Security Technology for Realistic Measurements with a Distinction on the Speech Recognition in Arabic and English Language Using LMS, Spectral subtraction and A/D Conversions Techniques

IOP Conf. Series: Materials Science and Engineering **518** (2019) 042023 doi:10.1088/1757-899X/518/4/042023

Security Technology for Realistic Measurements with a Distinction on the Speech Recognition in Arabic and English Language Using LMS, Spectral subtraction and A/D Conversions Techniques

Ismail Abduljabbar Hasan[1], Muhanad A. Ahmed[1] and Hassan Hayder Dawood[1]
[1]Department of Electrical Technology, Institute of Technology-Baghdad, Middle Technical University, Baghdad, Iraq

Abstract. The new human sound recognition technology is obtained by using an algorithm which filters and separates frequencies from each human speech and makes recognitions on an individual speech and then can enter to the system using digital codes for username and again voice recognition until the system allows the real voice. This technique is superior to the means of using the eye-print and fingerprint for the possibility of the presence of diabetes in that person or the eye injury to a disease and thus blocks the system.

1. Introduction

Voice or speech recognition is the possibility of the system or program to get interpretation and then dictation, or understanding and executing commands expressed using speech. In computers, analog audio signal must be converted to digital. This needs a conversion from peer to digital (Analog to Digital Converter ADC) after filtering by least mean square filter (LMS). Decoding such a signal it must have digital database, syllables, words or vocabulary and a quick way to compare these data with references [1].

Speech models are stored and loaded in memory while the program is running, then the computer checks the comparison of these stored styles against the output of the A / D adapter. Both voice and language models are important for recent statistics speech recognition algorithms (SSRA) [2]. These technologies didn't take care of noises and frequencies. In this research we will look at how to obtain a pure signal and consider the noise and frequency of the signals received.

Analogue signals can be converted to digital signal by ADC and then obtained the source signal this digital and filtered signal can be distinguished from each human speech and then stored as a hardware code to open or close the system.

2. Least Mean Squares (LMS) Filter Algorithm

LMS algorithms are a class of adaptation filter used to imitate a required filter by finding the filter parameters that relate to producing the least mean square of the error signal (difference between desired and actual signal) [3]. It is a random sloping origin method in that the filter is only adapted based on the error at the current time [4].As shown in figure (1) the speech signal ***x (n)*** is applied with noise to unknown system and adaptive LMS filter thus we obtain the output signal which is mixed of the speech signal and noise but are separated in time and frequency domain, thus the error e(n) is the deference between ***d (n)*** and $\hat{y}$ ***(n)*** [5]:

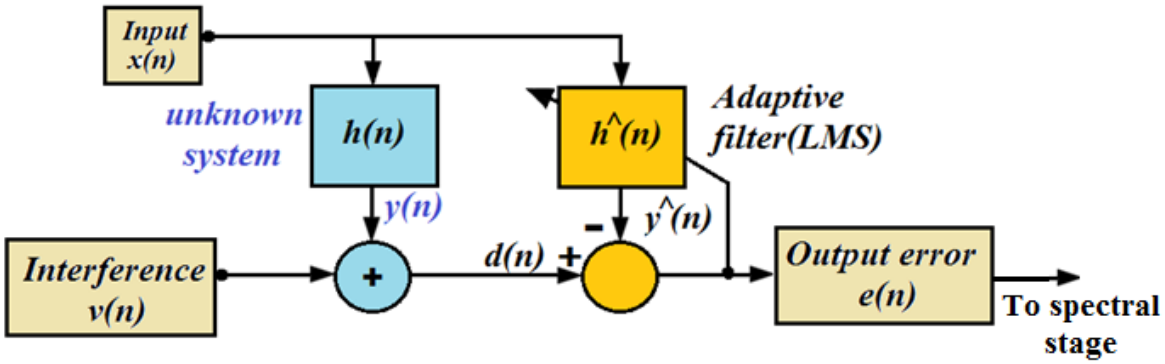

Figure 1. Block diagram for formulating LMS filter.

$$y\,(n) = x\,(n)\,h\,(n) \quad (1)$$
$$y\hat{}\,(n) = x\,(n)\,h\hat{}\,(n) \quad (2)$$
$$d\,(n) = y\,(n) + v\,(n) \quad (3)$$
$$e\,(n) = d\,(n) - y\hat{}\,(n) \quad (4)$$

3. Relationship with the filter of the least squares

The perception of the causal Wiener filter is very similar to the resolution of the least squares, except the signal processing field. The solution of the least squares, for the input of matrix **X** and the vector output **y** is [6]:

$$\hat{\beta} = (XT\,X) - 1\,XT\,y \quad (5)$$

The squares of FIR sources are associated with the Wiener filter, but do not reduce the error criterion for the first to the reciprocal link or automatic links. Resolve converges with Wiener filter solution. Most adaptive linear filtering problems can be written using figure 1. This means that the unknown system is ***h (n)*** to be defined. The adaptive filter tries to adjust the filter $\boldsymbol{h}^{\wedge}$ ***(n)*** to make it nearest to ***h (n)***, using only observable signals ***x (n), d (n)*** and ***e (n);*** but ***y (n), v (n),*** and ***h (n)*** cannot be observed directly.[7]

n is the number of the current input sample

p is the number of filter taps

$\boldsymbol{\{.\}^{H}}$ is (conjugate transpose)

$$X\,(n) = [x\,(n), x\,(n-1) \ldots x\,(n-p+1]\,T \quad (6)$$
$$h\,(n) = [h0\,(n), h1\,(n) \ldots hp-1\,(n)]\,T, h\,(n) \in \mathbb{C}^{p} \quad (7)$$
$$y\,(n) = h\,(n).X\,(n) \quad (8)$$
$$d\,(n) = y\,(n) + v\,(n) \quad (9)$$

$\boldsymbol{h}^{\wedge}$ ***(n)*** estimated filter interpret as the estimation of coefficients after n samples.

$$e\,(n) = d\,(n) - y\hat{}\,(n) = d\,(n) - h\hat{}\,(n)\,.X\,(n) \quad (10)$$

The basic idea of LMS filter is to reach the optimal filter weights $R^{-1}P$, by upgrading the filter weights in such a way that approximates the optimal filter weight. This depends on the regression algorithm. The algorithm assumed to be small weights (zero in most cases), at each step, then finds mean square error (MSE). [8]

If ***MSE*** ↑ ➡ ***e (n)*** ↑ ***if mean square error increases then the output error increases***

MSE = W$_n$↓ ➡ ***W$_n$*** ↓ ***if mean square error equals to the weights of n then the weights decreases***

MSE ↓ ➡ ***W$_n$*** ↑ ***if mean square error decreases then the weights increases***

Therefore, the basic equation for weight modernization is:

$$Wn + 1 = Wn - \mu\nabla\varepsilon[n] \quad (11)$$

Where $\boldsymbol{\varepsilon}$ represents the error of the mean square and $\boldsymbol{\mu}$ is the affinity coefficient. The negative sign shows that the slope error $\boldsymbol{\varepsilon}$ is going down, to find the filter weights $\boldsymbol{W_i}$ which minimize the error. The LMS algorithm for a ***pth*** order algorithm can be summarized as

Parameters: $\boldsymbol{p}$ = filter order

$\boldsymbol{\mu}$ **= step size**

Initialization: $\boldsymbol{h\hat{}\,(0)}$ **= zeros (*p*)**

Computation: For n = 0, 1, 2 …

$$X(n) = [x(n), x(n-1) \ldots x(n-p+1)]T \quad (12)$$

$$e(n) = d(n) - h\hat{}H(n)\,x(n) \quad (13)$$

$$h\hat{}(n+1) = h\hat{}(n) + \mu\, e*(n)\,x(n) \quad (14)$$

4. Algorithm and reasons for using digital signal processing (DSP) for the speech signal.

Security and safety in our contemporary world is very necessary, there are passwords, user names and security numbers in e-mails, messengers, and even in master or credit cards [9]. There are government banks and private sector banks withdraw or deposit money in international banks. And the transfer of funds for the purchase and sale of international remittances withdrawn or transferred to official or non-official [10]. All of these need security and security of the highest degree and with very high accuracy [11]. There are piracy factors where confidential numbers are accessed and hacked. But when you use voice recognition, you're far from hacking [12].

The work presented is related to research conducted in the development of a "speech recognition "using two languages to distinguish frequencies between them and the source of the audio signal (speaker) is determined [13]. To accomplish that, it should be very important to develop tools which allow sophisticated search in speech processing, analysis, estimation and recognition [14]. In the development of such a system it is necessary to conduct several tests of different system models. The units range from voice signal processing to extraction, voice activity Combine combination, classification styles, grading algorithms, etc., must be joined for execution speech recognition. Thus, the main drawback in this area of research is analysis, examination, demonstration, and integration of particular tasks necessary for speech discrimination. [15].

5. Noise filtration of speech signal.

Ear recognition is one of the biometric security systems actually used define speaker based on acoustic characteristics. Select speaker depends on various audio features like density analysis, sound level analysis, and audio feature extraction etc. [10]. This process is also reacted by the recognition of various elements such as noise in the background, hardware noise etc. We will define the effective noise access which is considered to identify the process of identifying the active speaker. In this access, durability the recognition system will be reformed by defining the integrated layers model. The durability will be accomplished for background noise and hardware noise [16]. The filtration can be divided into two main blocks. In the first stage, filtration is at a high level above noise is implemented to remove noise in the background. To implement this high-level separation method by using subtraction of spectrum. In the subsequent phase is to remove the noise of devices by using the linear probability coding policy [17].

This will reduce the effective noise which is above the pure signal. Additive white Gaussian noise (AWGN) is the main statistical noise in speech signal and information which has the same probability density function (PDF) as Gaussian distribution (GD) [18]. Figure (2) shows how can the noise be eliminated from the speech signal in two languages with male speech [19]. Figure (3), (4) and (5) shows filtering male and female speech also in two languages. [20]

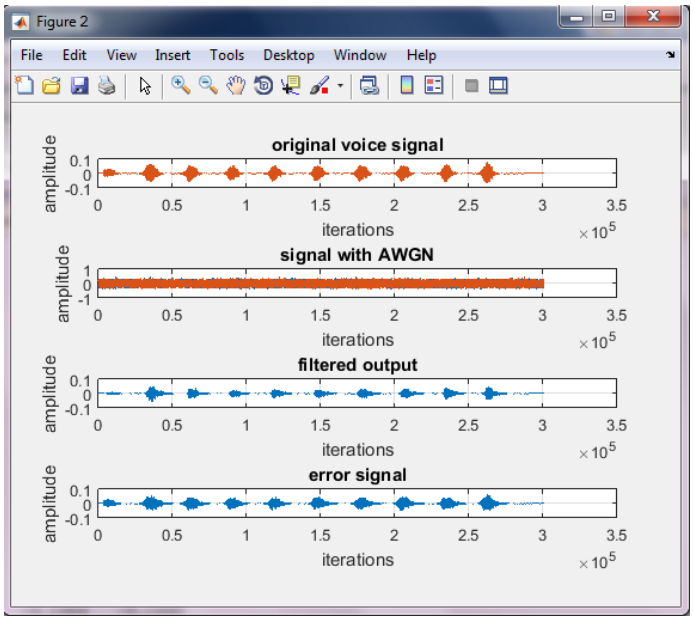

Figure 2. Filtering speech signal of number one in Arabic Language "male speech signal"

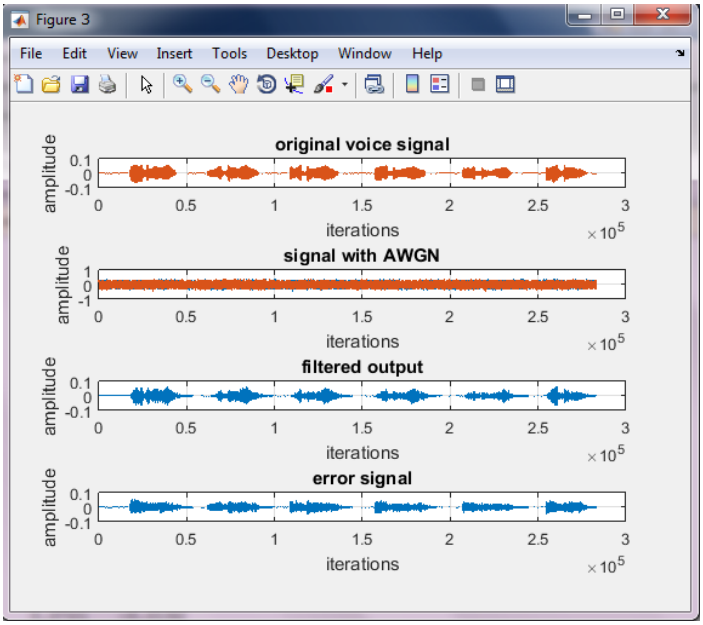

Figure 3.Filtering speech signal of number one in English Language "male speech signal"

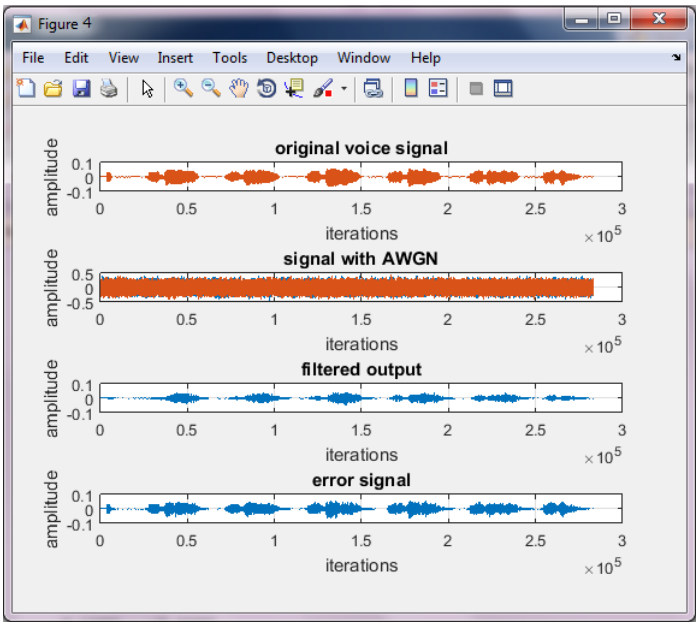

Figure 4. Filtering speech signal of Hello in Arabic Language "female speech signal"

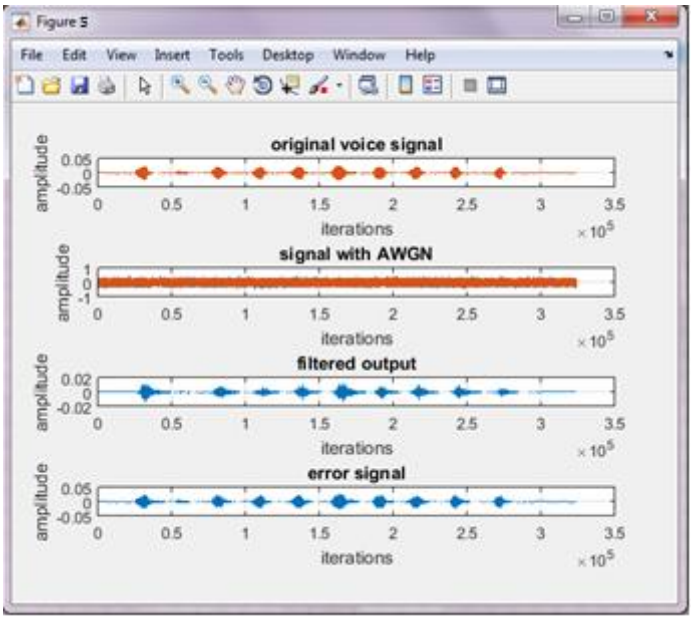

Figure 5. Filtering speech signal of Hello in English Language "female speech signal"

6. Spectral subtraction

Other technique beside LMF filtration is the fundamental method of spectral subtraction for de-noising, fast Fourier transform FFT to the noisy speech if existed and FFT to a pure noise then subtract the magnitude of these two spectra and do inverse FFT (IFFT) to reconstruct the transient signal by add the phase information of noisy speech. [21]

Speech algorithms in many information theories operate in the Discrete Fourier Transform (DFT) domain [22] assume that the real and imaginary part of the clean speech DFT coefficients can be modeled by different speech enhancement algorithms. In Fourier domain, we can write y (n) as

$$Xe[w]\,p = |Y[w]\,|p - |De[w]\,|p \quad (15)$$

Xe[w] is the estimation of clean speech Magnitude signal spectrum

Y[w] = Y |(w)| e $^{j\,\emptyset\,y}$ Where ***|Y(w)|*** is the magnitude spectrum and ***Ø*** is the phase spectra of the corrupted noisy speech signal. As shown in figures 6,7,8,9 how FFT, IFFT, NFFT and DFT produce

the pure speech signal. Single-Sided Amplitude Spectrum (SSAS) and their frequencies for figures 2, 3, 4 and 5 are plotted in figures 6, 7, 8, 9. All frequencies
f= 1x262145 double, sampling frequency fs=100, T=0.0100 and NFFT=524288.

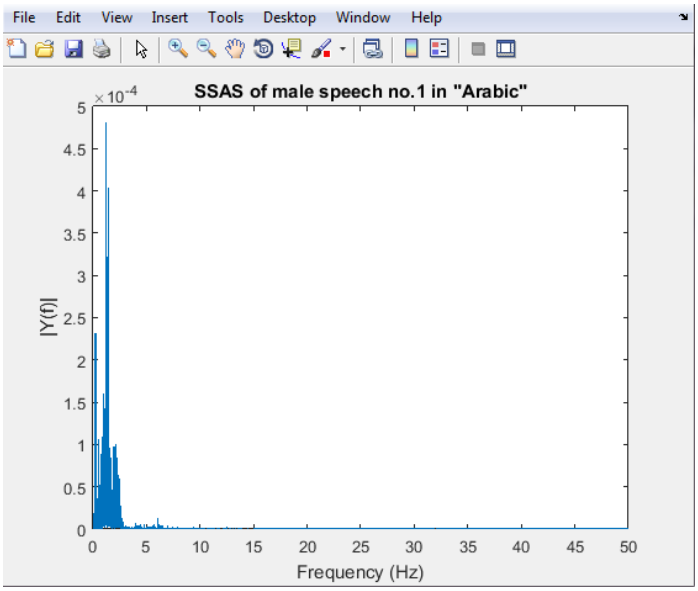

Figure 6. SSAS of male speech number one in Arabic English Language (L=300672)

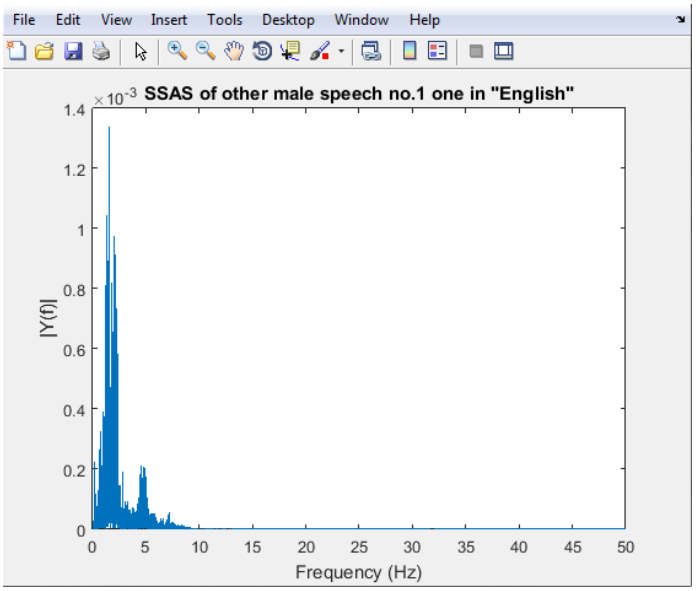

Figure 7. SSAS of other male speech number one in Language (L=323712)

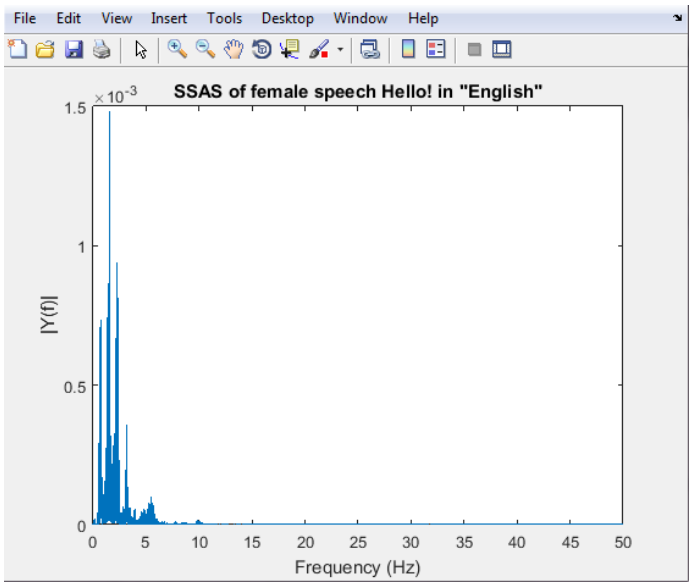

Figure 8. SSAS of female speech Hello! in Arabic Language (L=334080)

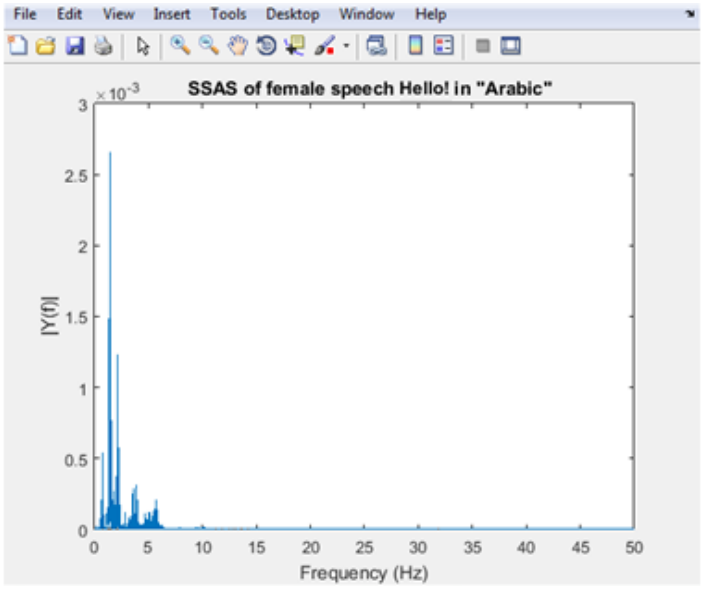

Figure 9. SSAS of female speech Hello! in English language (306432)

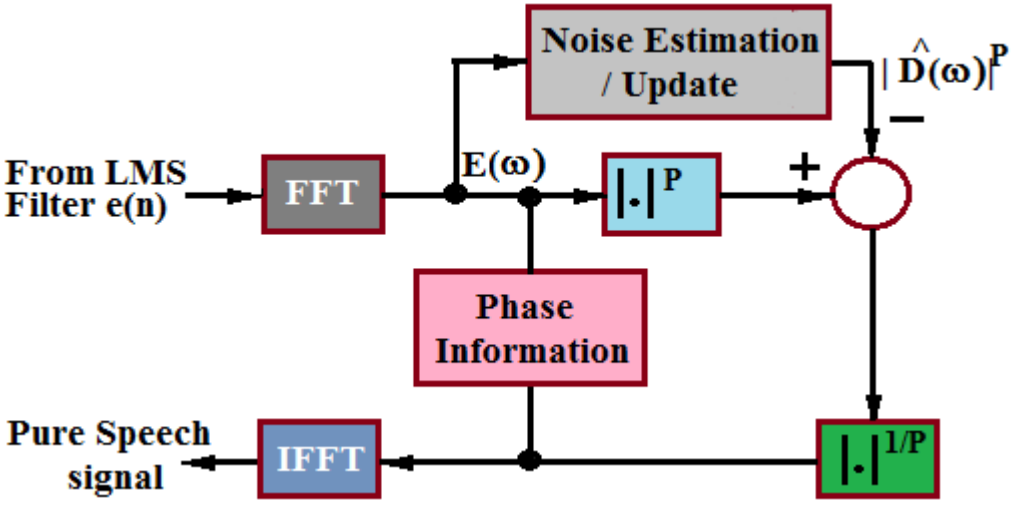

Figure 10. The spectral subtraction techniques to council noise in the speech sgnal

|De[w]| is the average value or estimated noise ***|De (w)|*** computed during non-speech activity that is during speech pauses.

P is the power exponent¸ the general form of the spectral subtraction, when ***p=1*** that's mean the magnitude spectral subtraction algorithm, ***p=2*** means the power spectral subtraction algorithm. The general form of the spectral subtraction algorithm is shown in figure 10. [23]

7. Description for conversion analog to digital (a/d) speech signal techniques

This technique converts a continuous-time and amplitude of analog signal to a discrete-time and amplitude of digital signal using quantization of the input pure speech signal and always the sampling input and limits input, and limits the allowable bandwidth. Sampling is to measure signal at periodic time intervals, and by using Nyquist-Shannon theorem that fs > 2fmax. We should take care for aliasing that introduction of false (alias) frequencies in the process of sampling or reconstruction that were not present in the original speech signal. To avoid aliasing the low pass filter (LPF) to band limit the analog speech signal (input signal) prior to sampling (antialiasing filter). As shown in figure 11 the complete security system.

After quantization and filtering the pure speech signal, the system in figure 12 allows the real user and opens the data.

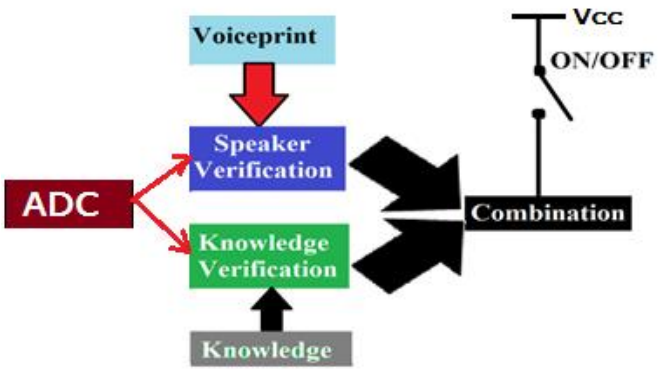

Figure 11. the complete security system of the speech signal

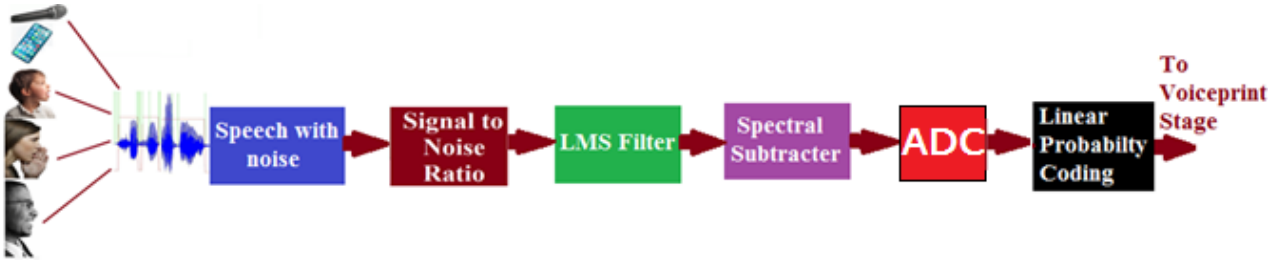

Figure12. how to enter to the system with real individual speech signal.

The voiced speech of a typical adult male will have a fundamental frequency from **85 to 180 Hz**, and that of a typical adult female from **165 to 255 Hz**. Thus, the fundamental frequency of most speech falls below the bottom of the "voice frequency" band as defined above.

8. Conclusion

The main task of making use of the voice recognition system is the safety ratio it achieves. Although sound discrimination is approximately secure, it still suffers from shortages and failures. In order to achieve a very high security, this system is integrated with traditional security features to provide an additional layer of security. This may include the use of other biometrics or security techniques like RSA or PINs or a collection of various different techniques. Finally, and for best development, voice discrimination can be one of very important, authenticated and secure applications in the future with multiple firewall and more security.

References

[1] Bahdanau D 2016 End-to-End Attention-based Large Vocabulary Speech Recognition.
[2] Deng L, Li J, Huang J, Yao K, Yu D, Seide F et al. 2013 Recent Advances in Deep Learning for Speech Research at Microsoft ICASSP.
[3] Priyanka G, Mukesh P, Pragya N 2015 Performance Analysis of Speech Enhancement Using LMS NLMS and UNANR algorithms *IEEE International Conference on Computer Communication and Control*, September.
[4] Hadei A and lotfizad M 2010 A Family of Adaptive Filter Algorithms in Noise Cancellation for Speech Enhancement *International Journal of Computer and Electrical Engineering* **2**, No. 2, April.
[5] Górriz J M et al. 2009 A Novel LMS Algorithm Applied To Adaptive Noise Cancellation *IEEE Signal Processing Letters* **16**, No. 1.
[6] Weifeng L, Jose P and Simon H 2010 *Kernel Adaptive Filtering: A Comprehensive Introduction,* John Wiley.
[7] Craig V 2005 China's opening to the world: what does it mean for US banks?" Federal Deposit Insurance Corporation's Banking Review **17** No. 3.
[8] Amarnag S 2006 Speech Modeling with Magnitude-Normalized Complex Spectra and Its *Application to Multisensory Speech Enhancement*‖, ICME, IEEE.
[9] Yu, R-L, Zhang J-C 2009 Speaker recognition method using MFCC and LPCC features *Computer Engineering and Design* **30** 1189–1191.
[10] Aldhaheri R, Al-Saad F 2014 Robust text-independent speaker recognition with short utterance in noisy environment using SVD as a matching measure Comp. Info. Sci., **17**.
[11] Furui S 2005 50 Years of progress in speech and speaker recognition, in: Proceeding of the *International Conference Speech and Computer SPECOM'05*, Patras, Greece, pp. 1–9.
[12] Esfandiar Z 2007 Noisy Speech Enhancement Using Harmonic-Noise Model and Codebook-Based Post-Processing *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING* IEEE.
[13] Naseem I, Deriche M 2006 A new algorithm for speaker identification using the Dempster–Shafer Theory of evidence, in: *Proceedings of the IPCV'06*, pp. 47–51.

[14] Shaban A, Sultan M, Aljebory K Speaker identification 2007 a hybrid approach using neural networks and wavelet transform J. *Comput. Sci*. **3**.
[15] Djemili R, Bedda M, Bourouba H 2007 A hybrid GMM/SVM system for text independent speaker identification *Int. J. Comput. Inform. Sci. Eng*. **1**.
[16] Saeed K 2006 A note on biometrics and voice print: voice-signal feature selection and extraction – a Burg–T?eplitz approach, *in: Proc. of the 10th IEEE Workshop on Signal Processing, SP'06, Poznan, Poland*, pp. 7–12.
[17] Alkanhal M, Alghamdi M, Muzaffar Z 2007 Speaker verification based on Saudi accented Arabic database, in: Proceeding of the ISSPA *International Symposium on Signal Processing and its Applications*, Sharjah, UAE, February, pp. 1–4.
[18] Hesham T 2011 A high-performance text-independent speaker identification of Arabic speakers using a CHMM-based approach *Alexandria Engineering Journal* **50** 43–47.
[19] Ramachandran R, Farrell P, Ramachandran K R, Mammone R R J 2002 Speaker recognition—general classifier approaches and data fusion methods *Pattern Recognition* **35** 2801–2821.
[20] Schmidhuber J 2015 Deep Learning Scholarpedia **10** (11).
[21] Phillips C L, 2007 Speech enhancement theory and practice 1st ed. Boca Raton, FL.: CRC. *Releases Taylor & Francis*
[22] Yi H and Philipos C L 2007 Subjective comparison and evaluation of speech enhancement algorithms *IEEE Trans. Speech Audio Proc*. **49** 588–601.
[23] Paliwal K and Alsteris L 2005 On usefulness of STFT phase spectrum in human listening tests *Speech Commmun*. **45** 153-170.