Case Study of Voice Based Security System

Sonawane Pushkar[1], Ahire Mrunal[2], Mathpati Vaishnavi[3], Jagtap Sanket[4], Chavan Sanskruti[5]

[1-5]Department of Mechanical Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, India.

Abstract: *Voice/speech is the most effective and prominent system for communication among all the human beings. Every human being is gifted by different voices i.e. no two persons in this world can have the same voice. By using such benefit of the unique voice of every person, huge inventions took place in the engineering field keeping this uniqueness of the voice at the centre of those inventions. Voice has tremendous potential to interact with computer. Voice/speech started using in mobile phones, railway technology, banking securities, defence field, confidential institutions, biometric security systems and many more. In this article, we will be going through the science and engineering behind this 'Voice based security system'. Some biological aspects are there which makes the voice different from all the other people so this paper also going to analyze those aspects deeply and the engineering and technological facts which effectively used those biological aspects in 'Voice based security system'.[1]*

Keywords: Speech, security, pitch, frequency, amplitude, analysis, system, voice, source

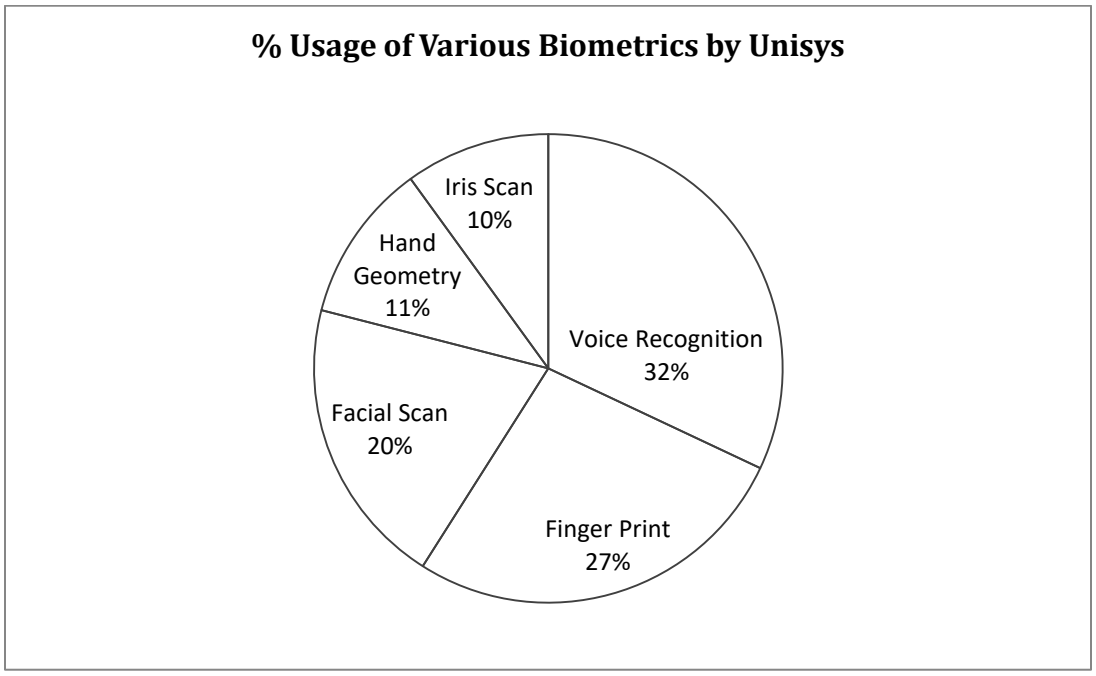
INTRODUCTION:

Biometrics refers to the physiological aspects as well as behavioral aspects which are related to human characteristics. This is used in computer science for identification of user's availability and access[1]. Biometrics, these days, has tremendous applications and scope because of various reasons such as:

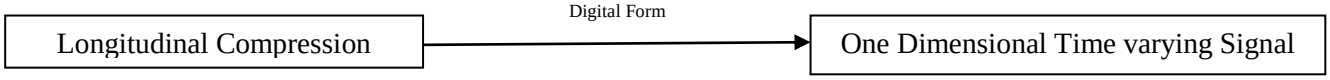
1. The person to be identified at the place of biometrics should be present at that place physically.
2. There is no need to remember password for longer time and change it continuously time to time for security.
3. Less time required to get access as compared to other security provisions such as passwords and patterns.
4. Proxy methods can be avoided.

According to the Survey of Unisys, the poll from the consumers for using various biometric applications can be listed in the following table –

Type of Biometrics	Percentage consumed
Voice Recognition	32 %
Finger Print	27%
Facial Scan	20%
Hand Geometry	11%
Iris Scan	10%



As the statistics given by Unisys, voice recognition has 32% applications in various sections of biometrics. Voice recognition basically works on the phenomenon of speech. Speech is generated by the excitation of acoustic tube which is known as Vocal Tract in biological language. The vocal tract has two ends i.e. the end of the vocal tract at lips and the starting of the vocal tract which is connected to glottis. It generates a longitudinal wave. If one converts this longitudinal wave into digital form, the speech signal will be one dimensional time varying signal.



The speech generated from the vocal tract can be classified in three domains:

1. Voiced Sounds- A speech sound accompanied by sound from the vibrations of the vocal cords.
2. Fricative Sounds- The energy of this type of sound is gathered at the peak of frequency band and it is like a noise in its appearance.
3. Plosive Sounds – Strong energy appears at many frequency bands as an explosion.

Following are some of the methods to do the speech analysis:

1. Oscillogram:
   Oscillogram is the most general method used for speech recognition. It is also called as waveform. Speech signal is the series of pressure variation between the source of the speech and the receiver. In this system, the time axis is horizontal axis and the pressure variation of the speech is plotted on it weather it is increasing or decreasing. The drawback of such system is that it is very difficult to extract a suitable structure from the large amount of information and to reach at any conclusion. [11]
   https://academo.org/demos/virtual-oscilloscope

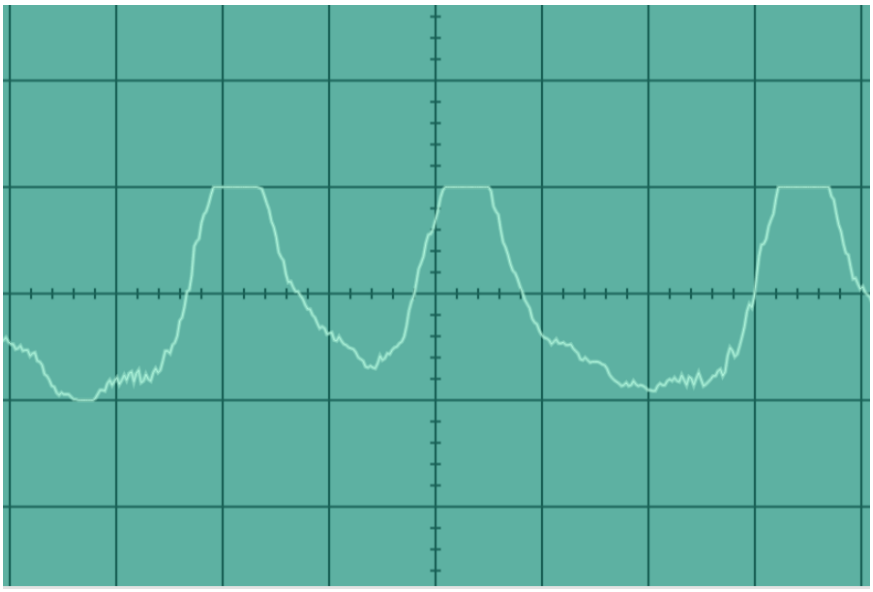


Fig 1 Oscilloscope Plot

2. Fundamental Frequency (Pitch):
Speech signal can also obtained by pitch analysis. This results about how listener perceives the stress and accent of the speaker. Speech is physical process consisting of two parts i.e. a product of a sound source (vocal cords) and filtering by tongue, lips, and teeth. In pitch analysis, what is done that it tries to capture the fundamental frequency which produced by the vocal cords by analyzing the final speech pronunciation. The fundamental frequency is the dominating frequency produced by the vocal cords. Several algorithms are developed to do the analysis of such frequency as it is quite difficult to perform.
https://phet.colorado.edu/sims/html/wave-on-a-string/latest/wave-on-a-string_en.html

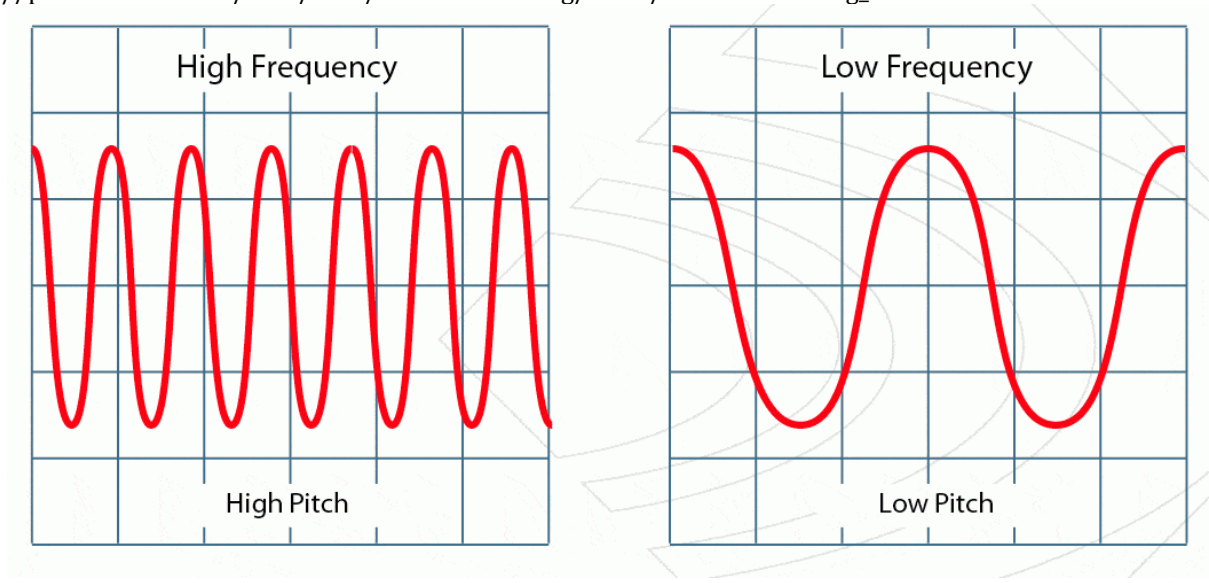


Fig 2 Frequency Pitch Analysis

3. Spectrum:
The spectrum[13] gives a picture of distribution of frequency and amplitude at a moment in time. The horizontal axis represents frequency and the vertical axis represents the amplitude. We can plot the spectrum in three dimensional diagrams called 'Spectrogram' if we plot it as a function of time.
https://academo.org/demos/spectrum-analyzer

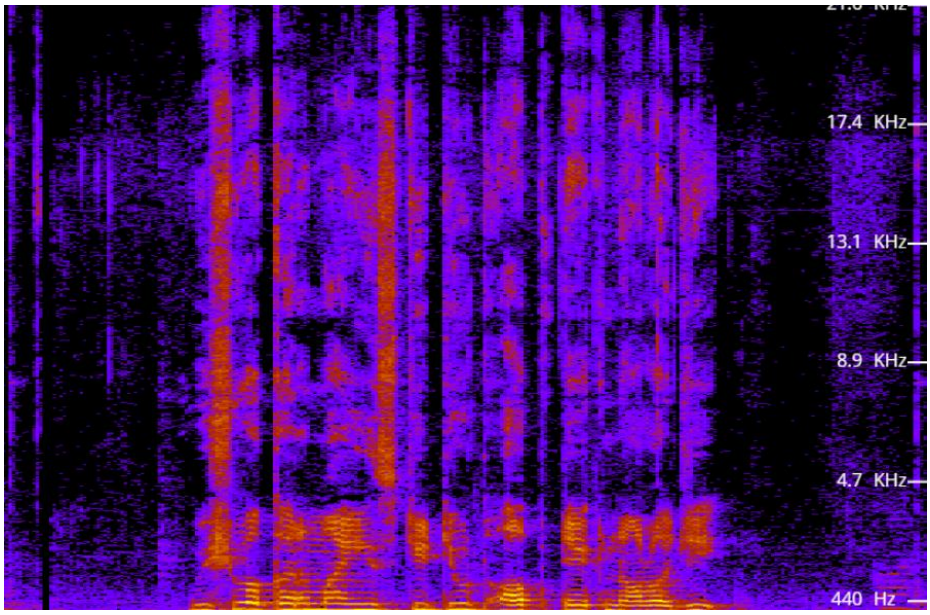


Fig.3 Spectrum Analysis Plot for "Vishwakarma Institute of Technology" words

4. Spectrogram:
In the spectrogram, the time axis is the horizontally plotted with the vertical axis of frequency. The amplitude, which is the third axis, represented the shades of darkness in the plot. The dark spots in the spectrum forms horizontal bands across the spectrogram. These bands are also called as formants. The position of formants is different for different sounds. The bands represent the frequencies where the mouth produces resonance to the sound.
https://musiclab.chromeexperiments.com/Spectrogram

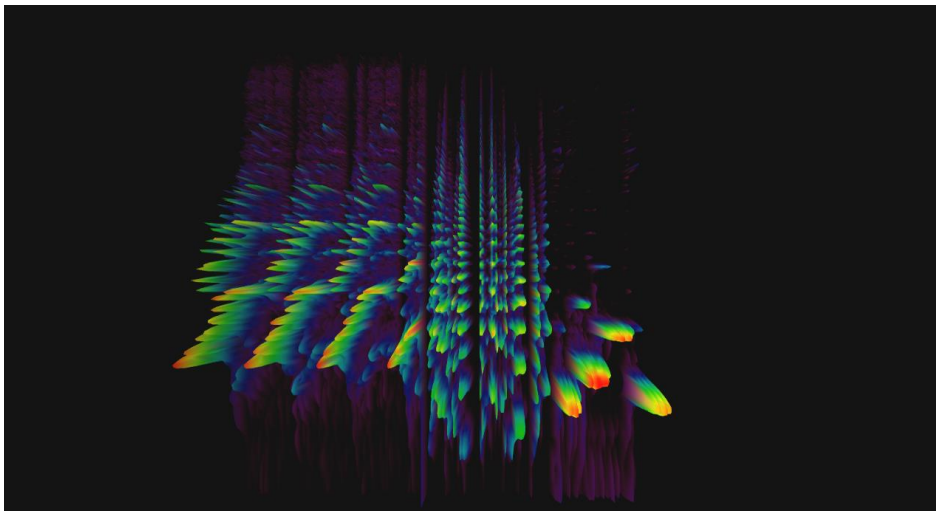


Fig.4 Spectrogram Plot for Flute, Violin and Glass Sound

5. Cepstrum:
This speech analysis is done by spectral representation of the signal. It can be done by using the relation –
r = v * s
where,
v = Filtering by person's voice box and throat.
r = what we can record.

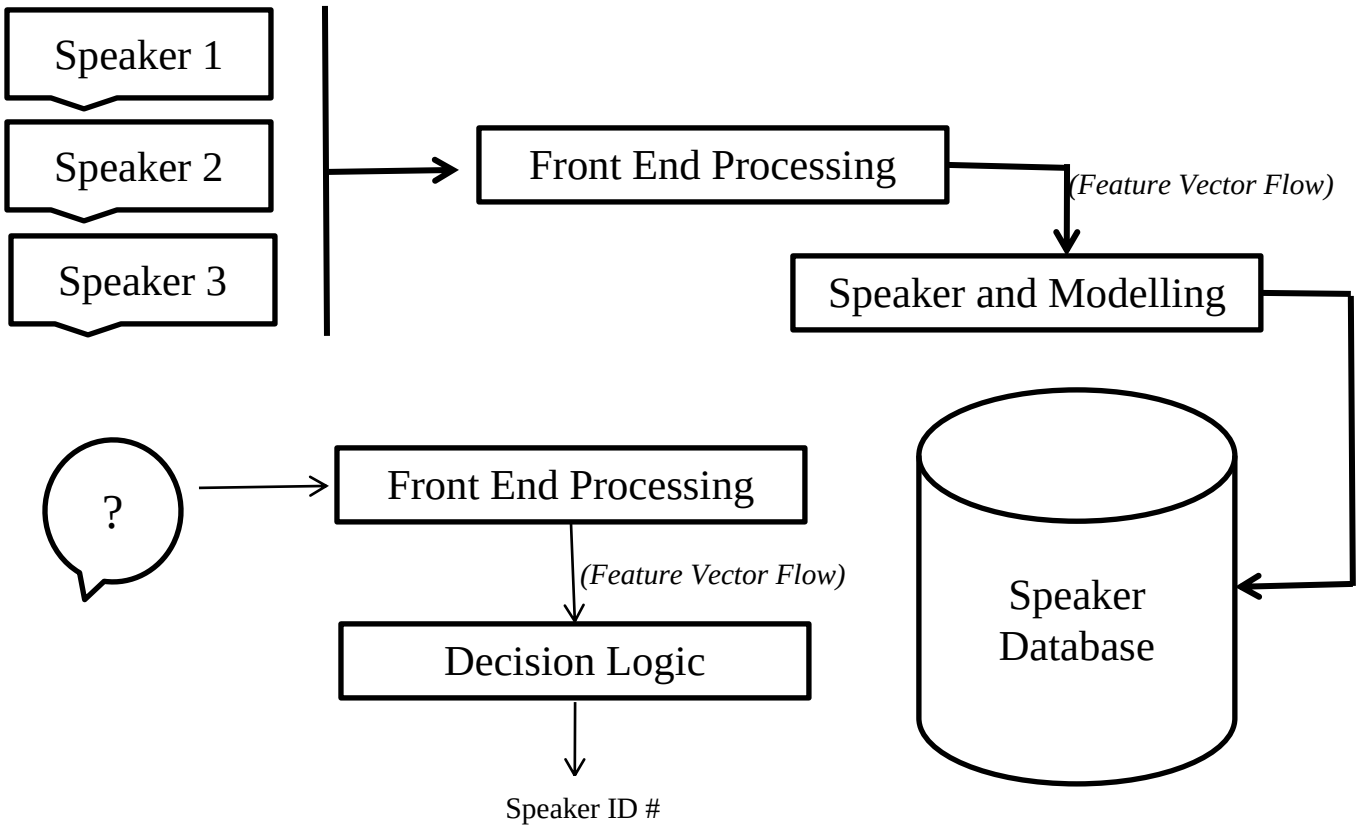
s = speech.

Description of VBSS:

To unlock any data after complete recognizing a voice of user or authorizer is the main function of Voice Based Recognition System. VBSS systems are only accessed if the user is in conscious or stable state. If he/she is in any unusual state such as worried, horrified then intensity of voice changes and this will not match with the voice which is already stored in system then it will not access system. Such systems provide instant messages to authorizer/main owner if anything goes wrong. This system sends messages to main owner in cases like[2]

1. When user allows anyone to access the system
2. When Unknown person tries to break lock
3. If someone tries to play with such systems.

Layout of Voice Recognition System:



The main problem/ work faced by this system is to recognize a particular speaker based on his/her voice input. Two main difficulties is to

- Identifying features in a known person's voice.
- Searching new samples for future.

There are basically two types/subtasks of speaker recognition system:

1. Speaker identification – In this, system checks and conform that the particular person is who he claims to be.
2. Speaker verification – the system defines and verifies the particular person is from the group of persons or not.

A speaker recognition system consists of different modules. Those modules are as follows:

1. **Front end processing:**
   This is a signal processing part, which converts the sampled signals or inputs into different set of vectors. This part characterizes different properties of speech/input that can separate different speakers. Front end processing is mainly performed in both training phase and recognition phase.
   This process contains some steps:

   Acquisition :
   Speech signature data acquisition:
   A user can store samples at different times of the day rather than all at one go to get good variability. As user tend to speak in a very prototyped manner if they have to repeat the same thing many times in a short period of time. As much speaker should speak clearly into microphone without any emphasis on any particular word. The recording should be done in a situation where there is less ambient noise or disturbance.

   Microphone issues :
   Overall microphones measure something roughtly proportional to intensity waveform of the sound or its first derivative by transuding movement of air into electric voltage using different diaphragrames. Which type of microphone used and then its distance from mouth als affects the response of systems. Each individual microphone exibits a slightly different and usually nonlinear transformation from true pressure signal to its output voltage. Ambient noise may affect the working of microphones which will give errors.

   Processing :
   Once the input is provided, theres need to remove any dead air at the beginning and the end. If user break up the voice sample into small frames then an overlap will occur. Each frame then capture certain part of specch only, which is windowed using a hamming window.

   Windowing :
   A Hamming window lays more emphasis on the center of frame and lesser emphasis on the edges side, thus minimizes the spectral distortion.
   Hammering windows generated using following equation:

   $$w(n) = (54 - 46*\cos(2*pi*N/(n-1)))/100;$$

   Filtering:
   The bandwidth of filter is adopted as parameters to determine the coefficient of respective filters. In VBSS different High Pass filters are used.

   Correlation:
   In system all the analog signals are converted into digital signals and then digital signals/response will get on digital filter. A voiceprint for each sampled word is formed by using analysis function. Main function of comparator is to compare voice print of tested sample with the one which is already stored voice sample.

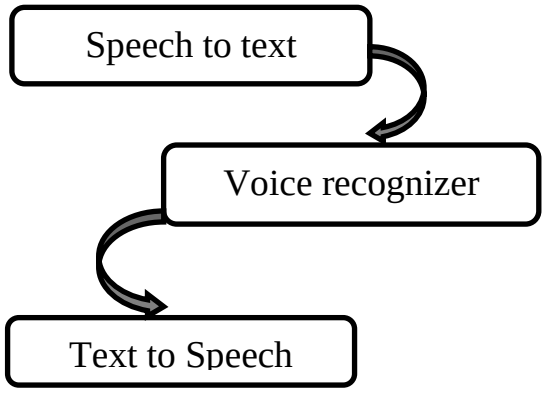
2. **Speaker modeling:**
   This part of speaker recognition system performs operations on the feature data by modeling the distributions of feature vectors.

3. **Speaker database:**
   The different speakers feature vectors are stored in this speaker database.
4. **Decision logic:**
   This unit makes the final decision regarding the identification of original speaker by comparing unknown feature vectors in all models stored in the speaker database and selecting the best matching model, or comparison with that particular user, as may be the case.

Controlling of system by using voice recognition system:

Different electronic devices can be controlled using voice recognition system. There are different modes of such systems.

- Speech to text
- Recognition of voice
- Text to speech.



Different templates are already stored in particular dictionary according to sample speech. Various parameters of voice are considered for templates such as pitch, intensity, amplitude etc. Nyquist Sampling theorem is the criterion to sample the voice.

In such systems, filters are also used like band pass filter, after sampling of analog signal that samples voice is allowed to pass through the filter to avoid or eliminate noise.

Difference between speaker recognition and verification

Speaker Recognition	Speaker verification
1. Voice samples are obtained and features are extracted from them and stored in database.	1. User supplies speakers identity and records users voice.
2. Samples are compared with various other stored ones and using methods of pattern recognition the most probable speaker is identified.	2. The goal of the speaker verification is to confirm the claimed identity of a subject by exploiting the individual differences in their speech.
3. As the number of speakers increase, this process becomes more troublesome on computer.	3. The feature extracted from the voice sample are matched against stored samples corresponding to the given user, therefore verifying the authenticity of the user.
4. When number of user increase, it becomes difficult to find unique features for each user. It may lead to wring identification.	4. In some cases, password protection is accompanies the speaker verification process for more security and reliability of system.

Applications:

- Contact centers
  Here, it is used to recognize the voice of the customer and open the profile of particular customer.
- Fraud prevention
  This technology is used to safeguard the profile of any of the person in any of the field to prevent the fraud.
- Financial services
  In banking sector, this VBSS technology is used to safeguard the digital accounts.
- Healthcare services
  In many hospitals, some specialized rooms and laboratories have the VBSS system.
- Mobile Phones
  Now days, generally every mobile phone has VBSS for security and privacy purposes.
- Public Services
  It is used to establish caller identity in a wide range of public service applications.
- Workforce Management
  In a large organization having distributed workforce, VBSS plays crucial role in worker identification for salary account, absenteeism, distributed work identification.

Conclusion:

By performing this home assignment activity which is based on case study of Voice Based Security System (VBSS),

a. We studied about the generation of speech and which was the biological aspect of this case study.
b. We came to know about various methods of speech analysis and their actual working by using online simulator.
c. We saw the effect of varying the amplitude and frequency in the simulator on the sound waves.
d. We came across the processing of the sound waves for identification with the help of general layout.
e. We studied about the front end processing and back end processing of Voice Based Security System.
f. We went through the basic difference between speaker recognition and speaker verification and its processing.
g. We studied the various applications as well as components along with VBSS are used.

References:

1. 'Voice Based Security System': paper by Abhishek Mitr, Saurabh Bisht, Vikas Ranjan: January, 2002.
2. 'Blog on Voice Biometrics' : By Aculab written on 27 August,2019.
3. 'Voice Authentication' By AWARE.
4. Security System Using Biometric Technology: Design and implementation of Voice Recognition System; Paper from International Conference of Computer and Communication Engineering; June, 2008.
5. Development of Alarm System Based on Speech Recognition : Qingsong Zhu1, Hongzhao Li, Ruyan Ye, Meide De, Fayue Zheng; Beijing Polytechnic, Beijing 100176;
   7th International Conference on Education, Management, Computer and Medicine (EMCM 2016).
6. Voice recognition : Benefits and challenges of this biometric application for access control; By SourceSecurity.
7. The National Center for Voice and Speech provides some basic information on preserving "Vocal Health" on their website.
8. Voice Users Mailing List: Detail in Q1.4.html of the FAQ.
9. " Talking to Computers Has its Hazards", by Gordon Arnaut in The Globe and Mail
10. "Talking to a PC May Be Hazard To Your Throat", by Julie Chao in the Wall Street Journal.
11. Simulator for Oscilloscope by academo.org.
12. Simulator for Frequency Pitch by phet.Colorado.edu.
13. Simulator for Spectrum analysis by academo.org.
14. Simulator for Spectrogram by musiclab.chromeexperiments.