DOI: http://dx.doi.org/10.26483/ijarcs.v8i9.4910

ISSN No. 0976-5697

Volume 8, No. 9, November-December 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

AUTOMATIC SPEECH RECOGNITION APPROACH FOR DIVERSE VOICE COMMANDS

Mehak Mehraj
M. Tech Student
Department of Electronics and Communication,
Swami Devi Dyal Inst. of Engg. & Technology
Kurukshetra University, Kurukshetra

Ms. Arti goel
Assistant Professor
Department of Electronics and Communication,
Swami Devi Dyal Inst. of Engg. & Technology
Kurukshetra University, Kurukshetra

Muheet Ahmed Butt
Scientist,
PG Deptt. Of computer science,
University of Kashmir,Srinagar

Majid Zaman
Scientist,
Directorate of IT&SS.
University of Kashmir,Srinaga

Abstract: To underseek and resolve the algorithms of speech recognition is the aim of this work done. MATLAB is used to program and simulate the forth put algorithm . The two systems are created in this work. First system rely on the information of shape of cross-correlation plotting and second is used in finishing successfully the speech recognition by using the Weiner Filter. In the simulation, the spoken words are recorded using microphones. If the speaker is the same person for three time recordings, and the success for this approach is very high. Thus, the designed systems work accurately for the basic speech recognition.

Key words: (Speech recognition, Cross-correlation, Wiener Filter, Simulation)

1. INTRODUCTION

Speech recognition is considered as one of the great inventions using present day Computer Systems and other pervading devices. It has created new world of opportunities and contingency for software and hardware developers around the globe especially those building IVRs and other telephony applications. Building such a huge speech recognition applications has given rise to many internal and external confrontation .Rather pressing buttons of a computer or a smart pervasive device and envisaging output on a computer screen, the modern users must speak to the computer via a microphone, and this creates level of uncertainty in the input itself, as automatic speech recognition process using uncertainties or likelihood to arrive at certain speech recognition.These processes and methods have joined many strengths and weaknesses with the speech recognition process. The most obvious weakness is the uncertainty in the speech input process, namely the potential for misrecognition. It requires a substantial judging, effort and care in developing a piece of speech recognition software module, but still there are always instances when the application misrecognizes user speech input. This needs greater error handling mechanisms to put in the speech recognition module in comparison to other software applications. If the confidence score on a specific recognition process is low, it becomes important to confirm the user input speech. The system may have to ask users to repeat the input speech to the corresponding speech recognition applications so as to enhance the confidence score. Many a times user'sspeech will not be understood by the system reason being a noisy environment. If a speech engine returns low confidence values for the same user several times, it may be imperative to transfer that user to a human operator so the user can conduct his or her transaction.Speech recognition has become a house hold application nowadays. Speech recognition devices are equipped in modern electronic gadgets. Internet is flooded with audio data and software for speech detection and recognition. Speech makes it more convenient to operate electronic systems instead of typing with the keyboard or operating with buttons. Voice recognition system nowadays has numerous applications which requires interconnection such as automatic call processing, query-based information systems, weather reports etc[1]. With the speech recognition systems,the lives of human is getting better in modified manner. The dramatic progress in voice recognition technology has been seen by past decade , to such an range that high-performance algorithms as well as the systems have become approachable. The efficiency of the daily life raises as well as makes people's life more modified.

Speech recognition is the technology with the help of which a computer can associates the components of human speech[3]. Speech recognition is one of the many available biometric recognition schemes [4].The process begins by capturing the spoken utterance using a microphone and to end with the notable words being output bythe system. Speech is technically defined as a sequence of basic units called phonemes [5]. Automated Speech Recognition (ASR) systems executes conversion of analog speech signals received through microphones to the digital signals which are then segmented to regain phonemes.The ASR system

refers to the vocabulary and grammar rules to decode words or phrases using the phoneme sequence.

.

2. SPEECH SIGNAL REPRESENTATION

The speech signal has a feature that it not only gives the information regarding the words or message being spoken but also the identity of the speaker. This speaker identification is done by representingspeech signal in terms of certain features which are grouped into feature vectors that serve to decreasedimension and redundancy in the input to the speaker identification system, while retaining the speaker-specific information. However the irrelevant information with regards to speaker discrimination is a common problem for all feature sets,it is the topic of ongoing research which strives to determine feature sets of very less complicacythat can be applied to speaker identification[15]. The nature of these feature set depends on which part of a speech signal the features are expected to potray and thus the type ofinformation which is to be extracted. Thus due to this reason feature sets can be grouped as source basedfeatures or the system based features. The source is described as being the actualsound wave that is transmitted from the diaphragm through the glottis and so these feature are involved to determine the characteristics of the vocal cords, where this waveform is shaped. The fundamental frequency is the most feasible parameter that can be determined. The system characteristics can be extracted for the vocal tract, the nasal cavity and the lip radiation. These features model the filter characteristics of the vocal tract which can be derived from information contained in both voiced as well as unvoiced speech. The physiology of the speaker is reflected by the system features. For every feature extraction method, it is important to know that exactly what is being extracted to avoid defect of accuracy and ambiguity.When performing the signal processing analysis, the information of the DC level for the target signal is not that useful except the signal is applied to the real analog circuit, such as AD convertor, which has the requirement of the supplied voltage. When analyzing the signals in frequency domain, the DC level is not that useful. Sometimes the magnitude of the DC level in frequency domain will interfere the analysis when the target signal is most concentrated in the lower frequency band. In WSS condition for the stochastic process, the variance and mean value of the signal will not change as the time changing. So the author tries to reduce this effect by deducting of the mean value of the recorded signals. This will remove the zero frequency components for the DC level in the frequency spectrum[2].

A spectrogram is used which is a short-time Fourier transform that shows the energy of a signal as a function of positive time and frequency [8], thus allowing us to locate areas of energy in the speech signal. It only represents the amplitude of the speech signal, as no phase information is retained. phase information is not important for discrimination between speakers[7], so it can be omitted for making calculations simple, i.e., the magnitude of the spectrum of the speech signal is used.

3. SPEECH RECOGNITION PROCESS

Speech recognition process requires frequency analysis. The frequency analysis is carried out in MATLAB using the following processes.

i. Spectrum Normalization

Normalization maintain the measurement standard as comparing spectrums in different measurement standardscould be difficult when comparing the differences betweendifferent speech signals. Hence the method of normalization reduces the error when comparing thespectrums. When the normalization of the absolutevalues of FFT is performed, the next step in programming the speech recognition is observing the spectrums of the recorded signals. At last the algorithms are compared based on thedifferences between the test or target signal and the trainingsignals or reference signals [10].The error is reduced by normalization when comparing the spectrums, which is good for the speech recognition [11]. So before subjecting the spectrum differences for different words, the first step is of using the linear normalization to normalize the spectrum. The linear normalization is represented by equation given below:

$$y=(x\text{-}MinValue)/(MaxValue\text{-}MinValue)$$

After normalization, the values of the spectrum $|X(\omega)|$ are set into interval [0, 1]. The normalization only changes the values' range of the spectrum, but does not change shape or the information of the spectrum itself. So for spectrum comparison normalization is better. The change in spectrum by the linear normalization is shown below in example. Firstly, record a speech signal and do the FFT of the speech signal. After that take the absolute values of the FFT spectrum. The FFT spectrum without normalization is as below:

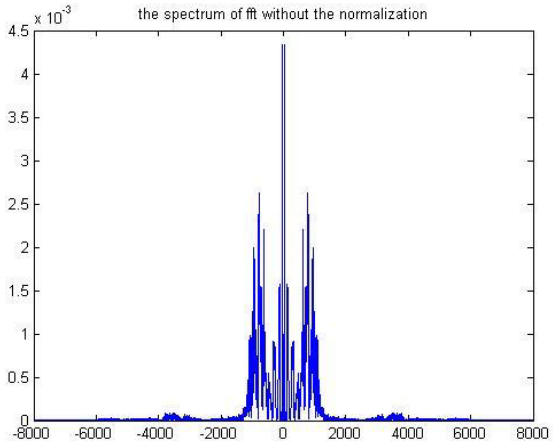

Figure 1: Absolute values of the FFT spectrum without normalization

Using linear normalization for normalizing the above spectrum, the normalized spectrum is as below:

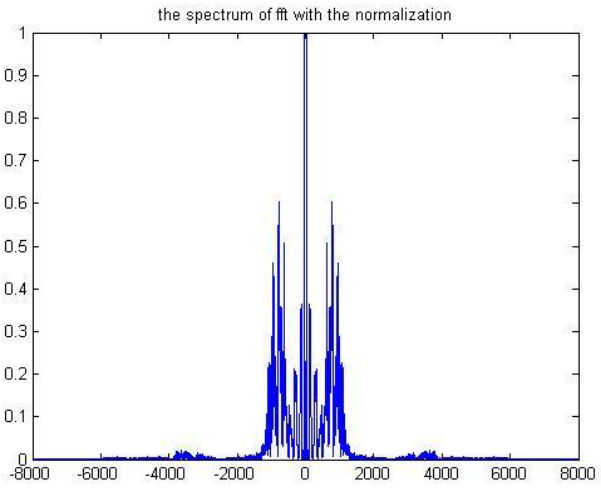

Figure 2: Absolute values of the FFT spectrum with normalization

From the Fig.1 and the Fig.2, the difference between two spectrums lies only in the interval of SpectrumX(ω) values, which is changed from [0, 4.5×10-3] to [0, 1]. Other information ofthe spectrum is not changed. After the normalization of the absolute values of FFT, the nextstep is observing spectrums of the three recordedspeech signals and finally finding the algorithms for comparing differences between the third recordedtarget and two recorded reference signals[9].

ii. The Cross-correlation Algorithm

There is a significant quantity of data on the frequency of the voice fundamental (F0) in the speech of speakers who differ in age and gender[12]. For the same loudspeaker system, the different words also have the different frequency bands which are referable to the different vibrations of the vocal cord.And the forms of thespectrums are also dissimilar. These are the foundations of this thesis for the speech recognition. In this thesis, to earn the speech recognition, there is a need to compare spectrums between the third recorded signal and the first two recorded reference signals. By checking which of two recorded reference signals better matches the third recorded signal, the system will make the judgment that which reference word is again read at the tertiary time. When thinking about the correlation of two signals, the first algorithm that will be considered is the cross-correlation of two signs.The cross-correlation function method is very useful to estimate shift parameter [13]. Here the shift parameter will be referred as frequency shift. The defining equation of the cross-correlation of two signals is as under:

$$r_{xy} = r(m) = \sum_{n=-\infty}^{\infty} x(n)y(n+m), m = 0, \pm1, \pm2, \pm3, ...$$

From the equivalence, the principal idea of the algorithm for the cross-correlation is about 3 steps:

Firstly, specify one of the two signals x (n) and switch the other signal y (n) left or right with some time units.

Secondly, multiply the value of x (n) with the shifted signal y (n+m) position by position.

At terminal, take the sum of all the multiplication results for x (n) · y (n+m). For instance, two sequence signals x (n) = [0 0 0 1 0], y (n) = [0 1 0 0 0], the lengths for both signals are N=5. Hence the cross-correlation for x (n) and y (n) is as the following figures shown:

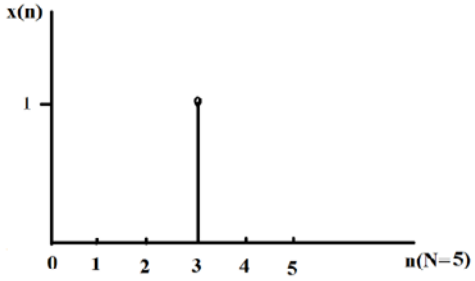

Figure 3: The signal sequence x(n)

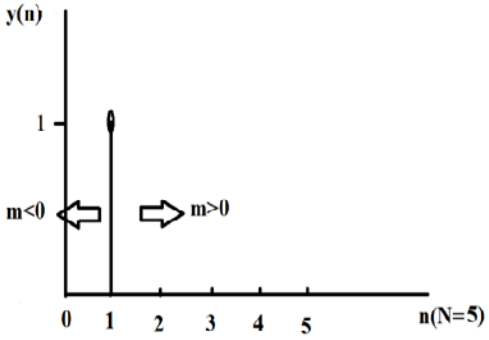

Figure 4: The signal sequence y (n) will shift left or right with m units

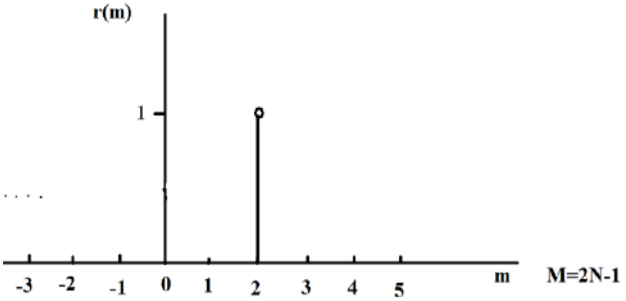

Figure 5: The results of the cross-correlation, summation of multiplications

As the case presented, in that respect is a discrete time shift about 2 time units between the signals x (n) and y (n). From Fig5, the cross-correlation r **(**m) has a non-zero result value, which is equal 1 at the position m=2. So the maxis of Fig.5 is no longer the time axis for the sign. It is the time-shift axis. Since the lengths of two signals x (n) and y (n) are both N=5, so the length of the time-shift axis are2N. When using MATLAB to execute the cross-correlation, the duration of the cross-correlation is still 2N. But in MATLAB, the plotting of the cross-correlation is from 0 to 2N-1, not from –N to +N anymore. Then the 0 time-shift point position will be shifted from 0 to N. Thus when two signals have no time shift, the maximum value of their cross-correlation will be at the position m=N in MATLAB, which is the center level position for the total duration of the cross-correlation.

iii. The Auto-correlation Algorithm

The autocorrelation can be treated as computing the cross-correlation of the signal and itself instead of two different signs. The auto-correlation is the algorithm to assess how the signal is self-correlated with itself. The equation for the auto-correlation is:

$$r_x(k) = r_{xx}(k) = \sum_{k=-\infty}^{\infty} x(n)x(n+k)$$

The figure below is the graph of plotting the autocorrelation of the frequency spectrum |*X*(☐)|

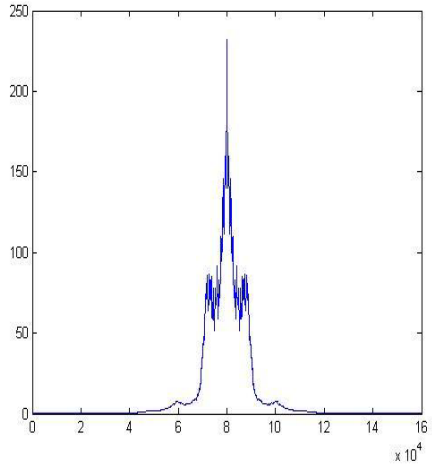

Figure 6: The autocorrelation for *X*(☐)

iv. The FIR Wiener Filter

The FIR wiener filter is shown below in fig 7

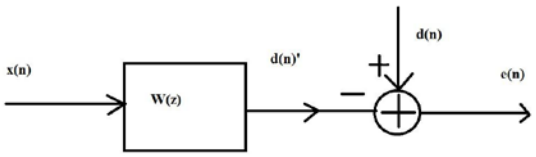

Figure 7: Wiener filter

The input signal of Wiener filter is x (n). Assume the filter coefficients are w (n). So the output d (n)' is the convolution of x (n) and w (n):

$$d(n)' = w(n) * x(n) = \sum_{l=0}^{p-1} w(l)x(n-l)$$

Then the estimation error is given as:

$$e(n) = d(n) - d(n)' = d(n) - \sum_{l=0}^{p-1} w(l)x(n-l)$$

The Wiener filter is used in choosing the suitable filter order and as well as in finding the filter coefficients with which the system can get the best estimation. In other words, with the proper coefficients the system can minimize the mean-square error:

$$\xi = E\left\{\left|e(n)\right|^2\right\} = E\left\{\left|d(n) - d(n)'\right|^2\right\}$$

In order to get the suitable filter coefficients we minimize the MSE, there is a sufficient method for doing this is to get the derivative of ξ to be zero with respect to w*(k).
Then we get:

$$E\left\{e(n)x^*(n-k)\right\} = 0 \quad , \text{k=0, 1...., p-1}$$

The above equation is known as *orthogonality principle* or *the projection theorem* [14].
After some proper rearrangement, the final equation becomes:

$$\sum_{l=0}^{p-1} w(l)r_x(k-l) = r_{dx} \quad ; \quad \text{k=0, 1... p-1}$$

This equation may be written in matrix form:

$$\begin{bmatrix} r_x(0) & r_x^*(1) & \cdots & r_x^*(p-1) \\ r_x(1) & r_x(0) & \cdots & r_x^*(p-2) \\ r_x(2) & r_x(1) & \cdots & r_x^*(p-3) \\ \vdots & \vdots & & \vdots \\ r_x(p-1) & r_x(p-2) & \cdots & r_x^*(0) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \\ w(2) \\ \vdots \\ w(p-1) \end{bmatrix} = \begin{bmatrix} r_{dx}(0) \\ r_{dx}(1) \\ r_{dx}(2) \\ \vdots \\ r_{dx}(p-1) \end{bmatrix}$$

The matrix equation is actually Wiener-Hopf equation [6] of:

$$R_x w = r_{dx}$$

From the above equation, the input signal x (n) and the desired signal d (n) are the only things that need to know.Then using x (n) and d(n) finds the cross-correlation r_{dx}. At the same time, using x (n) gives the auto-correlation $r_x(n)$ and this $r_x(n)$ forms the matrix R_x in MATLAB. When having theR_x and r_{dx}, the filter coefficients can be obtained. Using the filter coefficients the minimum mean square-error ξ can be obtained. The minimum mean square-error is:

$$\xi_{\min} = E\left\{e(n)d^*(n)\right\} = E\left\{\left[d(n) - \sum_{l=0}^{p-1} w(l)x(n-l)\right]d^*(n)\right\} = r_d(0) - \sum_{l=0}^{p-1} w(l)r_{dx}^*(l)$$

4. LITERATURE SURVEY

The Literature review on speech recognition systems orders consideration towards the finding of Alexander Graham Bell regarding the methodused for transforming sound waves into electrical impulses and the first speech recognition system matured by Davis et al. [6] for finding telephone superiority digits spoken at normal speech rate. This attempt for (ASR) automatic speech recognition was mainly centered on the abstract structure of an electronic circuit for revealing ten digits of telephone superiority. To obtain a 2-D plot of formant 1 vs. formant 2 , spoken words were inspected. For finding the greatest correlation coefficient among a set of novel incoming data for pattern matching a circuit was developed. These features are grouped into feature vectors whose goalis decreasing redundancy as well as dimensionality in the input to the speaker recognition system.An indication circuit was invented to display the spoken digit that was already discovered. The proposed way lays stress on acknowledging speech sounds andproviding suitable labels to these sounds. In last five decades various approaches and types of speech recognition systems came into state of being. This evolution has led to a noticable impact on the growtht of speech recognition systems for various languages worldwide. The exact nature of the feature set relyon which part of a speech signal the features are expected to potray and thus what type of information is to be sunder out. In process of conversion of speech to text , the output of the system shows the text which is used to apprehend the speech. Automatic speech recognition system has been created using language whichis a portion of total around 7300 existing languages which are Hindi, English, Tamil, Bengali, Russian, Japanese, Portuguese, Sinhala, Chinese, Malayalam, Vietnamese, Spanish, Arabic, Filipino, Hindi are well-known among them. Maximum work for

recognition is done for English language. Since 1930s, a simple speech machine that answers to a limited small set of words was invented. This proposed machine is capable to take actions on spoken words and create the speech. From that time, it has become popular area of research for invention of speech recognition system. The best example for this is done by Olson and Belarin 1950 in RCA Laboratories who build a system to identify 10 syllables of a single talker (Olson et al., 1956) and at MIT Lincoln Lab, Forgie and Forgie built a speaker-independent 10-vowel recognizer (Forgie et al., 1956). It is continued by the middle of 70's. The new system of speech recognition depends on LPC methods. These were forthput by Itakura, Rabiner and Levinson (Itakura 1975; Rabiner et al., 1979) and others. This research bringsmain benefits where research shift the methodology from the more spontaneous template-based approach towards a more accurate statistical modeling outline (Juang et al., 2004) in 1980s.

5. PROPOSED METHODOLOGY

In this work, two designed systems are used for speech recognition. The learning in the theory part of this thesis were used by these two systems,which has been introduced at an earlier time. The two designed systems were tested by the author and her friends. For running the system codes at each time in MATLAB, MATLAB will ask the operator to record the speech signals for three times. The reference signalsconsists of the first two recordings and the third time recording is used as the target signal.

Algorithm for Design System 1:

1. Set the sampling frequency16 kHzafter assigning the variables.
2. Get returned matrix signals after processing the recorded signal.
3. Get the frequency spectrum by swapping the input signal.
4. Normalize the signal by process of Linear Normalization,whose equation is given as:

$$y=(x\text{-MinValue})/(\text{MaxValue-MinValue})$$

5. Execute the cross-correlation of the targeted signal with the first two reference signals separately.
6. Check the frequency shift of the cross-correlations.
7. Do the comparison by the symmetric property for the cross-correlations of the matched signals. The cross correlation of two signals is given by:

$$r_{xy} = r(m) = \sum_{n=-\infty}^{\infty} x(n)y(n+m), m = 0, \pm 1, \pm 2, \pm 3, \ldots.$$

Algorithm for Design System 2:

1. Assign the variables and set the sampling frequency equal to 16KHz .
2. Record 3 voice signals. Make the first two recordings as the reference signals and the third recording as the target voice signal.
3. Get returned matrix signals by processing the recorded signal.
4. Get the frequency spectrum by interchanging the input signal.
5. Use the linear normalization for normalizing the frequency spectrum.
6. Compute the auto-correlations of 3 signals:
7. Using wiener filter mode compute the filter coefficient.
8. Compute the minimum mean square-error for each reference signal which is given by

$$\xi_{\min} = E\left\{e(n)d^*(n)\right\} = E\left\{\left[d(n) - \sum_{l=0}^{p-1} w(l)x(n-l)\right]d^*(n)\right\} = r_d(0) - \sum_{l=0}^{p-1} w(l)r_{da}^*(l)$$

.

9. Smaller the minimum mean square-errors ,the superior is the estimation value.

6. RESULTS

The first two recordings in the process of speech recognition are used as reference signals. The third recording is used as the target signal for which MATALB should give the judgment. In the following results, the author uses "reference signals" to stand for the first two recordings and uses "target signal" to stand for the third recording. The words in the quotes stand for the contents of recordings. The author tried to test designed systems for both easily recognized words and difficultly recognized words. "From time 1 to time 10, 'on'" in the following of the thesis means the operator simulated 10 times and the third recording word is "on" in the first 10 times simulations. Both the contents of the reference words and the target word are known, the author wants to test if the judgment that is given by MATLAB is correct as we know. The statistical simulation results will be put in tables and will also be plotted. In this Simulation Result part, only the plotted results will be shown in the following content.

The information of the first statistical simulation results for system 1 :

Reference signals: "on" and "off":

Target signal: From time 1 to time 10, "ON".

From time 11 to time 20, "OFF".

Speaker: Speaker 1 for both reference signals and the target signal.

umbworld around: 'ALMOST NO NOISE'

Frequency spectrums for three recorded signals is potrayed in figure 8, but the axis is not the real frequency axis since the figure is got by STFT.

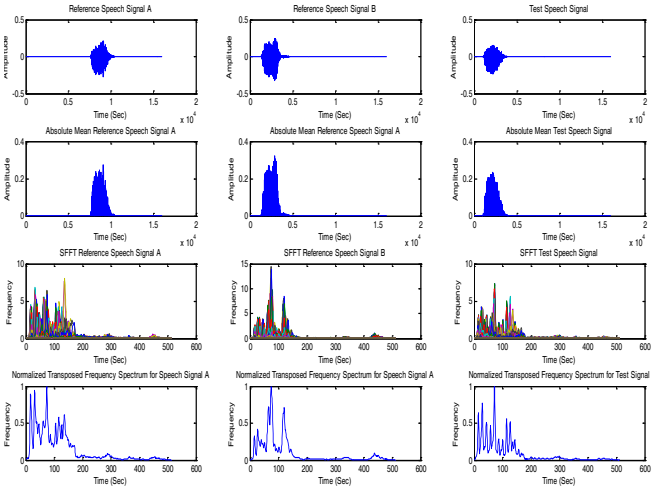

Figure 8: Frequency spectrums for three speech signals: "on", "off", "on"

.

The figure 9 shows the cross-correlations between the target signal “on” and the reference signals wherethe reference signal of the left plotting is “on”; the reference signal of the right plotting is “off”:

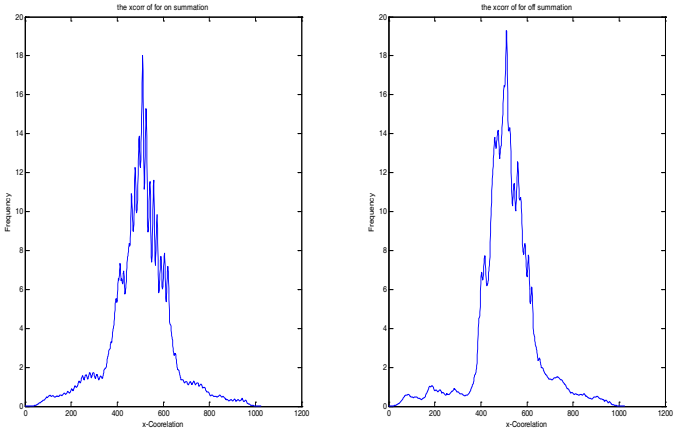

Figure 9: Cross-correlations between the target signal “on” and reference signals

There is no large difference between two graphs as potrayed in figure 9 above, since the pronunciations of “on” and “off” are close.

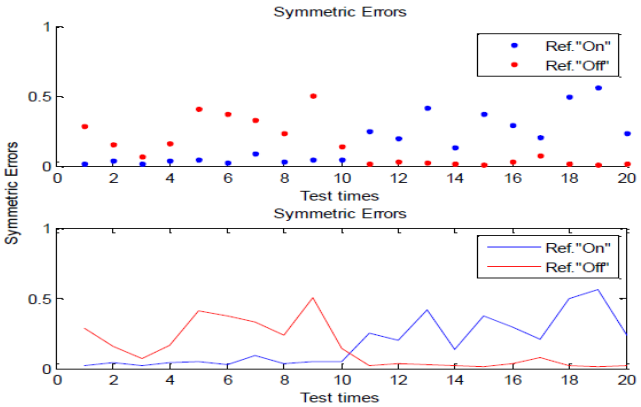

Figure 10: Symmetric errors in 20 times simulations for reference “on” and “off”

In Fig. 10, when the reference speech word is “on”, the simulated result is shown by blue curve. The red curve is the simulated result when the reference speech word is “off”. As information given at the start, the target speech word is “on” in the first 10 times simulations and the target speech word is “off” in the second 10 times simulations.

From Fig. 10, it is shown that the reference signal “on” curve has lower value for the first 10 times and for the second 10 times the reference “off” curve has lower value. The results have potrayed that the symmetric errors are smaller when the reference speech signal and the target speech signal are matched. The judgments are totally correct.

The information of the second statistical simulation results for system 1 is as following:

Reference signals: “Door” and “Key”:

Target signal: From time 1 to time 10, “Door”.

From time 11 to time 20, “Key”.

Speaker: Speaker 1 for both reference signals and the target signal.

Umbworld around: ALMOST NO NOISE.

The frequency spectrums for three recorded signals potrayed in figure 11 is got by the same way as in the figure8.

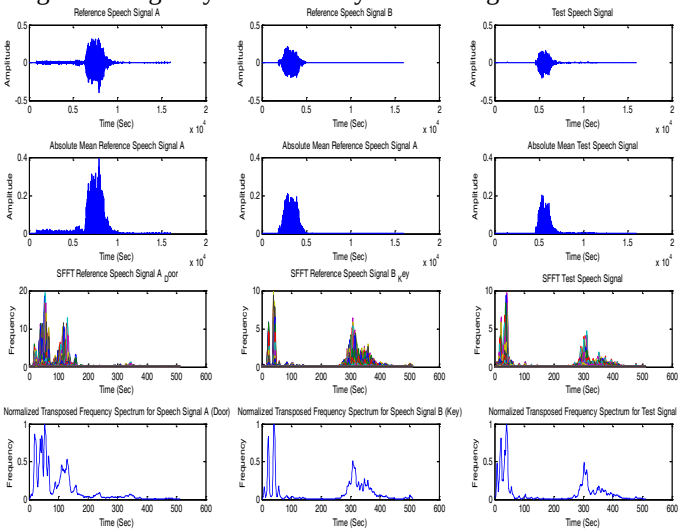

Figure 11: Frequency spectrums for three signals: “Door”, “Key”, and “Door”

The figure of cross-correlations for the target signal “Door” with reference signals ‘key’is as below.The reference signal of the left plotting is “Door”; and the reference signal of the right plotting is “Key”):

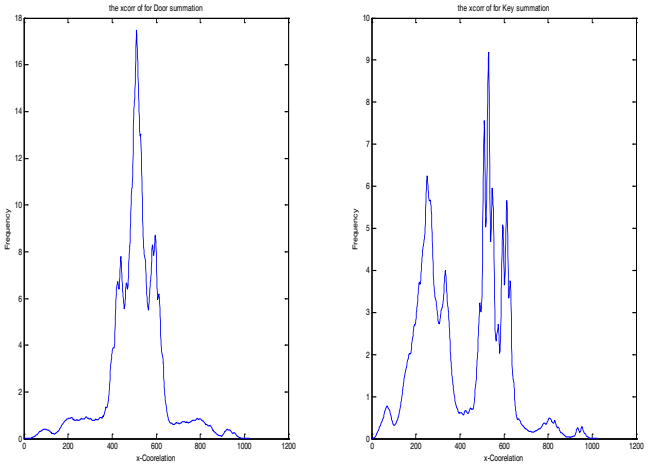

Figure 12: Cross-correlations for the target signal “Door” with reference signals

There is huge difference between two graphs as shown above in figure12. Since the pronunciations of word “Door” and “Key” are different. As introduced in theory part, the better matched signals have better symmetric property of the cross-correlation. The Figure.12 proved this point.

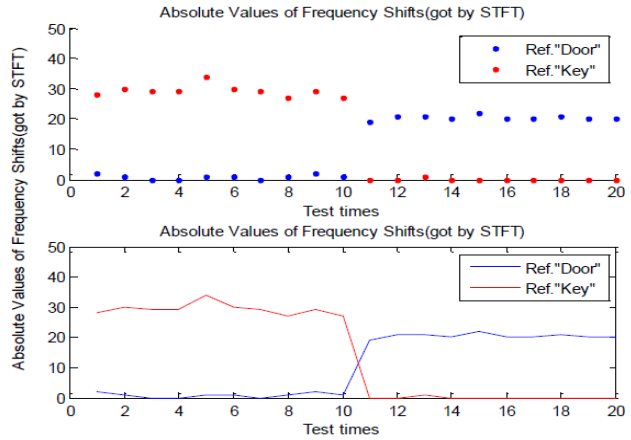

Figure 13: Frequency shits in 20 times simulations for reference “Door” and “Key”

From Fig.13, we can see that there are huge differences in the frequency shifts. So the designed system will directly give the judgments according to the frequency shifts.

3. The information of the third statistical simulation results for system 1 is as following:

Reference signals: “on” and “off”:

Target signal: From time 1 to time 10, “on”.

From time 11 to time 20, “off”.

Speaker: Speaker 2 for both reference signals and the target signal.

Umbworld around: appearance of some noise at sometimes

Since “on” and “off” have small frequency shift difference, so the designed system will only give the result with symmetric errors. These symmetric errors in 20 time simulations for reference ON and OFF are shown in figure 14 .

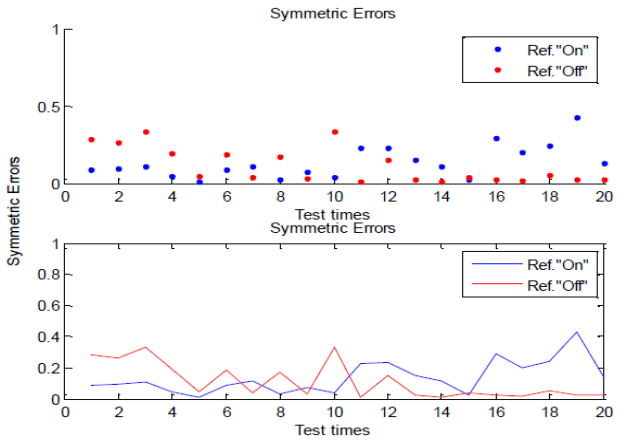

Figure 14: Symmetric errors in 20 times simulations for reference “on” and “off” (noisy)

As shown in Fig.14, the blue curve is simulated result when the reference speech word is “on”. The red curve is the simulated result when the reference speech word is “off”. As the information given at the outset, the target speech word is “on” in the first 10 times simulations and the target speech word is “off” in the second 10 times simulations.From Fig. 14, it is shown that in the first 10 times simulations the reference “on” curve has a lower value and in the second 10 times the reference “off” curve has a lower value. The outcomes have indicated that when the reference speech signal and the target speech signal are fitted, the symmetric errors are smaller. The assessments are completely right.

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

Reference signals: “Door” and “Key”:

Target signal: From time 1 to time 10, “Door”.

From time 11 to time 20, “Key”.

Speaker: Speaker 2 for both reference signals and the target signal.

umbworld around: existence of some noise sometimes .

The plotted simulation result is as below:

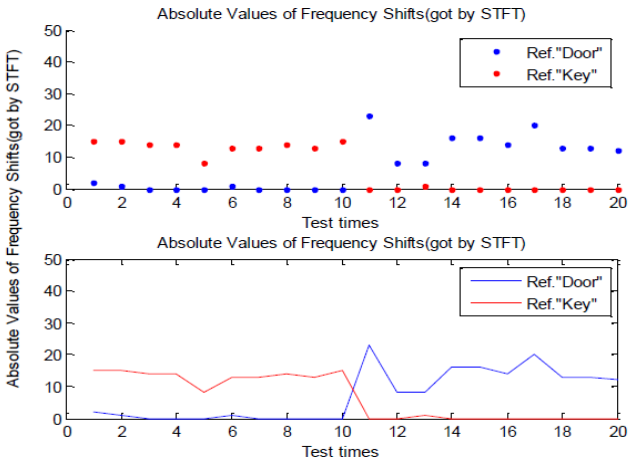

Figure 15: Frequency shits in 20 times simulations for reference “Door” and “Key” (noisy)

Table 1 points out the simulation results for reference signals “Door” and “Key” as the information given at the beginning of this section.

Table 1: Simulation results for speech words “On”, “Off”, “Door” and “Key”

Test times	frequency_on_shift	frequency_off_shift	Error1	Error2	**Final judgments**
1	2	8	No need	No need	on
2	7	8	0.2055	0.4324	on
3	8	9	0.2578	0.2573	off
4	9	17	No need	No need	on
5	8	9	0.2304	0.3640	on
6	0	0	0.3268	0.6311	on
7	0	0	0.3193	0.3210	on
8	0	0	2.2153	0.9354	off
9	0	0	0.4603	0.1481	off
10	0	0	0.1189	0.0741	off
11	8	22	No need	No need	Door
12	8	0	No need	No need	Key
13	8	25	No need	No need	Door
14	8	24	No need	No need	Door
15	8	24	No need	No need	Door
16	-15	0	No need	No need	Key
17	-15	0	No need	No need	Key
18	-14	0	No need	No need	Key
19	-14	0	No need	No need	Key
20	-15	0	No need	No need	Key
Total successful probability(total in 20 times)			**80%**		

Since the simulation results are not good as they were expected. So only the table results are shown here.

7.CONCLUSION

For indefinite conclusions, the noise easily diverts the designed systems for speech recognition, which can be observedfrom Table 1. For the designed system 1, the better

matched signals have the better symmetric property of their cross-correlation. For the designed system 2, if the reference signal is the sameas the target signal, there will be smaller errors in using this reference signal to model the target signal. This outcome can be demonstrated by all the assumed results for the designed system 2. When two of reference signals and the target signal are recorded by the same person, two systems work well for distinguishing different words, no matter where this person is from. But if the reference signals and the target signal are recorded by the different people, the execution of both systems is not well. So in order to improve the performance of designed systems to make it work better, we need to increase the exemption of system against noise and to find the common characteristics of the speech for the different people. Contrarily, the effect of input noise can be minimized by designing some analog and digital filters for processing the input signals which can also be used to form the large data base of the speech signals for different words. Studying more progressive algorithms for signal modeling can stipulate a lot of help to actualize the better speech recognition.

8. REFERENCES

[1] Lawrence R. Rabiner & B.H. Juang, Automatic Speech Recognition – A Brief History of the Technology Development, 2004

[2] M. Vikas, Deepak “Speech Recognition using FIR Wiener Filter”, International Journal of Application or Innovation in Engineering & management (IJAIEM),pp.204-20,2013.

[3] Christine Englund, Speech recognition in the JAS 39 Gripen aircraft - adaptation to speech at different G-loads.

[4] ” The Importance of a Biometric Authentication System”by Sushil Phadke , The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), Vol. 1, No. 4, September-October 2013.

[5] Speech Recognition Technology & Patent Landscape, iRunway 2015.

[6] Biddulph, R., Balashek and Davis, K., , S., “Automatic Recognition of Spoken Digit,” J. Acoust. Soc. Am. 24: Nov 1952, p. 637.

[7] S. Furui, .Digital speech processing,. Synthesis and Recognition, New York, Marcel Dekker, 2001.

[8] Auditory Toolbox ,Malcolm Slaney,.Version 2, Technical report, Interval Research Corporation, 1998.

[9] Dimitris G. Manolakis, John G. Proakis Digital Signal Processing, Principles, Algorithms, and Applications, 4th edition,Pearson Education inc., Upper Saddle River.

[10] Sons ,Inc ,John Wiley,,. Statistical Signal Processing And Modeling, Monson H Hayes,Georigia Institute of Technology.

[11] Eduardo Lleida, Luis Buera, Antonio Miguel ,Oscar Saz, Alfonso Oretega, “Robust Speech Recognition with On-line Unsupervised Acoustic Feature Compensation”, Communication Technologies Group (GTC),13A, University of Zaragoza, Spain.

[12] Anders Eriksson ,Hartmut Traunmüller, “The frequency range of the voice fundamental in the speech of female and male adults”, Institution enförlingvistik, Stockholmsuniversitet, S-106 91 Stockholm, Sweden.

[13] JiwanGupta, Jian Chen,,“Estimation of shift parameter of headway distributions using crosscorrelation function method”, Department of Civil Engineering, The University of Toledo.

[14] Buera, Luis, et al. "Cepstral vector normalization based on stereo data for robust speech recognition."*Audio, Speech, and Language Processing, IEEE Transactions on* 15.3 (2007),pp, 1098-1113,2007

[15] J. L. Gauvain; L. Lamel; H. Schwenk; G. Adda; L. Chen; F. Lefevre Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conferenceon,2003.