Chapter one
Introduction

1.1 Back ground

Speech recognition is apopular topic in today's life. The applications of speech recognition can be found everywhere, which make our life more effective.

For example the application in the mobile phone, instead of typing the name of the person who people want to call, people can just directly speak the name of the person to the mobile phone. And the mobile phone will automatically call that person. If people want to send some text messages to someone, people can also speak messages to mobile phone instead of typing. Speech recognition is a technology that people can control the system with their speech [1].

Instead of typing the keyboard or operating the buttons for the system, using speech to control system is more convenient. It can also reduce the cost of the industry production at the same time. Using the speech recognition system not only improves the efficiency of the daily life, but also makes people's life more diversified [2].

1.2Research Problem

A human can easily recognize a voice and a program cannot recognize this easily so this task is very difficult for the computer to recognize the different features of voice. Many complex problems arise while writing algorithm for different kind of voice recognition system this problem may arise because word at different times becomes difficult. Some of the factors that are vary or change time in the human speech that how fast word is being spoken.

1.3 Objectives

In general, the objective of this thesis is to investigate the algorithms of speech recognition by programming and simulating the designed system in MATLAB. At the same time, the other purpose of this thesis is to utilize the learnt knowledge to the real application.

1.4 literature review

The article by Bolt et al. (2010), summarize methods used for speaker identification and their validity.

Speech spectrograms portray short-term variations in intensity and frequency in graphical form. Thus they give much useful information about speech articulation is similar but not identical. Thus spectrograms of their speech will show similarities but also differences. However, there are also differences when the same speaker repeats a word.

Our auditory system exhibits an amazing ability to identify speakers, especially if the voice is well known to us, even in the presence of substantial interference. However, wrong identifications are within the experience of all of us. Careful studies, in fact, have shown that listening provides more dependable identification of the speaker than the examination of spectrograms of the same utterances dose (Stevens et al. 2013). Work is being done on the design and evaluation of methods for objective voice identification using completely automatic procedures but, at this time at least, they do not inspire confidence in the use of voiceprint for error-free speaker identification [3].

1.5 Thesis lay out

This thesis consists of three chapters, chapter one is the introduction, and chapter two is devoted for theory back. While the contribution is in chapter three.

Chapter two
The sound

2.1 Introduction

In this chapter we have defined sound as a longitudinal mechanical wave in an elastic medium, the source and properties of sound, our auditory system exhibits an amazing ability to identify speakers, Thus, The mechanical of speech and hearing.

Voice print can be thought of as the task of finding who is talking from a set of known voices of speakers.

2.2 The sound

The word sound is used to describe two different things [3]:

1-An auditory sensation in the ear

2-The disturbance in a medium that can cause this sensation

Sound is a longitudinal disturbance consisting of a succession of compression and rarefactions to which the ear is sensitive [4].

2.2.1 Type of wave motion

Energy can be transferred from one place to another by several means; energy propagation by means of disturbance in a medium instead of the motion of the medium itself is called wave motion.

1-A mechanical wave is a physical disturbance in an elastic medium

2-Electromagnetic wave is no physical medium is necessary for the transmission [5].

2.2.2 Type of waves

Waves are classified according to the motion of a local part of a medium with respect to the direction of wave propagation [5].

1-Transverse wave, the vibration of individual particles of the medium is perpendicular to the direction of wave propagation

2-Longitudinal wave, the vibration of the individual particles is parallel to the direction of wave propagation

2.2.3 Source of sound

Sound can be produced by a number of different processes [3].

1-Vibrating bodies, when a drumhead or piano sound board vibrates, it displaces the air next to it and causes the local air pressure to increase and decrease slightly, these pressure fluctuations travels out word as a sound wave

2- Changing air flow, when we speak or sing, our vocal folds (cards)alternately open and close so that the rate of air flow from our lungs increases and decreases, resulting in a sound wave.

2.3 properties of sound

2.3.1 Speed of sound

The speed of sound depends on the matter through which it travels. Sound wave travel faster through solids and liquids, the speed of sound through a material increases as the temperature of the material increases, in air at 0C°, soundtravels at a speed of 331m/s [5], frequency, wavelength, and velocity are related by [6]:

$$V = \lambda f$$

V = speed of sound
λ= wavelength
f = frequency

2.3.2 Loudness of sound

Loudness of sound is a physiological unit and cannot be measured directly, it is quite obvious that a sound described as loud by one person may not be described in similar terms by another [7].

2.3.3 Intensity of sound

Intensity of sound is a physical quantity and can be measured with suitable apparatus of great sensitivity [7], the power per unit area [8].

$$I = P/A$$

I = intensity
A = area
P = power

2.3.4 Diffraction of sound

Thespreading out of waves when they encounter a barrier or pass through a narrow opening [3].

2.3.5 Reflection of sound

Reflection on abrupt change in the direction of wave propagation at change of medium [3], we call the reflection of sound an echo, the fraction of sound energy reflected from a surface is large if the surface is rigid and smooth's [9].

2.3.6 Refraction of sound

A bending of waves when the speed of propagation changes either abruptly (at a change of medium) or gradually (sound waves in a wind of varying speed) [3].

2.3.7 Interference of sound

The interaction of two or more identical waves, which may support (constructive interference) or cancel (destructive interference) each other [3].

2.4 Power and intensity of sound

The energy of sound per unit time is the power of source (P)

The power per unit area at some distance r from the source is the intensity (3).

Consider a particle moving and osculating according to equation

$$x = A\sin(\omega t)$$

Where x is the displacement, A is the amplitude and w is the angular frequency. The velocity V is given by:

$$V = \omega A\cos(\omega t)$$

The force is given by:

$$F = -kx = -m\omega^2 x$$

Thus the potential energy takes the form:

$$V = -\int Fdx = m\omega^2 \int xdx = \tfrac{1}{2}m\omega^2 x^2 = \tfrac{1}{2}m\omega^2 A^2 sin^2\omega t$$

But the kinetic energy is given by:

$$T = \tfrac{1}{2}\, mv^2 = \tfrac{1}{2}\, m\omega^2 A^2\, cos^2\omega t$$

Thus the total energy takes the form:

$$E = \tfrac{1}{2}m\omega^2 A^2[sin^2\omega t + cos^2\omega t] = \tfrac{1}{2}m\omega^2 A^2$$

For small particle in a medium of density, one gets

$$\Delta E = \tfrac{1}{2}\Delta m\,(\omega A)^2 = \tfrac{1}{2}(\rho A\Delta x)(\omega A)^2$$

Thus the power is given by:

$$P = \frac{dE}{dt} = \tfrac{1}{2}\,\rho A\left(\frac{dx}{dt}\right)(\omega A)^2 = \tfrac{1}{2}\rho Av(\omega A)^2$$

The wave intensity takes the form:

$$I = \frac{P}{A} = \tfrac{1}{2}\rho v(\omega A)^2$$

Decibel (dB)

The decibel (dB) is a logarithmic unit used to measure the intensity of a sound, one decibel is one tenth of one bell, named in homer of Alexander Graham Bell

$$\beta = 10\log\left(\frac{I}{I_\circ}\right)$$

$I_\circ$ is the reference intensity, taken to be at the threshold ofhearing ($I_\circ = 1.00\times 10^{-12}\ w/m^2$)

I is the intensity in watts per square meter at the soundlevelβis measured in decibels (dB).

2.5 Physics of the ear and hearing

The human ear is complex organ that can detect a wide range of sounds, the ear can be divided into three parts [3], [7].

1-The outer ear,is a sound collector; it consists of the part that you can see and the ear canal. The visible parts collect sound waves and direct them into the ear canal

2-The middle ear, is a sound amplifier; it consists of the ear drum and three tiny bones (hammer, anvil, and the stirrup),sound waves that pass through the ear canal cause the eardrum to vibrate, theses vibrations are transmitted to the three small bones, which amplify the vibrations

3-The inner ear,cochlea is the main part of inner ear, which transforms pressure variations into neural impulse [7], the cochlea is filled with fluid and is lined with tiny hair-like cells. Vibrations of the stirrup bone are transmitted to the hair cells. The movement of the hair cells produces signals that travel to your brain, where they are interpreted as sound.

2.6 Sound wave

Sound waves are compressional waves traveling through a compressible medium; there are three categories of longitudinal mechanical waves that cover different ranges of frequency [6]:

1-Infrasonic waves are longitudinal waves with frequencies below the audible range.

2-Audible waves are sound waves that lie within the range of sensitivity of the human ear, typically, (20-20,000) Hz.

3-Ultrasonic waves are longitudinal waves having frequencies above the audible range.

2.7 Mechanics of speech

The human vocal organs,as wellas arepresentation of the main acoustical features. The lungs serve as both areserver of air and an energy source.

In speaking, as in exhaling,air is forced from the lungs through thelarynx into the three main cavities of the vocal tract; they pharynx and the nasal and oral cavities, from the nasal and oral cavities the air exits through the nose and mouth, respectively.

Air can be inhaled or exhaled with little generation of sound if desired. In orderto producespeech sounds, the flow of air is interrupted by the vocal cards or by constriction in the vocal tract (made with the tongue or lips). The sounds from the interrupted flow are appropriately modified by various cavities in the vocal tract andare eventually radiated as speech from the mouth and, in some cause, the nose [10].

2.8 Voice print

Humans have used body characteristic such as face,odor, voice, etc. Forthousandof year to recognize each other. The voice print the a sound recording of sound waves of the human voice and compares it to the recordings of several people to determine the voice of particular person.

Any human physiological and / or behavioral characteristic can be used as abiometric characteristic as long as it satisfies the following requirements [11]:

- **Universality**: each person should have the characteristic
- **Distinctiveness**: any two persons should be sufficiently different in terms of the characteristic
- **Permanence**: the characteristic should be sufficiently invariant (with respect to the matching criterion) over a period of time
- **Acceptability**:which indicates the extent to which people are willing to accept the use of a particular biometric identifier (characteristic) in their daily live
- **Circumvention**: which reflects how easily the system can be fooled using fraudulent methods
- **Performance**: which refers to the achievable recognition accuracyand speed, the resources required to achieve the desired recognition accuracy and speed, as well as the operational and environmental factors that affect the accuracy and speed

A practical biometric system should meet the specified recognition accuracy, speed, and resource requirement, be harmless to the users, be accepted by the intended population, and be sufficiently robust to various fraudulent methods and attacks to the system.

2.8.1 Biometric system

A biometric system is essentially a pattern recognition system that operates by acquired data, and comparing this feature set against the template set in the database.

Depending on the application context, a biometric system may operate either in verification mode or identification mode [12].

- **Identification mode**

- Match a person's biometrics against a database to figure out his identity by finding the closest match
- Commonly referred to as 1:N matching
- 'criminal watch-list' application scenarios

- **Verification mode**

- the person claims to be ‘Noha’, system must match and compare his / her biometrics with Noha’s stored biometrics
- If they match, then user is ‘verified’ or authenticated that he is indeed ‘Noha’.
- Access control application scenarios
- Typically referred as 1:1 matching

2.9 Applications of biometric systems

The applications of biometrics can be divided into the following three main groups:

- Commercial applications (have used knowledge-based systems) such as computer network login, electric data security, internet access, physical access control, medical records management, distance learning, etc.
- Government applications (have used token-based systems) such as nationalID card, correctional facility, social security, passport control, etc.
- Forensic applications (have relied on human experts to match biometric features) such as corpse identification, criminal investigation, terrorist identification, etc.

Chapter three
Programming steps and results

3.1 Introduction

In this chapter there is designed system (M file of MATLAB) for speech recognition. For running the system code at each time in MATLAB, MATLAB will ask the operator to record the speech signals for three times. The first two recordings were used as reference signals. The third time recording was used as the target signal.

3.2 Programming language MATLAB

Programming language is coded language used by programmers to write instructions that a computer can understand to do what the programmer (or the computer user) wants.

MATLAB is a language for technical competing that combines numeric computation, advanced graphics and visualization, and a high-level programming language. MATLAB is the natural environment for analysis, algorithm proto typing and application development.

3.3 Time domain to frequency domain: DFT and FFT

3.3.1 DFT

The DFT is an abbreviation of the Discrete Fourier Transform. So the DFT is just a type of Fourier Transform for the discrete-time instead of the continuous analog signal, which means transforming the signals from the time domain into the frequency domain.

3.3.2 FFT

The FFT is an abbreviation of the Fast Fourier Transform. Essentially, the FFT is still the DFT for transforming the discrete-time signal from time domain into its frequency domain. The difference is that the FFT is faster and more efficient on computation. And there are many ways to increase the computation efficiency of the DFT, but the most widely used FFT algorithm is the Radix-2FFT algorithm [13].

The main idea for Radix-2FFT is to separate the old data sequence into odd part and even part continuously to reduce approximately half of the original calculations.

3.4 Frequency analysis in MATLAB of speech recognition

3.4.1 Spectrum normalization

After doing DFT and FFT calculations, the investigated problems will be changed from discrete-time signal to frequency domain signal. The spectrum of the frequency domain signal is the whole integral or the summation of the all frequency components [14]. The equation of linear normalization is as below:

Y= (x-min value) / (max value-min value)

After normalization, the values of spectrum are set into interval [0, 1]. The normalization just changes the values' range of the spectrum, but not changes the shape or the information of the spectrum itself.

3.4.2 The cross-correlation algorithm

There is a substantial amount of data on the frequency of the voice fundamental in the speech of speakers who differ in age and sex [15]. For the same speaker, the different words also have the different frequency bands which are due to the different vibrations of the vocal cord. And the shapes of spectrums are also different.

The cross-correlation function method is really useful to estimate shift parameter [16].

3.5 MATLAB code

```
clear;
on=0;
off=0;
fs=16000;
duration=2;
fprintf('press any key to start %g second of
recording.on.\n',duration);
pause;
fprintf('recorging...\n');
r=wavrecord(2*fs,fs);
r=r-mean(r);
fprintf('press any key to start %g second of
recording.off.\n',duration);
pause;
fprintf('recorging...\n');
y=wavrecord(2*fs,fs);
y=y-mean(y);
fprintf('press any key to start %g second of
recording.v.\n',duration);
pause;
fprintf('recorging...\n');
```

```
v=wavrecord(2*fs,fs);
v=v-mean(v);
nfft=min(1023,length(r));
s=specgram(r,nfft,fs,hanning(511),380);
s2=specgram(y,nfft,fs,hanning(511),380);
s3=specgram(v,nfft,fs,hanning(511),380);
%take abs make it real and ease to plot
absolute=transpose(abs(s));
absolute2=transpose(abs(s2));
absolute3=transpose(abs(s3));
%get time-frequency related spectrum
a4=sum(absolute);
a5=sum(absolute2);
a6=sum(absolute3);
%the spectrom and also decrease the noise effect
a4_norm =(a4-min(a4))/(max(a4)-min(a4));
a5_norm =(a5-min(a5))/(max(a5)-min(a5));
a6_norm =(a6-min(a6))/(max(a6)-min(a6));
%transpose row to colume vector
f4=transpose(a4_norm);
f5=transpose(a5_norm);
f6=transpose(a6_norm);
%frequency spectrum:comper
[x3,lag3]=xcorr(f6,f4);
[mx3,indice3]=max(x3);
frequency_on_shift =lag3(indice3);
[x4,lag4]=xcorr(f6,f5);
[mx4,indice4]=max(x4);
frequency_off_shift =lag4(indice4);
%plotting
subplot(2,3,1),plot(abs(s));
subplot(2,3,2),plot(abs(s2));
subplot(2,3,3),plot(abs(s3));
subplot(2,3,4),plot(f4);
subplot(2,3,5),plot(f5);
subplot(2,3,6),plot(f6);
figure(2)
subplot(1,2,1),plot(x3);
title('xcorr of reference signal and target signal');
subplot(1,2,2),plot(x4);
title('xcorr of reference signal and target signal');
```

3.6 The simulation results

(1) The information of the first statistical simulation results for system is as following:

Reference signal: “on” and “off”.

Target signal: “on”.

Speaker: Noha for both reference signals and the target signal.

Environment around: almost no noise

The figure 1 is about frequency spectrums for three recorded signals.

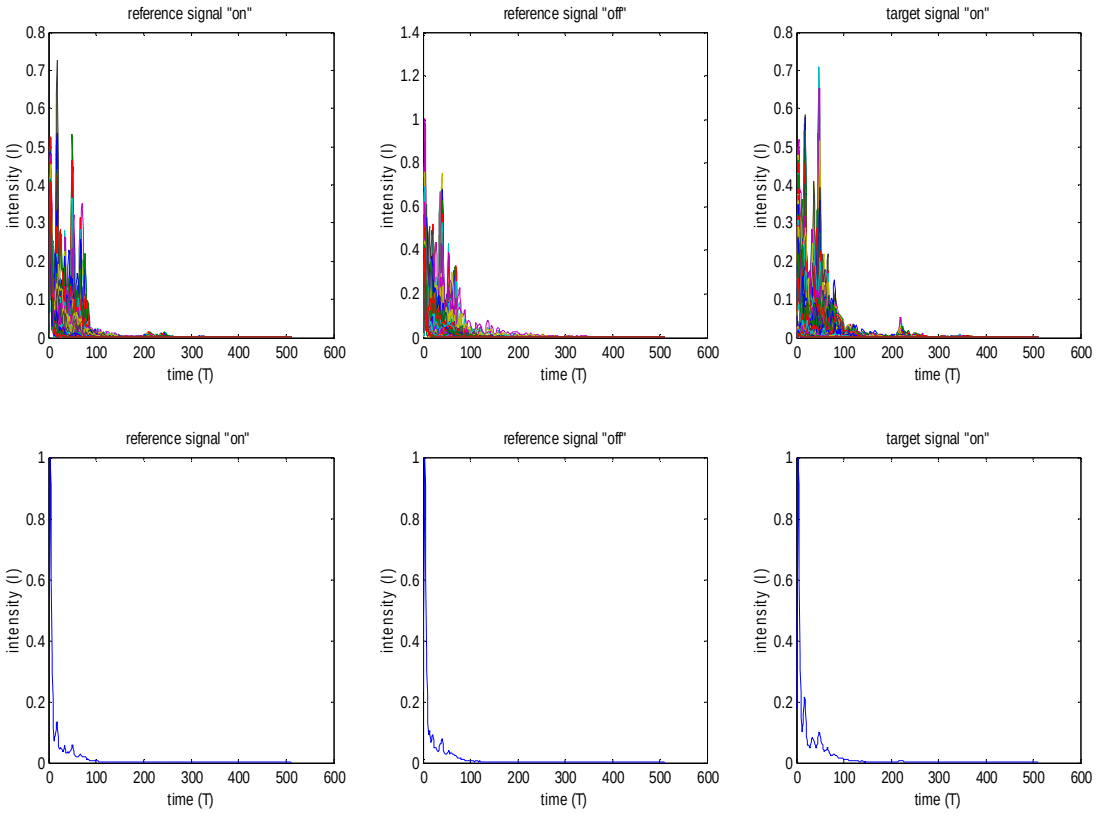

Figure 1: frequency spectrums for three speech signals: “on”, “off”, “on”.

Table 1

I_{max}, T for reference signals “on”, “off” and target signal “on”

Imax = max intensity

T= time period

Signal	Imax	T	Imax/T×10^{-3}
Reference signal of“on”	0.72	500	1.44
Reference signal of“off”	1	500	2
Target signal “on”	0.71	500	1.42

The figure of cross-correlations between the target signal “on” and the reference signals is as below (the reference signal of left plotting is “on”; the reference signal of the right plotting is “off”):

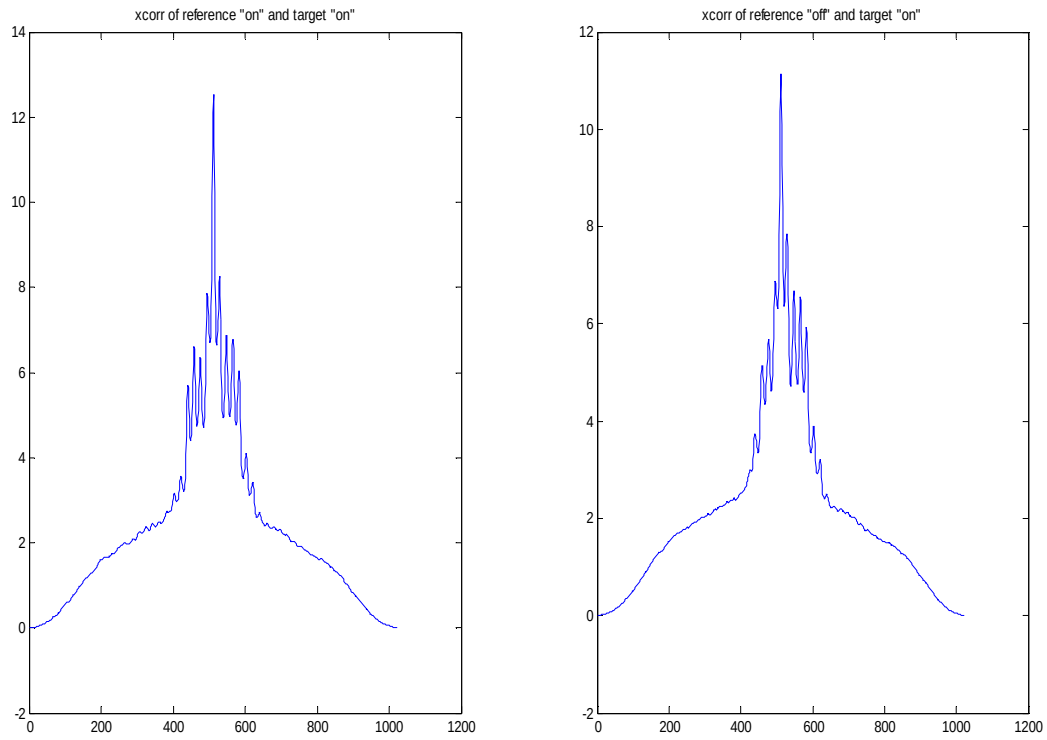

Figure 2: cross-correlations between the target signal “on” and reference signals

Table 2

Compare the height of beaks of cross-correlations between the target signal “on” and reference signals

Signal	Height of beaks				Mean of beaks	Result
	1	2	3	4		
Reference signal of “on”	2.5	4.3	6	7.8	5.15	Reference signal and target signal are close
Target signal “on”	3	5	7	8	5.75	
Reference signal of “off”	3.7	5	5.8	7	5.37	Reference and target signal are different
Target signal “on”	3	4	6	6.2	4.8	

(2) The information of the second statistical simulation results for system is as following:
Reference signal: “dog” and “cat”.
Target signal: “dog”.
Speaker: Noha for both reference signals and the target signal.
Environment around: almost no noise
The figure 3 is about frequency spectrums for three recorded signals are got by the same way as the figure1.

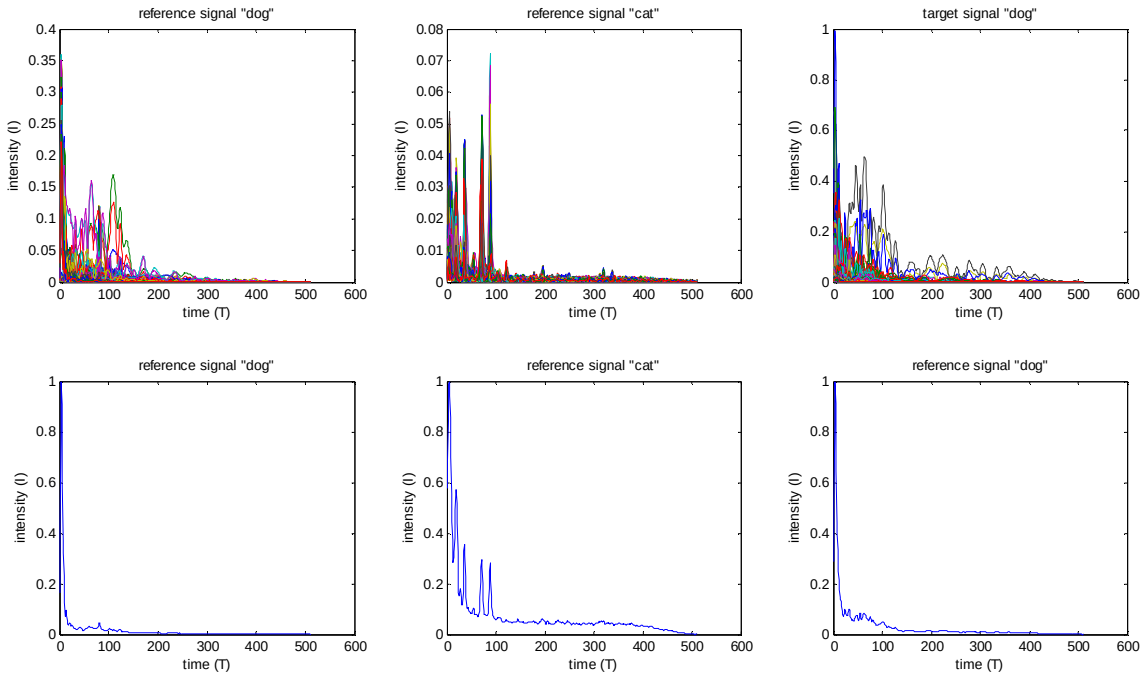

Figure3: frequency spectrums for three signals: “dog”, “cat”, and “dog”

Table 3

I_{max}, T for reference signals “dog”, “cat” and target signal “dog”

Imax = max intensity

T = time period

Signal	Imax	T	Imax/T×10^{-3}
Reference signal of “dog”	0.37	500	0.74
Reference signal of “cat”	0.07	500	0.14
Target signal “dog”	0.95	500	1.9

The figure of cross-correlations for the target signal “dog” with reference signals is as below (the reference signal of the left plotting is “dog”; the reference signal of the right plotting is “cat”):

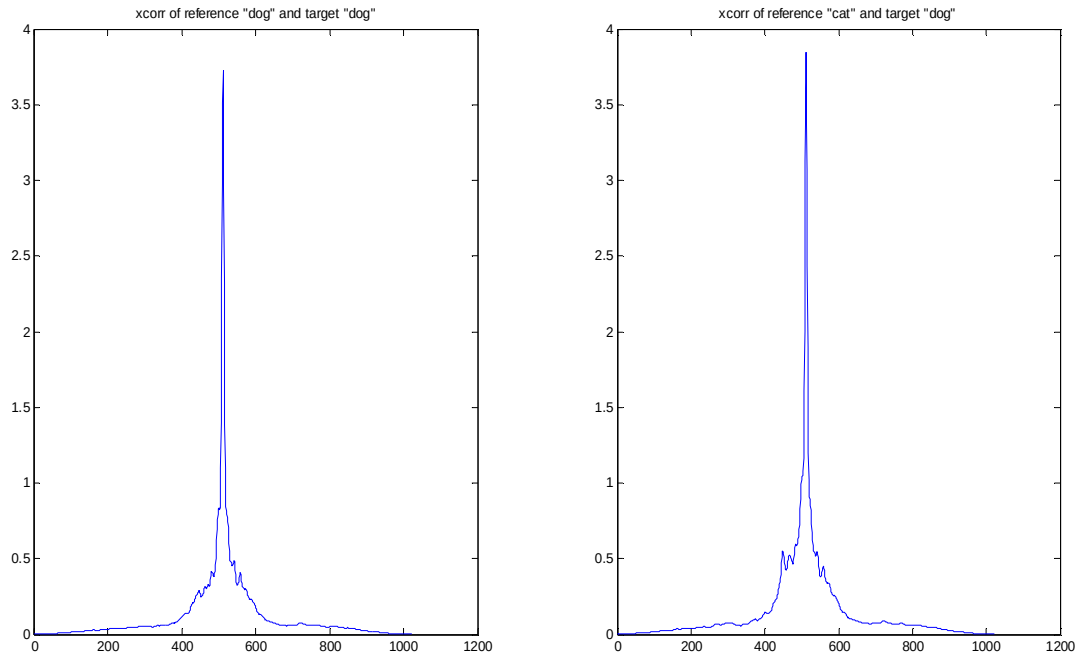

figure4: cross-correlations for the target signal “dog” with the reference signals

Table 4

Compare the height of beaks of cross-correlations for target signal “dog” with the reference signals

Signal	Height of beaks				Mean of beaks	Result
	1	2	3	4		
Reference signal of “dog”	0.4	0.5	0.8	2.8	1.125	Reference signal and target signal are the same
Target signal “dog”	2	2.4	3.5	5.6	1.125	
Reference signal of “cat”	0.6	0.5	1	3.3	1.35	Reference and target signal are close
Target signal “dog”	0.4	0.5	0.7	3.2	1.2	

(3) The information of the third statistical simulation results for system is as following:

Reference signal: “on” and “off”.

Target signal: “on”.
Speaker: Noha for both reference signals and the target signal.
Environment around: there is some noise

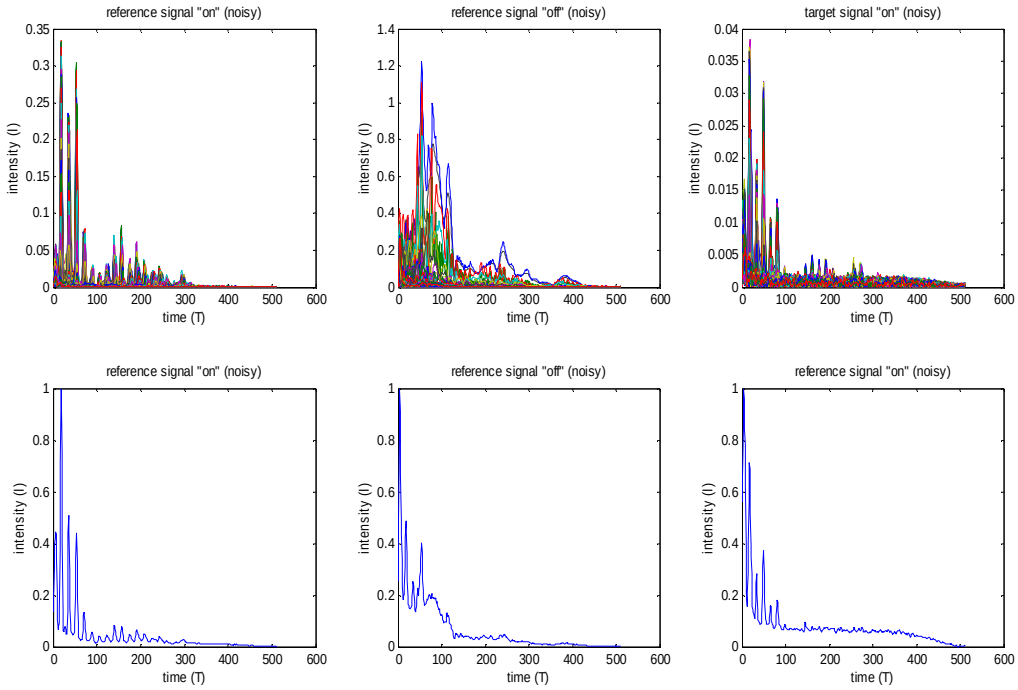

Figure5: reference “on” and “off” (noisy)

Table 5
I_{max}, T for reference signals “on”, “off” and target signal “on” (noisy)
Imax = max intensity
T = time period

Signal	Imax	T	Imax/T×10^{-3}
Reference signal of “dog”	0.34	500	0.68
Reference signal of “cat”	1.2	500	2.4
Target signal “dog”	0.038	500	0.076

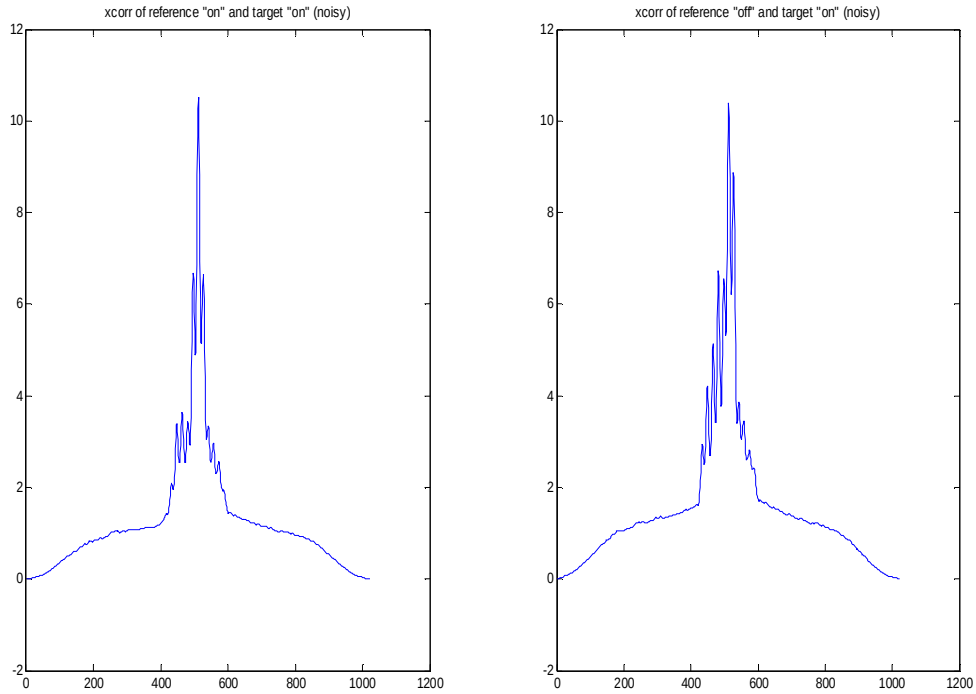

figure6: cross-correlations for target signal “on” with reference signals (noisy)

Table 6

Compare the height of beaks of cross-correlations for target signal “on” with reference signals (noisy)

Signal	Height of beaks				Mean of beaks	Result
	1	2	3	4		
Reference signal of “on”	2	3.4	3.5	6.4	3.825	Are the same
Target signal “on”	2	2.8	3.4	6.5	3.675	
Reference signal of “off”	2.4	4	5	7	4.6	Reference and target signal are different
Target signal “on”	2.3	3	8	9	5.575	

3.7 Discussion

Mat lab program was used to enable for voice recognition. Two reference signals for the words “on” and “off” were displayed and stored as reference signals. The spectra of the voice of “on” and “off” are in the form of signal relating intensity ‘I’ and time ‘T’.

The program was tested by target signal. In fig (1) and table (1) the target signal “on” is compared with the reference signals. It is found that the target signal “on” resemble the reference signal “on”. The comparison between reference signals and target signal “on” is also displayed in fig (2) and table (2). Again the target signal “on” resemble the reference one “on”.

Again the program was tested for two words “cat” and “dog” by displaying their reference signal in fig (3) and table (3). These signals are compared with the target signal “dog”. It is found again that the target signal resemble that of “dog”. The two reference signals were compared graphically in fig (4) their mean beak height I are compared in table (4). The comparison shows that the target signal “dog” resemble the reference one “dog”.

The reference signals “on” and “off” are displayed in fig (5) and (6) beside tables (5) and (6). And are compared with target signal “on”, but here in the presence of noise. The result shows again that the target signal “on” resemble that of reference one “on”.

3.8 Conclusions

It is clear that the Mat lab program for noise recognition can recognize the voice easily. This means that this program can be used for security purposes.

3.9 Recommendation

-Avoid the noise.

References

[1] R. KamelWassef, Physics for life, Dar Al-waffaa, Cairo, 1996.

[2] S.B.H.M Chandran, J vigneshwar, M Fazil Ahmed, E Balaji, GagatriKaur and SunishaSugunan, A survey on the impact of radiation emitted by the cell phone tower on human subjects, International Journal of life Science Biotechnology and Pharma Research, Vol. 1, 2012.

[3] Thomas D.Rossing, F.Richard moor and wheeler,The science of sound,3edition, Addison Wesly, USA, 2002.

[4]Holt, Rinettart and Winston, Modern physics,Holt, Rinettart and Winston,USA, 1990.

[5]PoulE.Tippens, Physics, seventhedition,McGraw-Hill Higher Education, London.EC4, 2007.

[6]Raymond A. serway, Physics for scientists and engineers with modern physics, 2edition, Saunders College Publishing,USA, 1986.

[7] D.H.FENDER, B.SC, PH.D, General physics and sound to Advanced and scholarship level, The English Universities Press LTD, England, 1965.

[8]Raymonal A. serway, Physics for scientists and engineers with modern physics, fourth edition, Saunders College Publishing, USA, 1996.

[9] Paul G. Hewitt, John Suchocki and Leslie A. Hewitt, Conceptual physical science, 3 edition, Pearson Addison Wesley, San Francis co, 2004.

[10]Paul W. Zitzewitz and James T. Murphy, physics principles and problems, Merrill Publishing Company Columbus, Ohio, USA, 1990.

[11] Anil k.jain, Arun Ross and salilprabhakar, An Introduction to Biometric Recognition, 2006.

[12]J.L.Wayman, Fundamentals of Biometric Authenticatian Technologies, InternationalJournal ofImage and Graphics, Vo1.1, No.1, PP.93-113, 2001.

[13]John G.Proakis, Dimitris G. Manolakis, Digital signal Processing, Principles, Algorithms, and Applications, 4th edition, Pearson Educatianinc, Upper saddle River, 1987.

[14]Luis Buera, Antonio Miguel, Eduardo lleida, Oscar saz, Alfonsooretega, Robust speech recognition with On-line Unsupervised Acoustic Feature Compensation, Communication Technologies Group (GTC), 13A, University of Zaragoza, Spain, 1989.

[15]HartmutTraunmuller, Anders Eriksson, The frequency range of the voice fundamental in the speech of male and female adults,Institutionen for lingvistik, StockholmsUniversitet, S-10691 Stockholm, Sweden, 2002.

[16]Jian Chen, JiwanGupta,Estimation of shift parameter of headway distributions using cross correlation function method, Department of Civil Engineering, The university of Toledo, 2007.