Advance Digital Signal Processing Methods for Speech Recognition Systems

Ramdan A. M. Khalifa[1], KHALED SEGHAYER ELGANTRI[2], Adel Issa Ben Issa[3]

1 High Institute of science and Technology-Suk Algumaa. / Almarifa University Tripoli, Libya

2 High Institute of science and Technology-Suk Algumaa. Tripoli, Libya

3 Higher institute of engineering technologies Tripoli. Tripoli, Libya

ramadanamharee@gmail.com

الملخص

أصبح التعرف على الصوت نهجًا أساسيًا كجزء من العديد من التقنيات والأنظمة المتطورة في عصرنا الحديث حيث تم اعتماده من قبل شركات تصنيع التكنولوجيا المختلفة لأغراض مختلفة. تحتاج منظومات التقنيات الحديثة ، التي تستخدم أنظمة التعرف على الكلام ، إلى خوارزميات أكثر دقة لأداء وظائف التعرف على الكلام الحقيقي. في هذا البحث تم اقتراح نظام فعال للتعرف على الكلام لتحقيق نتائج أفضل مقارنة بالطرق الكلاسيكية. تم اعتماد تقنيات معالجة الإشارات المتقدمة لتحليل الكلام المسجل ، والتي تقيس المعلمات المطلوبة قبل تنفيذ مهمة التعرف علي بصمة الصوت. تم تسجيل الكلام باستخدام برنامج MATLAB عن طريق ضبط مسجل قناتين بجودة 32 بت ومن ثم يتم استخدام هذا الكلام كعينة لإكمال هذه الدراسة. يشتمل النظام على قسم قابل للإزالة من الضوضاء لإزالة المقاطع بدون صوت بحيث يتم استغراق وقت أقل لإجراء التعرف على الكلام. تم حساب الارتباط التلقائي لمطابقة الإشارة المسجلة مع إشارة الاختبار ، ثم تم تحديد فترة الملعب لضمان هوية المتحدث. تمت ملاحظة النتائج من خلال إجراء التعرف على الكلام لمعرفة ما إذا كان النظام يمكنه تحديد محتويات الحديث وتحديد السماعة ، لذلك يُسمح بالوصول إذا كان المتحدث نفسه بنفس المحتوى يدخل صوته. لهذا الغرض ، تم استخدام برنامج MATLAB لإنشاء النظام كما تم عرض النتائج خلال هذه الورقة العلمية.

ABSTRACT

Voice recognition has become a primary approach as part of many advanced technologies and systems in our modern age as it has been adopted by different technology manufacturers for different purposes. Modern technologies, which use speech recognition systems, need more accurate algorithms to perform real speech recognition functions. In this paper, an effective speech recognition system has been proposed to achieve better results compared to the classical methods. Advanced signal processing technologies are adopted to analyze the recorded speech, which measures the parameters required before performing a voiceprint recognition task. The speech was recorded using MATLAB software by setting a 32-bit two-channel recorder and then this speech is used as a sample to complete this study.

The system includes a noise-removable section to remove syllables without sound so that it takes less time to perform speech recognition. Autocorrelation was calculated to match the

recorded signal with the test signal, and then the pitch period was determined to ensure the identity of the speaker. The results are observed by performing speech recognition to see if the system can identify the contents of the speech and identify the speaker, so access is allowed if the same speaker with the same content enters his or her voice. For this purpose, MATLAB software was used to build the system and the results were presented throughout this paper.

Keywords - Hidden Markov Model, Automatic, Speech Recognition.

I. INTRODUCTION

This project tried to recognize the speaker's identification when he/she speaks with others. For smooth and sophisticated SR, the Matlab digital signal-processing tools box and programmable functions are very helpful. Speakers may input their spoken language through the Matlab workstation by recording it in a digital format such as WAV or MP3. The channel type such as single channel or double channel may also be chosen with the speech rate for the best signal presentation. A reference signal can be saved, which may be chosen for reference while performing the comparison on further stages:

The frequency is evaluated by applying the Fast Fourier Transformation (FFT) to the signal. It also produces the information that constructs a speech signal while disturbances including microphone and circumstantial noises are linked with it. Digital filters must be utilized for removing noise components in the signal. Two microphones were used and the other microphone recorded the noise signal. After the signal settlement, we compared it with the reference database for evaluating the characteristics of the speaker.

As a result, the system recognizes the speaker's information including the name and designation, for example: "YOU ARE: KHALED ADEL RAMADAN, 37 AGED, DIRECTOR" or "YOU ARE NOT AUTHORIZED TO ACCESS THE SYSTEM."

II. Different kinds of voice recognition

There are three different kinds of ASR. The different bases on which scholars have worked include speakers' vocabulary size and bandwidth; these categories are categorized as follows: Isolated word, connected world, continuous word, and spontaneous word are the four types of words.[1] The focus of this study is on solitary word speech recognition. The user must pause after each syllable when using an isolated word speech recognition system. The system is divided into two phases: training and recognition. For each word said by the user, a training vector is generated during the training phase. For differentiating distinct classes of words, the training vectors extract spectral information. Each training vector can be used to create a template for a single word or a group of words. These training vectors (patterns) are saved in a database and used in the recognition phase later. During the recognition phase, the user can say any word that the system has been trained to recognize. A test pattern is created for that word and using a pattern comparison technique, the appropriate text string is displayed as the result.[2]

III. DSP approaches

This section demonstrates some key tools required for digital signal processing in both time and frequency forms.

A. **Signal Energy and Power**

Signal energy and signal power are phrases used to describe a signal. They aren't the energy or power meters. Any signal x(t), including signals with complex values, is included in the definition of signal energy and power [3]. The signal's signal energy x(t) is:

$$E = \int_{-\infty}^{\infty} |x(t)|^2 . dt \qquad (1)$$

The signal power in the signal x(t) is:

$$P = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |x(t)|^2 . dt \qquad (2)$$

If 0 <E< ∞, The signal x(t) is thus referred to as an energy signal. There are, however, signals where this criterion is not met. It is considered the power for such communications. If this is the case, the signal is referred to as a power signal. Note that the power for an energy signal is zero 0 <P< ∞ (P = 0) and that the energy for a power signal is infinite (E = ∞); some signals are neither energy nor power signals.[4]

B. Noise and disturbances, presence of noise

The following nominees can be used to categorize noise: White noise: as described in [5]. The electronic noise generated by the thermal agitation of the charge carriers (typically electrons) inside an electrical conductor at equilibrium (thermal noise, Johnson noise, or Nyquist noise) occurs regardless of the applied voltage. The fluctuation-dissipation theorem is a statistical physical derivation of this noise that uses generalized impedance or generalized susceptibility to characterize the medium. Thermal noise in a perfect resistor is white, indicating that the power spectral density is essentially constant across the frequency spectrum. as a figure (1).[6]

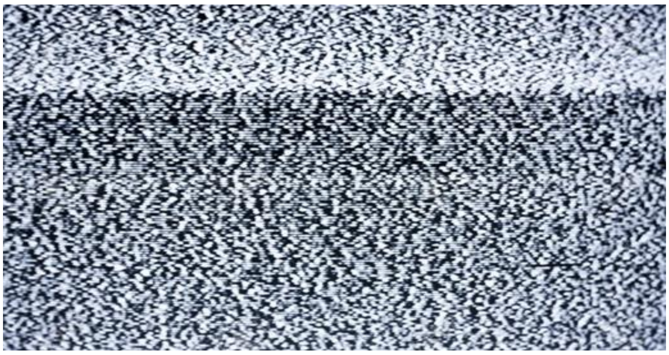

Fig (1). Show the thermal noise.

Gaussian noise: assumed quantity where it can produce when a random variable having Gaussian mean and variant is applied to the transmitted signal. The reason behind this production can be attributed to the channel condition where the signal is traveling from the source to the destination. This quantity can be described based on the variant and the standard deviation of the noise signal. [7]

$$Pn = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(n(t)-\mu)^2}{2\sigma^2}} \qquad (3)$$

The grey level, the mean value, and the standard deviation are all represented by n (t).

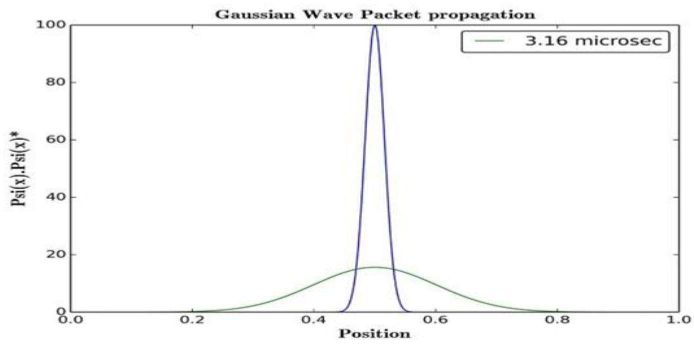

Fig (2). Show the Gaussian noise waveform.

1. Frequency representation of signals

Frequency representation of the signal is quite important to perform any analysis on that signal. The common technique for frequency domain conversion is Fast Fourier Transformation (FFT), wavelet transform; where the time domain of a signal lasting for 10 seconds is converted into the amplitude of this signal versus the frequency of occurrence. Normally frequencies of different components associated with the signal such as noise frequency can be waived off by establishing pre-designed filtering techniques. For the continuous signal x(t); the Fourier transform can be evaluated from the following.[8]

$$X(f) = \frac{1}{T} * \int_{-\infty}^{\infty} x(t) * e^{(-jwt)}.\, dt \qquad (4)$$

The resulting signal is attributed to the frequency (f) as shown and depicted in figure (4),

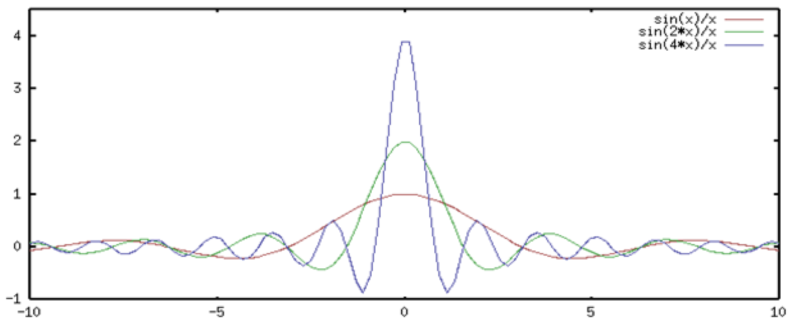

Figure 4: Fourier Transform depicted for a signal where the right-hand side is the real part and the left-hand side is the imaginary part of the resulting signal in the frequency domain.

where the original time domain signal is continuous over the period from minus infinity to infinity which is a general expression of Fourier Transform, the actual signal could be limited to a pre-defined period i.e. -10<t<+10. The point of interest is the discrete signal Fourier Transform where the signal is sampled with a sampling rate of double frequency (F sampling = 2* F). The DFT term for discreet Fourier transformation could be evaluated from the following expression[9]

$$X[n] = \sum_{-\infty}^{\infty} x[n] * e^{-jnwt} \qquad (5)$$

IV. Simulation and results

To make accurate speech recognition, many subsystems are required to be implemented. Speech recording and converting the analog speech signal into digital form is one of the important stages in SRS. It is necessary to understand the characteristics of the speech signal and the methods to detect the useful information from the said test signal and those methods used to remove the unused parts or noise interference as well. To make accurate speech recognition, many subsystems are required to be implemented as shown in Figure 5.

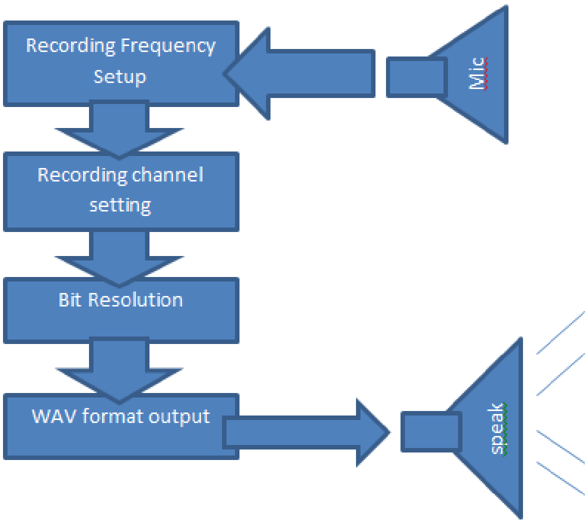

Fig (5). Explain the speech generation scheme.

Speech recording and converting the analog speech signal into digital form is one of the important stages in SRS. It is necessary to understand the characteristics of the speech signal and the methods to detect the useful information from the said test signal and those methods used to remove the unused parts or noise interference as well. The signal is brought into Matlab from the microphone by recording the speech of the person as shown in Figure 6, recording is needed to be done with high quality so many norms need to be followed to achieve the required quality.

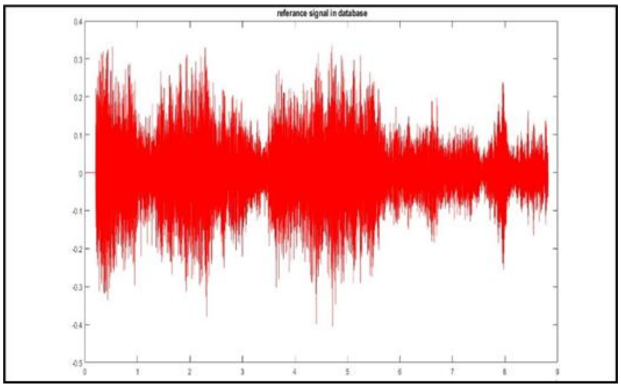

Fig (6) Show the Recorder signal as database input

The channels of a recording i.e. signal channel or double channel (Stereo or Mono), the frequency of voice signal that is specified while recording; normally 44100 Hz and 16-bit resolution are used for the recording of voice/sound signal. Figure 7 depicts the data input to Matlab. A normal speech phrase signal has two parts: one carries the speech data, and the other provides silence or noise sections in between the utterances that contain no verbal data.

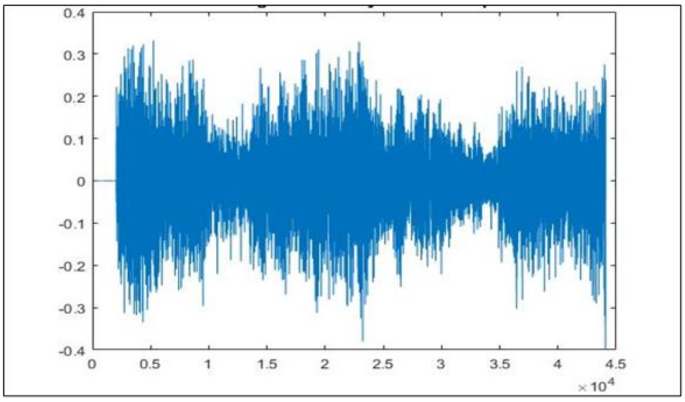

Fig (7). Show the Signal after recording with no silent parts

The verbal (informative) component of communication is further separated into two types: voiced and unvoiced speech. Vowel sounds make up the majority of voiced speech. It is created by forcing air through the glottis and adjusting the tension of the vocal cords, which results in the cords opening and shutting and the creation of practically periodic air pulses. The vocal tract is stimulated by these pulses. Psychoacoustics research demonstrates that this region of the speech contains the majority of the information and hence holds the keys to identifying a speaker. Turbulence is created by forcing air through a constriction produced at a location in the vocal tract (typically around the mouth end), resulting in unvoiced speech passages. It's critical for speech signal analysis to be able to distinguish between the three. The work is done to implement a speech recognition system that is capable to distinguish the voices of a variety of people; however, the program is made in MATLAB environments and language independent solution for voice recognition. Advanced digital signal processing is adopted in this work to achieve the best recognition results, the implemented system is attributed to lesser time complexity and quick response to the user. Data is always entered into Matlab in two parts, the reference input, and the external input. The database is loaded with reference speech signal and DSP is applied to extract the pitch frequency as in Figure 8.

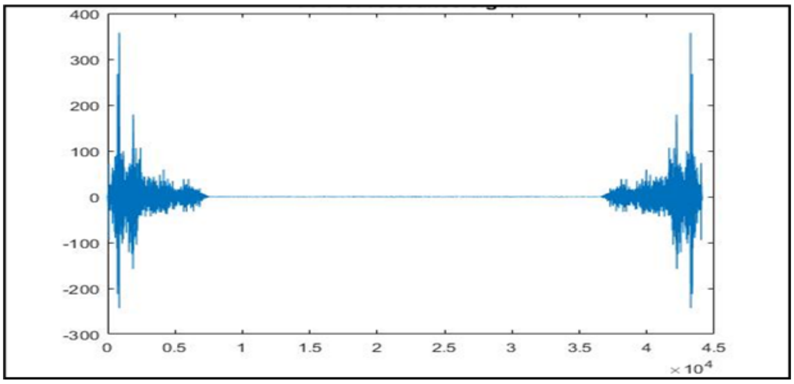

Fig (8). Show the Two-sided Fast Fourier Transform

The program is demanding another input to be the recognition. Normally such a system is providing a great contribution to data or system security as in Figure 9.

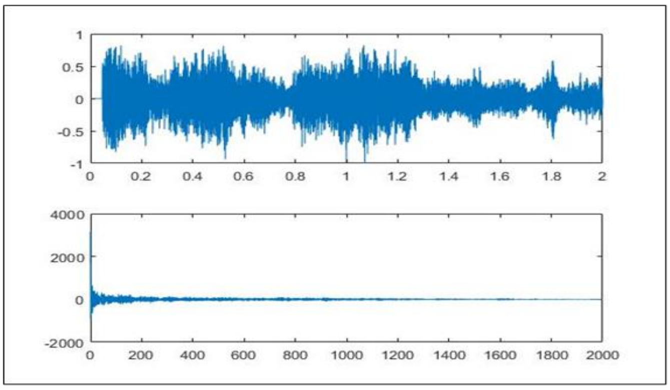

Fig (9). Show the first reference input with auto-correlation results

The user is needed to be the same and the voice signal is required to be the same so that the role of "same man and same password" is validated. Nobody can get access to the system but the main person who has fed the data earlier into the database. Such work is usually required to be implemented in the applications which demand that level of security where only the same person/authorized person can access the data. In a future expansion, the system may use a large database with a large number of users and all of them will be able to record a password and then use the same for entry.

V. Conclusion

As far as the early signal processing stage is concerned, many theories have been created to analyze analog signals such as signals of heat and light sensors. These approaches are used for analog-digital conversion (ADC) for making analog signals appropriate for digital environments specifically because the digital world is providing great facilities for quick and less expensive solutions. A speech is recorded in the form of an analog signal, which emerges out of vocal cord vibrations. Nowadays, speech is used for security and personal identity. It is a unique attribute in humans that every person has a different speech frequency and because of that, different people have different voices.

Advanced digital processing of signals was used in this project to achieve the best recognition results. The utilized system showed attributes such as less time complexity and quicker response to the user. For this kind of experiment, the data is always entered in Matlab into two parts, the reference input, and the external input. The database is loaded with a reference speech signal when the DSP is applied to extract its pitch frequency. At that time, the program demands another input for recognition. Normally, such a system has a great contribution to the data or system security. The user should be the same because the voice signal is required to be the same so that the condition “same person and same password” is validated. In this way, nobody other than the intended user will be able to access the system and feed the data into the database. Such work should be done with applications, which require a high level of security allowing only the authorized person to access the data. For future research, the mentioned system should be tested for larger databases allowing a large number of users all of them will be able to enter a password and access the system.

VI. References

[1]. ASEEM SAXENA, AMIT KUMAR SINHA, SHASHANK CHAKRAWARTI, and SURABHI CHARU, "SPEECH RECOGNITION USING MATLAB," International Journal of Advances In Computer Science and Cloud Computing, Volume 1, Issue 2, Nov-2013, ISSN: 2321-4058. Gender, Fate, and effect of oil and routes of exposure to commercial fisheries report June (2006), pp.9-14.

[2]. Atheer Tahseen Hussein, “Analysis of Voice Recognition Algorithms using MATLAB”, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 IJERTV4IS080082 www.ijert.org (This work is licensed under a Creative Commons Attribution 4.0 International License.) Vol. 4 Issue 08, August 2015

[3]. "SECURE SPEECH CONTROLLED ROBOT USING MATLAB AND ARDUINO," International Journal of Modern Trends in Engineering and Research (IJMTER) Volume 02, Issue 04, [April – 2015]. Hardik Chhatbar, Janak Trivedi, Rahul Chauhan, Darshan Bhatt, "SECURE SPEECH CONTROLLED ROBOT USING MATLAB AND ARDUINO," International Journal of Modern Trends in Engineering and Research ISSN: 2349–9745 (online); 2393–8161 (print).

[4]. Lavin Jalan, Rahul Masrani, Roshan Jadhav, Tejaswini Palav, “SPEECH RECOGNITION BASED LEARNING SYSTEM”, International Journal of Engineering Trends and Technology- Volume4Issue2- 2013.

[5]. Er. Abhishek Thakur, Assistant Prof. Neeru Singla, “Design of Matlab-Based Automatic Speaker Recognition and Control System”, Er. Abhishek Thakur* et al. / (IJAEST) INTERNATIONAL JOURNAL OF ADVANCED ENGINEERING SCIENCES AND TECHNOLOGIES Vol No. 8, Issue No. 1, 100 – 106.

[6]. P. Yasotha1 N. Rajasekaran2 M. Nathiya3, “Automatic Spectral Analysis for Speech to Text Conversion with Embedded Execution”, IJSRD - International Journal for Scientific Research & Development| Vol. 2, Issue 09, 2014 | ISSN (online): 2321-0613.

[7]. Siddhant C. Joshi1, Dr. A.N.Cheeran2, "MATLAB Based Back-Propagation Neural Network for Automatic Speech Recognition", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 7, July 2014.
[8]. Shiv Kumar, Member, IACSIT, IAENG Aditya Shastri and R.K. Singh, "An Approach for Automatic Voice Signal Detection (AVSD) Using Matlab", International Journal of Computer Theory and Engineering, Vol. 3, No. 2, April 2011 ISSN: 1793-8201.
[9]. Neha Sharma, Shipra Sardana, "Designing a Real-Time Speech Recognition System using MATLAB", International Journal of Computer Applications (0975 – 8887) National Conference on Latest Initiatives& Innovations in Communication and Electronics (IICE 2016).