\textsuperscript{20}.

Proponents however claim that the introduction of Aadhaar will save the government huge amounts of money. However, the headline claims about savings that accrue from the use of Aadhaar is controversial. While the government claims that till date it has already saved about Rs. 14,672 crore by using Aadhaar through various DBT programmes, a recent study by a Canadian agency claimed that the government actually incurred a loss of Rs. 97 crore.

However, it is too early to effectively compute or understand the costs or benefits of Aadhaar since the project is yet to be fully operational. For the government, the economic benefits will be far fewer than what has been thought till date. The best case scenario for Aadhaar is that it will allow de-duplication and will facilitate a one-time readjustment in the cost of expenditure of the government.

The major benefit for the government is that it will allow cash transfer. Aadhaar will lead to a change in system of subsidy delivery

\textsuperscript{19} http://www.epw.in/journal/2016/52/web-exclusives/questioning-%E2%80%9Cphenomenal-success%E2%80%9D-aadhaar-linked-direct-benefit
from the present system of providing tangible commodities and services including those related to social protection services. Though this will invariably have a political cost, it will lead to a major change in the manner in which administration will be carried out in the country. In the case of others like the public distribution system, it will create more inconvenience for the beneficiaries. However, the long-term benefits of DBT on the poor are as yet largely unstudied and most of the expectations are based on theoretical assumptions.

2. Looking Beyond Cost-Benefit Analysis

Calculating the cost benefit analysis of technology adoption is difficult and it is a problem that policy makers frequently grapple with not just in India but all over the world. Even after more than a decade of rapid technological change, there are no satisfactory methods to evaluate the contribution of technology in general and data in particular to economic output. Quantifying the use and advantage of biometric is even more problematic. Nandan Nilekani has pointed out that only in about five percent of the cases, fingerprint authentication is used\(^{21}\).

Andhra Pradesh government, which publishes monthly UID authentication data, indicates that out of the total 1.968 crore Aadhaar authentications in June 2017, only 3.03 lakhs used iris. The success rate was 96.77 percent with the failures being 6.34 lakh authentications (or 3.22 percent) as indicated in the table below\(^{22}\).

---

\(^{21}\) [Website](https://qz.com/957607/nandan-nilekani-aadhaar-is-being-demonised-because-its-so-transparent/) (Website last visited 4 July 2017.

\(^{22}\) AP government data sourced from Chief Minister Dashboard: [Website](http://epostest.ap.gov.in/eposcmdashboard/AuthenticationAbstract.jsp) (Website last accessed 05 July 2017).
### Andhra Pradesh Biometric Authentication Details June 2017

<table>
<thead>
<tr>
<th>District</th>
<th>Total Authentifications</th>
<th>Biometric Success</th>
<th>Total Success</th>
<th>Success %</th>
<th>Total Failure</th>
<th>Failure %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Single</td>
<td>Fusion</td>
<td>IRIS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ananthapur</td>
<td>1605647</td>
<td>577917</td>
<td>963706</td>
<td>6785</td>
<td>1548408</td>
<td>96.43</td>
</tr>
<tr>
<td>Chittoor</td>
<td>1710509</td>
<td>652288</td>
<td>994166</td>
<td>7558</td>
<td>1654012</td>
<td>96.69</td>
</tr>
<tr>
<td>East Godavari</td>
<td>1764576</td>
<td>729001</td>
<td>921332</td>
<td>53625</td>
<td>1703958</td>
<td>96.56</td>
</tr>
<tr>
<td>Guntur</td>
<td>1642327</td>
<td>690885</td>
<td>877730</td>
<td>7675</td>
<td>1576290</td>
<td>95.97</td>
</tr>
<tr>
<td>Krishna</td>
<td>2346712</td>
<td>1249318</td>
<td>1044212</td>
<td>8371</td>
<td>2301901</td>
<td>98.09</td>
</tr>
<tr>
<td>Kurnool</td>
<td>1557576</td>
<td>620042</td>
<td>892435</td>
<td>2787</td>
<td>1515264</td>
<td>97.28</td>
</tr>
<tr>
<td>Nellore</td>
<td>1103817</td>
<td>407011</td>
<td>641675</td>
<td>4356</td>
<td>1053042</td>
<td>95.4</td>
</tr>
<tr>
<td>Prakasam</td>
<td>1091964</td>
<td>440197</td>
<td>593569</td>
<td>1449</td>
<td>1035215</td>
<td>94.8</td>
</tr>
<tr>
<td>Srikakulam</td>
<td>1536167</td>
<td>630387</td>
<td>828288</td>
<td>40094</td>
<td>1498769</td>
<td>97.56</td>
</tr>
<tr>
<td>Visakhapatnam</td>
<td>1364607</td>
<td>559185</td>
<td>695267</td>
<td>57932</td>
<td>1312384</td>
<td>96.17</td>
</tr>
<tr>
<td>Vizianagaram</td>
<td>982497</td>
<td>409785</td>
<td>492761</td>
<td>54347</td>
<td>956893</td>
<td>97.39</td>
</tr>
<tr>
<td>West Godavari</td>
<td>1957632</td>
<td>833673</td>
<td>1005954</td>
<td>67908</td>
<td>1907535</td>
<td>97.44</td>
</tr>
<tr>
<td>YSR Kadapa</td>
<td>1021312</td>
<td>381853</td>
<td>597554</td>
<td>7368</td>
<td>986775</td>
<td>96.61</td>
</tr>
<tr>
<td></td>
<td>19685343</td>
<td>8181542</td>
<td>10548649</td>
<td>320255</td>
<td>19050446</td>
<td>96.77</td>
</tr>
</tbody>
</table>

*Source: AP CM Dashboard (data accessed on July 05 2017)*

The failure rate can be considered to be high because Andhra Pradesh uses various methods of authentication including fingerprints and iris for the public distribution system. Hence, it is safe to assume that the authentication failure varies from about 2-7 percent per month and the averages can often hide widespread problems which can be as high as nearly 10 percent in some
districts\textsuperscript{23}. In such a scenario, it becomes all the more difficult to calculate the impact of a transformation process such as the spread of Aadhaar on livelihoods.

The cost-benefits of Aadhaar raises a number of issues, especially since the user agencies and even UIDAI tend to be more secretive about disclosing information about data generated, the manner in which it has been used and the nature of users. The UIDAI places very little such information in the public domain and there are no independent agencies like the Comptroller and Auditor General (CAG) that audits these claims. UIDAI discloses very little information and most of the gleanings about its efficiency have to be based on the few statistics available in the public domain\textsuperscript{24}. This counter-productive approach that black out any information is often the answer to criticism that Aadhaar and biometric authentication compromise on privacy and security.

The case of Andhra Pradesh is illustrative of the issues surrounding the importance of Aadhaar and its ability to facilitate the delivery of government benefits in the last mile. This is assuming that obstacles to greater last mile access like connectivity are available. The data indicates\textsuperscript{25} that the statistics may be masking larger problems related to access of rations. Though the numbers indicate a halving of failures from a high of 7.14% in January 2017 to 3.56% in June 2017, this is accompanied by a commensurate 61% fall in the number of authentications in June as compared to January.

\textsuperscript{23} There are wide variations in the failure rates within districts and in months. February 2017 had a failure rate of 6.31% while January 2017 had 7.12% for the state as a whole. It varied from 2.37% (February 2017) failure rate in Krishna district to about 9.86% in Vizagaram in the January 2017.

\textsuperscript{24} https://scroll.in/article/833060/how-efficient-is-aadhaar-theres-no-way-to-know-as-the-government-wont-tell

\textsuperscript{25} Annexure I provides details of month wise Total Authentications and the failure rate.
Thus, the headline decrease in failures is because of the decrease in number of authentications rather than increased efficiency of Aadhaar and the process of authentication. Another peculiarity is the large increase and decrease in the total number of authentications in alternative months. One wonders why there is such a large increase or decrease in the total authentications in a neat pattern of alternate months.

Interestingly, the failure rate is highest in the districts which are known to have high levels of migratory labour. Similarly, the authentications and failures are highest when a large number of people are present in the villages (between December and February). This raises important questions about the efficiency of Aadhaar and the State Government’s claims about their introduction of ‘anywhere rations’ through their programme called “ePoS”. Further, studies need to be undertaken to throw more light on this aspect of Aadhaar and its role in the supply of ‘efficient’ rations.

The failure in Aadhaar authentication largely precludes any discussion of the quality of biometrics collected during the enrollment process. The quality of biometrics captured is not known. Anecdotal evidence seems to indicate that the inability to capture biometrics in three attempts led to the person employed in the enrolment centre often using the ‘manual’ override to force the system to capture the biometric irrespective of the quality. There is no way of cross verifying the quality of biometrics stored, especially by the person who has enrolled.

Hence, unless different authentic methods are available concurrently for authentication, there is a risk that citizens will be troubled due to the quality of biometrics captured and stored. In the worst case scenario, it could lead to ‘identity denials’ wherein a person can be denied the fact that they are who they are. This lack of quality could be the reason for high level of authentication failures even in states like AP. However, UIDAI claims that it uses three independent biometrics to ensure de-duplication and minimal errors; these
include fingerprints of all 10 fingers, iris (both left and right) and face\textsuperscript{26}.

It is in this context that the doubts raised by the statutory auditor to the government, Comptroller and Auditor General of India (CAG) are important. The CAG is reported to have observed that the savings in the case of petroleum products was due to the collapse of price rather than de-duplication\textsuperscript{27}. The savings claimed by the Central Government during the financial year 2014-15 and 2015-16 was Rs. 22,000 crore by way of direct cash transfer of subsidy and voluntary surrender of subsidy by better off consumers. In contrast, the CAG noted it was added up to less than Rs. 2,000 crore. Most of the savings was due to the drastic fall in prices of imported LPG prices due to global factors.

Anecdotal evidence from Kurnool district of Andhra Pradesh seems to indicate that the savings could well be due to the exclusion of genuine beneficiaries due to mistakes in the ration cards rather than de-duplication efforts. Independent verification of the claims related to importance and benefits of Aadhaar is lacking and is the need of the hour.

Assuming that only five percent are denied government benefits due to issues with Aadhaar, does it mean that this exclusion of a small minority is condonable in a democratic society? This exclusion in a country of more than a billion can mean that about 50 million people lose benefits. In the European context, 50 million could well be more than the population of most countries.

In India, there have been reports about large number of people being denied benefits because of Aadhaar. The case of Telangana, where about thirty six percent of the Aadhaar authentications failed

\textsuperscript{26} Unique Identification Authority of India, “Erring on Aadhaar”, Economic and Political Weekly, pp. 84-85, March 12 2016

\textsuperscript{27} “CAG Audit Nails Centre’s Claim on LPG Subsidy Saving” http://www.thehindu.com/news/national/CAG-audit-nails-Centre%E2%80%99s-claim-on-LPG-subsidy-saving/article14499201.ece, July 20, 2016
is a clear case of possible issues with biometric authentication\textsuperscript{28}. Most of these were blamed on biometric mismatch due to wear and tear of fingers for manual labourers in rural areas. Iris scanners were not used because they are expensive. These claims have not been clarified by UIDAI or other government agencies. Instead, the oft repeated claim is that the system has not been designed well and is contrasted with Andhra Pradesh\textsuperscript{29}. However, it is pertinent to note that till 2014, the same system designed was in vogue in the states that now comprise Andhra Pradesh and Telangana.

3. Aadhaar and Road Ahead: Negotiating the Challenges

Aadhaar faces a number of challenges over the short and long-term. The primary challenge is to protect the data from prying and excessive profit seeking excess of the business world. It is well-known that businesses are increasingly operating in a highly competitive world in which ethical boundaries are rapidly being pulled down. The problem is compounded because they have to satisfy their shareholders in a competitive business environment that rarely looks beyond the quarterly profits and the operational dynamics of stock market listing.

A more worrisome aspect is that Aadhaar is increasingly becoming the pivot for the operation of various economic and administrative activities. In an era when cyber threats are frequent, the major challenge for UIDAI is to protect the data under its control since the biometrics is now an important national asset which has huge

\textsuperscript{28} “Aadhaar Fails MGNREGS Test in Telangana”, http://www.livemint.com/Politics/UfSB33ZB2sYKpmLqwMke8O/Aadhaar-fails-MGNREGS-test-in-Telangana.html, April 07, 2017

\textsuperscript{29} Nandan Nilekani in an interview explained it as a problematic design in the programme and contrasted with AP. https://qz.com/957607/nandan-nilekani-aadhaar-is-being-demonised-because-its-so-transparent/ (Website last visited 4 July 2017)
ramifications for various government programmes and the banking system. Thanks to Aadhaar, for the first time in the history of India, there is now a readily available single target for cyber criminals as well as India’s external enemies. In a few years, attacking UIDAI data can potentially cripple Indian businesses and administration in ways that were inconceivable a few years ago. The loss to the economy and citizens in case of such an attack is bound to be incalculable.

A more important issue that Aadhaar has to grapple with in the near term is the legal challenge pending in the Supreme Court. The case largely deals with criticism that Aadhaar impinges on the privacy of the individual. The problem is now compounded with the linkage of bank account, driving license, phone number and a host of other services to Aadhaar. It means that the Aadhaar number will be available in databases of each and every service provider. Hence, any breach of the database of one provider has the potential to compromise the details of the Aadhaar numbers. Second, a more serious privacy concern is that UIDAI promises to archive the data for five years. Since UIDAI’s vision has no qualms about allowing private service providers, Indian and foreign, to participate in developing an ecosystem related to it, one wonders how security of data of companies that are essentially foreign in nature and those that have little legal and economic exposure or interest other than profit in India, can be ensured.

The third important issue is how businesses will use Aadhaar? During enrolment, the default option was the explicit approval of the citizen to share the data. Aadhaar when used with GST, which can generate as much as 320 crore transactions a month\(^{30}\) makes it a potent marketing tool for those who control data and giving those who access it an unfair advantage in the market place. Since most of these are linked to Aadhaar and PAN, it means the information

collected about transactions will be something hitherto unknown in the history of India. The fact that GSTN is a private company with its largest shareholders having conflict of interest due to their own business activities, there is a need to check and take measures that does not grant unfair advantage and will stop the creation of oligopolies in the market place.

One important reason why privacy advocates seem justified is that in all the cases of violation of the Aadhaar Bill, no penal action or at best only minimal action has been taken by UIDAI or the Government. Section 29 of the Aadhaar Bill requires that the information collected should not be posted or distributed publicly. However, the case of a government department in Jharkhand publicly disclosing the address and Aadhaar numbers of beneficiaries does not inspire much confidence.

The recent case of Aadhaar data breach by Axis Bank business correspondent Suvidha Infoserve and another service provider eMudhra, who attempted unauthorized authentication and impersonation using illegally stored Aadhaar biometrics indicates that the concerns about privacy are not misplaced. The same data breach also led to accusation of multiple transactions undertaken with the same fingerprint. The inability of UIDAI to take serve action against the powerful private sector bank leads to questions about the ability to protect consumers. It is insufficient for UIDAI to claim that the problem is under control. In both the cases cited above, the violators got away lightly, indicative of the fact that the larger the offender/violating agency, the less likely the possibility of penal action. This contrasts with the issue of notices by UIDAI to those privacy advocates who have been campaigning for more security.

The manner in which UIDAI has responded to these criticisms is that privacy safeguards are built into the database. One argument made by Aadhaar is that unlike a physical document, e-KYC or Aadhaar cannot be misused nor can it be used for other purposes. The facility wherein a person can “lock” their Aadhaar number is not a sufficient
remedy for concerns related to Aadhaar\textsuperscript{31}. In a country where literacy and awareness are minimal or non-existing in most circumstances, the possibility of misuse of Aadhaar is as high as that of a physical document.

Innumerable instances have come to light of short changing of citizens or organized fraud despite the use of biometrics. Hence, there is a need for UIDAI to assuage the privacy concerns in a far more proactive way rather than serve as a tool for businesses to exploit consumers. There is also a need for a more robust and comprehensive law that protects the use and misuse of the huge amount of data that is now being generated and collected.

A perceptive article pointed out four issues with Aadhaar, around which there is widespread mistaken perception including (a) it is not an address proof, (b) it is not proof of citizenship, (c) Aadhaar alone is not an identity proof and, (d) Aadhaar can but should not be made public\textsuperscript{32}.

4. Conclusion

The above overview of Aadhaar indicates that, while Aadhaar introduces a hitherto new feature – biometric authentication on a large scale – the benefits to the consumers have been mixed, with not much benefit to those in the last mile. In the realm of business and administration, it promises to have a larger impact than previously thought. There is a need for caution in the manner in which Aadhaar is used by the Government, especially as more


\textsuperscript{32} https://scroll.in/article/832595/privacy-security-and-equality-are-not-the-only-serious-problems-with-aadhaar-here-are-four-more
programmes and economic activities are linked to it. Only time will tell if the benefits outweigh the costs or vice versa.

**Bibliography**


## Annexure I

### Month wise Total Authentications and Failure Rates

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ananthpur</td>
<td>15009972</td>
<td>1638068</td>
<td>8.71</td>
<td>2798228</td>
<td>7.14</td>
<td>1626731</td>
<td>7.12</td>
<td>1880943</td>
<td>5.66</td>
</tr>
<tr>
<td>Chittoor</td>
<td>1397136</td>
<td>1707560</td>
<td>8.02</td>
<td>2643880</td>
<td>5.66</td>
<td>1653091</td>
<td>5.86</td>
<td>2177875</td>
<td>3.31</td>
</tr>
<tr>
<td>East Godavari</td>
<td>1962450</td>
<td>2529802</td>
<td>13.10</td>
<td>3548775</td>
<td>4.15</td>
<td>2852679</td>
<td>2.37</td>
<td>2678124</td>
<td>2.05</td>
</tr>
<tr>
<td>Guntur</td>
<td>1547980</td>
<td>2231923</td>
<td>8.42</td>
<td>2529802</td>
<td>8.42</td>
<td>2371542</td>
<td>5.01</td>
<td>2177875</td>
<td>3.31</td>
</tr>
<tr>
<td>Krishna</td>
<td>1668586</td>
<td>3258056</td>
<td>6.79</td>
<td>3548775</td>
<td>4.15</td>
<td>2965918</td>
<td>2.12</td>
<td>2947633</td>
<td>1.73</td>
</tr>
<tr>
<td>Kurnool</td>
<td>1513037</td>
<td>1849634</td>
<td>6.82</td>
<td>2614445</td>
<td>7.87</td>
<td>1212085</td>
<td>9.52</td>
<td>1931467</td>
<td>4.71</td>
</tr>
<tr>
<td>Nellore</td>
<td>1038905</td>
<td>1293054</td>
<td>6.88</td>
<td>1845755</td>
<td>7.70</td>
<td>1221503</td>
<td>4.03</td>
<td>1125639</td>
<td>4.20</td>
</tr>
<tr>
<td>Prakasham</td>
<td>1145775</td>
<td>1313152</td>
<td>9.58</td>
<td>2115466</td>
<td>8.84</td>
<td>1329808</td>
<td>6.02</td>
<td>1255874</td>
<td>4.73</td>
</tr>
<tr>
<td>Srikakulam</td>
<td>1132255</td>
<td>1367439</td>
<td>20.12</td>
<td>1977099</td>
<td>9.10</td>
<td>1331805</td>
<td>7.93</td>
<td>1405729</td>
<td>5.90</td>
</tr>
<tr>
<td>Visakhapatnam</td>
<td>1384016</td>
<td>1726915</td>
<td>18.96</td>
<td>2161268</td>
<td>7.19</td>
<td>1546082</td>
<td>7.40</td>
<td>1565105</td>
<td>5.89</td>
</tr>
<tr>
<td>Vizianagaram</td>
<td>955012</td>
<td>1156492</td>
<td>18.56</td>
<td>1513678</td>
<td>9.86</td>
<td>989827</td>
<td>7.01</td>
<td>1071210</td>
<td>5.33</td>
</tr>
<tr>
<td>West Godavari</td>
<td>1734464</td>
<td>2039546</td>
<td>15.68</td>
<td>2771902</td>
<td>5.11</td>
<td>2181273</td>
<td>5.68</td>
<td>2517626</td>
<td>1.99</td>
</tr>
<tr>
<td>YSR Kadapa</td>
<td>1006391</td>
<td>1213209</td>
<td>5.92</td>
<td>1723013</td>
<td>6.45</td>
<td>2181273</td>
<td>7.30</td>
<td>1223483</td>
<td>4.76</td>
</tr>
<tr>
<td>Total</td>
<td>17995979</td>
<td>23324850</td>
<td>10.39</td>
<td>3188290</td>
<td>7.12</td>
<td>2062087</td>
<td>6.31</td>
<td>2242677</td>
<td>3.54</td>
</tr>
</tbody>
</table>

Source: Compiled from Respective Months (http://epostest.ap.gov.in/eposcmdashboard/AuthenticationAbstract.jsp); Compiled on July 07, 2017
IDRBT Staff Paper Series

1. Mobile Banking - September 2015
2. Payment Systems - June 2016

Access these Staff Papers from www.idrbt.ac.in/staffpapers.html

Frameworks from IDRBT

2. FAQs on Cloud Adoption for Indian Banks
3. Digital Banking Framework
4. Cyber Security Checklist
5. IT Vendor Management: Principles & Practices
6. Data Quality Framework
7. Cloud Security Framework
8. Green Banking Framework
9. Social Media Framework
10. Information Security Framework
11. Information Security Governance
12. Holistic CRM and Analytics
13. Organizational Structure for IT
14. Micro-ATM Standards

Access these Frameworks from www.idrbt.ac.in/bestpractices.html