



## SPEAKER RECOGNITION SYSTEM USING SIGNAL PROCESSING TOOL IN MATLAB

Ledisi G. Kabari

Computer Science Department  
Ken Saro-Wiwa Polytechnic,  
Bori, Nigeria

Kebin Ogie

Computer Science Department  
Ignatius Ajuru University of Education  
Port Harcourt, Nigeria

**Abstract:** Speech Recognition is the process in which certain words of a particular speaker will automatically be recognized based on the information included in individual speech waves. Speech is one of the most important medium by which a communication can take place. The invention and widespread use of mobiles, telephones, data storage devices etc. have provided a major help in speech communication and its analysis. These also brought about ever-increasing need to authenticate and identify individuals automatically. Biometrics which identifies the physical traits and behavioural characteristics that make each of us unique therefore becomes necessary as a natural choice for identity verification. Advances in Biometric technology promises an effective solution to the world security needs as it can accurately identify or verify individuals based upon their unique physical or behavioural characteristics. This paper demonstrates a speech recognition system using signal processing tool in MATLAB. MATLAB programming was used to develop code that compares the pitch and formant vectors of a known speech signal with the bunch of other unknown speech signals and identifies appropriate match. The system was repeated several times to test its accuracy and was found to always make a perfect match.

**Keywords:** Speech Recognition, Signal Processing, MATLAB, Pitch Analysis, Speech Editing

### I. INTRODUCTION

In the modern world, there is an ever-increasing need to authenticate and identify individuals automatically. Securing personal privacy and deterring identity theft are national priorities. Biometrics, the physical traits and behavioral characteristics that make each of us unique, are a natural choice for identity verification [1]. It is an emerging technology that promises an effective solution to our security needs. It can accurately identify or verify individuals based upon their unique physical or behavioral characteristics. It is a key that can be customized to an individual's access needs opening doors for one while keeping others out. We can use a biometric to access our home, our account, or to invoke a customized setting for any secure area or application.

Speech recognition is a branch of computational linguistics that develops methodologies and technologies that enable the recognition and of spoken language. It is also known as automatic speech recognition (ASR). Speech is one of the most important medium by which a communication can take place. The invention and widespread use of mobiles, telephones, data storage devices etc. have provided a major help in setting up of speech communication and its analyzing. The term and the basic concept of speech identification was began in the early 1960's with exploration into voiceprint analysis which was somewhat similar to fingerprint concept. Your voice is unique because of the shape of your vocal cavities and the way you move your mouth when you speak. It was in 1984 that a science fiction called "Star Trek to George Orwell's," derived the concept that a machine can recognize the human voice[2].

Nowadays, with further growth and advancement in the field of speech recognition, the humans who are physically challenged such as blind and deaf can easily communicate with the machines. Speaker recognition uses vocal characteristics to identify individuals using a pass-phrase. The matching strategy may typically employ approaches based on hidden Markov model, vector quantization, or dynamic time warping[3]. A telephone or microphone can serve as a sensor, which makes it a relatively cheap and easily deployable technology. However, voice recognition can be affected by environmental factors such as background noise. This technology has been the focus of considerable efforts on the part of the telecommunications industry and the U.S. government's intelligence community, which continue to work on improving reliability.

### II. SPEECH RECOGNITION

Voice or speaker recognition uses vocal characteristics to identify individuals using a pass-phrase. The matching strategy may typically employ approaches based on hidden Markov model, vector quantization, or dynamic time warping. A telephone or microphone can serve as a sensor, which makes it a relatively cheap and easily deployable technology. However, voice recognition can be affected by environmental factors such as background noise. This technology has been the focus of considerable efforts on the part of the telecommunications industry and the U.S. government's intelligence community, which continue to work on improving reliability. Basic model of speech recognition is as demonstrated in figure 1

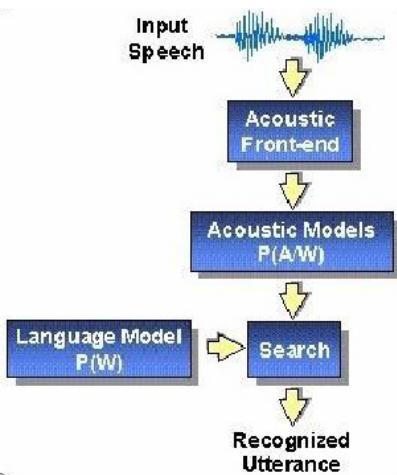


Figure 1. Model of Speech Recognition [4]

Basic principles that are involved in speech recognition are Speech Editing, Speech Degradation, Speech Enhancement, Pitch Analysis and Formant Analysis [2].

#### A. Speech Editing

In Speech Editing technique a set of the speech signal in ‘.wav’ (dot) wave format are recorded and taking a speech signal from the set of recorded speech waves, Speech Editing are then performed. Here the length of the vector representing this speech file must have a magnitude of 30,000. However, this vector is then divided into two separate vectors having equal length and in opposite order. Then with the help of MATLAB Programming and Tools a code is developed by which the given wave file is read and then the same file is played in reverse order.

#### B. Speech Degradation and Enhancement

Noise plays a vital role in speech enhancement as well as speech degradation. Thus noise estimation is one of the major parts while performing the speech recognition task. Therefore, it is understood if the estimated noise is low it will not affect the speech signal but if the noise is high then speech will get distorted and loss intelligibility. So to remove the noise required two techniques i.e. speech degradation and speech enhancement. The speech degradation technique involves the addition of gaussian noise to the original .wav format file with the help of MATLAB Function called `randn()`. Moreover, this process not only help us in making comparison between the clean file and the signal with the added gaussian noise, it also can be further viewed as that which filter in DSP (Digital signal processing) such as Chebysev Filter, Butterworth Filter etc. [2]

The speech enhancement technique enlightens upon the major use Speech Degradation technique i.e. removal of Gaussian noise from the original speech wave. In this technique firstly the degraded signal i.e. original signal mixed with Gaussian noise is first converted to the frequency domain with the help of Fast Fourier Transform(FFT) tool in MATLAB Programming. Then higher frequency noise components are then removed with the help of 3rd order Butterworth low pass filter. The butterworth filter is better as it has the capability to filter the Gaussian noise more closely and approximates an ideal low pass filter as the order, n, is increased. The resulting filtered signal was then scaled and plotted with the original noisy signal to compare the filtering result and the general representation.

#### C. Pitch Analysis

Pitch in terms of speech analysis can be defined as a technique which allows the ordering of sounds on a frequency-related scale. Pitch analysis helps us in identifying the state of speech of a person. The considered states are neutral, happy, sad. Therefore it is very important to understand the concept of pitch analysis. The calculation of the average pitch of the entire .wav format speech file that are recorded in the data base of different speakers would be done and found to have a certain value which can be used in voice recognition, where the differences in average pitch can be used to characterize a voice file.

#### D. Formant Analysis

In Formant Analysis technique, formant analysis will be performed on any of the .wav format speech file taken from the set of recorded .wav speech signal. With the help of MATLAB Programming we prepared a code for Formant Analysis. With the help of this code the first five formants that are present in .wav speech file are calculated, calculation of difference between the vector peak positions of these five formants, vector position of the peaks in the power spectral density were easy calculated and can be used to determine the speech file.

Voice or speaker recognition uses vocal characteristics to identify individuals using a pass-phrase. The matching strategy may typically employ approaches based on hidden Markov model, vector quantization, or dynamic time warping. A telephone or microphone can serve as a sensor, which makes it a relatively cheap and easily deployable technology. However, voice recognition can be affected by environmental factors such as background noise. This technology has been the focus of considerable efforts on the part of the telecommunications industry and the U.S. government’s intelligence community, which continue to work on improving reliability. The general block diagram of Speaker Verification is shown figure 2.

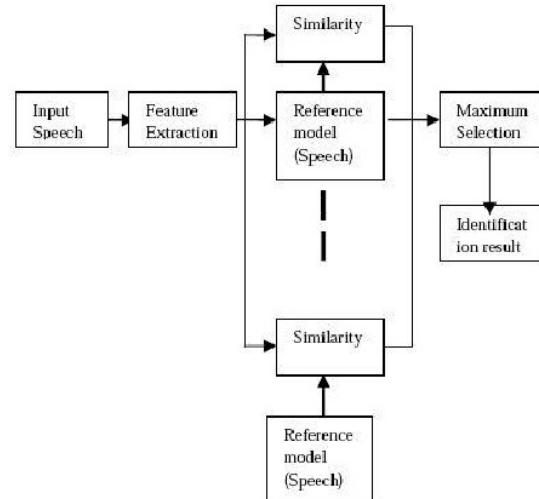


Figure 2: The general block diagram of Speaker Verification[5]

All speaker recognition systems have to serve two distinguishes phases. The first one is referred to the enrollment sessions or training phase while the second one is referred to as the operation sessions or testing phase. In the training phase, each registered speaker has to provide samples of their speech so that the system can build or train a reference model for that speaker. In case of speaker verification systems, in addition, a speaker-specific threshold is also computed from the training samples. During the testing phase (Figure 2), the input speech is matched with stored reference model and recognition decision is made [5].

### E. Speech Feature Extraction

The purpose of this module is to convert the speech waveform to some type of parametric representation (at a considerably lower information rate) for further analysis and processing. This is often referred as the signal processing front end. The speech signal is a slowly timed varying signal (it is called quasistationary). An example of speech signal is shown in Figure 3. When examined over a sufficiently short period of time (between 5 and 100 ms), its characteristics are fairly stationary. However, over long periods of time (on the order of 1/5 seconds or more) the signal characteristic change to reflect the different speech sounds being spoken. Therefore, short-time spectral analysis is the most common way to characterize the speech signal.

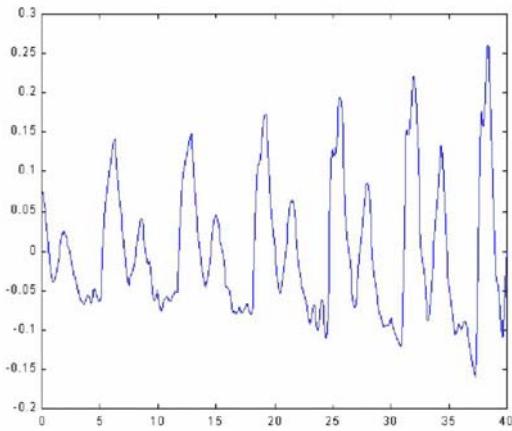


Figure 3. An example of speech signal[5]

## III. EXPERIMENT AND RESULTS

### A. Creating the Database

To recognize the uttered word of the speaker, a database was created to resemble the pronounced word. To create such database, we first recorded some numerals from one to five of five speakers(Speaker1, Speaker2, Speaker3, Speaker4, Speaker4, Speaker5).

### B. Training of the Voice

Speech recognition system is trained before use. We trained our speech samples at sampling frequency 8 kHz. The duration of the training was varied around 20s. After the training of the speech samples the system will separate the frames of speech signal with high energy and the speech signal with low energy.

### C. Experimental Testing

Our speech recognition system was a speaker dependent system. So it was dependent on the user's voice only. In the training of this system we created a database of five words. After the training of this system, a real time speech input was given to it through a good quality microphone. The system divided the real time speech sample into small segments of frames or continuous groups of samples.

The speech detection algorithm was developed by processing the prerecorded speech samples frame by frame within a simple loop. We divided the whole frame into the

segment of 160 samples and each of the samples was detected by the system. For the detection of each frame we used a combination of signal energy and a zero crossing rate. This calculation became very simple with the MATLAB mathematical and logical operators.

### D. Results

In this task we want to achieve a result recognizing voice in Speaker2 out of the five we recorded in the database.

In the command window we called "speechrecognition ('speak2.wav')" and figure 4, figure 5, figure 6, figure7 and figure 8 are observed.

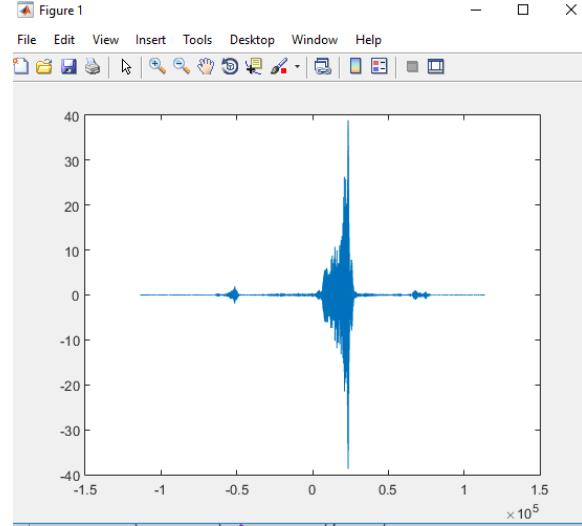


Figure 4: Showing speak1.wav as unrecognized voice of "Speak2.wav"

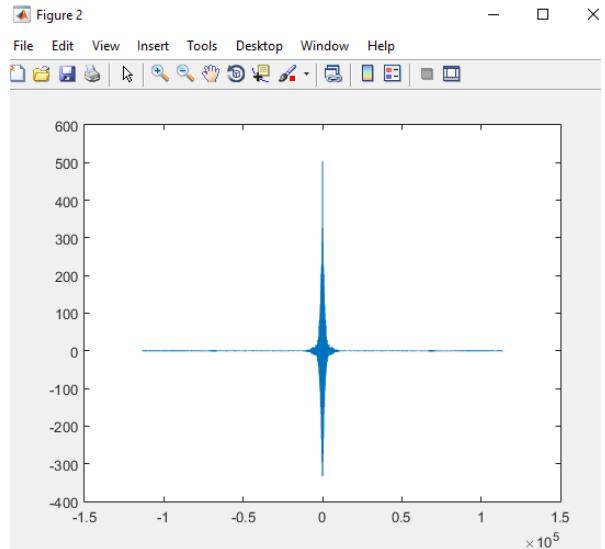


Figure 5: Showing the recognize voice of "Speak2.wav"

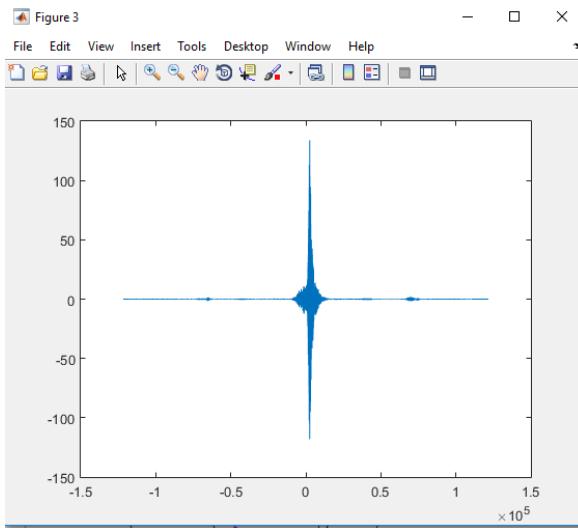


Figure 6: Showing speak3.wav as unrecognized voice of “Speak2.wav”

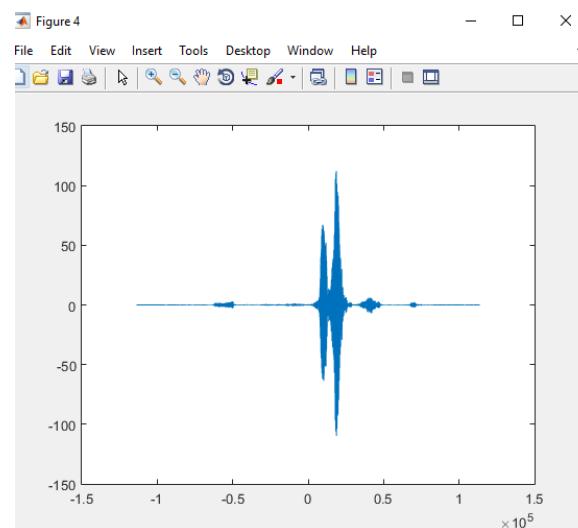


Figure 7: Showing speak4.wav as unrecognized voice of “Speak2.wav”

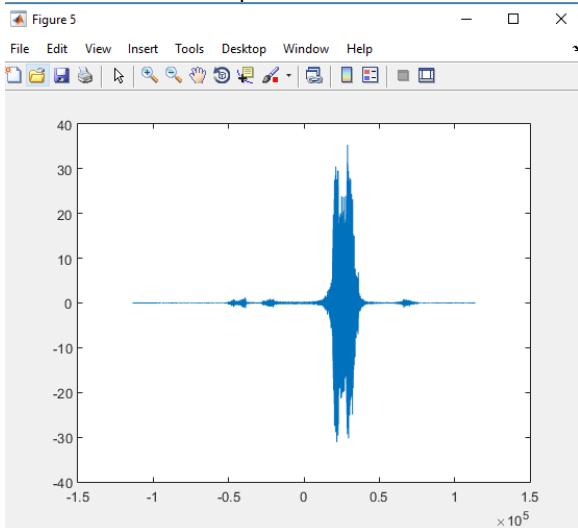


Figure 5: Showing speak5.wav as unrecognized voice of “Speak2.wav”

#### IV. CONCLUSION

In this project five words were collected and analyzed. Words were distinguished by energies associated with them. The system was able to separate the words according to their energies. Thus, the system was able to identify certain words of a particular speaker automatically by recognizing information included in individual speech waves. This paper brings closer the ever-increasing need to authenticate and identify individuals automatically. This paper also present MATLAB as a veritable tool in speech recognition. Further works in this area will consider using voice recognition in Automatic Teller Machines (ATM) for identification as Biometric technology promises an effective solution to the world security needs as it can accurately identify or verify individuals based upon their unique physical or behavioral characteristics.

#### V. REFERENCES

1. F. Monroe and A. D. Rubin, “Keystroke dynamics as a biometric for authentication, AT&T Labs-Research, Florham Park, NJ, USA, Future Generation Computer Systems 16 (2000) 351–359.
2. A. Jain, A. Singh, S. Raj Singh and S. Sikoria, “Speech Recognition Using MATLAB”, International Journal of Advances In Computer Science and Cloud Computing, ISSN: 2321-4058 Volume- 1, Issue- 2, Nov-2013
3. C. N. Solanki and M. G. Vayada, “Speech Recognition using Hidden Markov Model Algorithm”, JSRD - International Journal for Scientific Research & Development, Vol. 3, Issue 08, 2015, ISSN (online): 2321-0613
4. M.A.Anusuya , S.K.Katti “Speech Recognition by Machine - A Review” International journal of computer science and Information Security 2009, Vol. 6, No. 3, 2009
5. M. Saundade and P. Kurle, “Speech Recognition using Digital Signal Processing”, International Journal of Electronics, Communication & Soft Computing Science and Engineering ISSN: 2277-9477, Volume 2, Issue 6.