Interface of Microcontroller Implementation of Voice Recognition for Person Authentication

Asst. Lecturer Wail Ibrahim Khalil

الخلاصة

تم استخدام النبرات الصوتية في بحثنا هذا لغرض إعطاء التخويل. والسبب في استخدام نبرة الصوت ذلك لان خصائص صوت الانسان ممكن تحليلها والتعرف عليها بشكل سهل، بالاضافة الى ذلك استخدام نبرة الصوت تعطي المرونة في استخدامها في انظمة السيطرة وغيرها. في هذا البحث استخدم برنامج التعرف على نبرات الصوتية وكذلك تم تصميم دائرة الكترونية تحتوي على مسيطر دقيق لغرض عرض النتائج عن عملية التطابق على الشاشة. وهذا النوع من العمل قادر على تشخيص النبرات الصوتية بعد عملية التعليم للشخص الجديد وحفظ هذه النبرات الصوتية في قاعدة بياناته وبالتالي تنفذ لتعطي رسالة الى منفذ الاخراج ومنه يمكن ايضاً السيطرة على فتح الابواب لغرض اعطاء صلاحية الولوج للشخص المخول بالدخول وبنفس الوقت اشعار المراقب بمعلومات الشخص المخول. تم تصميم نظام يحتوي على برنامج تشخيص الصوت وكذلك دائرة الكترونية تحتوي على مسيطرة AT89S52 مع شاشة عرض LCD، وتم تدريب الايعازات للاشخاص لغرض التعرف واعطاء الاخراج الصحيح.

Abstract

Voice is used in this proposed for the authentication and display information of registered person and find whom is authorized to be access or not. The reason for choosing voice of human because is considered part of biometric system method and is easily being reproduced by human. Besides that, usage of voice gives control system that be effective and convenient to be used. This proposed involve a simple system that consists of voice recognition software and AT89S52 microcontroller with LCD to build up the system. This work is able to recognize the command trained by the user and successfully gives the correct output.

Keywords: voice recognition, Fourier Transform, Microcontroller AT89S52

1.Introduction

The data generated from voice signals are captured by a microphone attached to a PC. Signal generated by speech is an analogue signal, which must be digitized at a certain frequency. Typically, most moderate grade microphones employed have a sampling rate of approximately 32 kHz. So a 32 kHz sampling rate is generally more than sufficient for human voice patterns [1].

For reliable signal acquisition of voice data, the frequency and amplitude of the signal must be acquired with high fidelity. This is an issue with speakers with a large high frequency component, such as women and children. Typically, most modern recording devices are capable of digitizing voice data at 16 bits or more, providing more than sufficient dynamic range to cover human speech patterns [1].

In a speaker-independent system, the user's voice pattern is analyzed and compared to all other voice samples in the user database. There are a number of ways this comparison is made regarding the voice samples and these details are provided via closest match to the voice data presented for identification of the speaker. There are three possible outcomes: i) The speaker is correctly identified; ii) the speaker is incorrectly identified as another speaker; or iii) the speaker is not identified as being a member of the system. Clearly, the last two possibilities will avoid, which reflect the false acceptance rate (FAR) (type II error) and the False Rejection Rate (FRR) as much as possible [8]. When speakers attempt an authentication task, the speakers have provided some evidence of their identity, and the purpose of the voice recognition process is to verify that these persons have a legitimate claim to that identity. The result of this approach is a binary decision: either the person is verified as the claimed identity or not, and when be verified a message from voice software will be managed and sent to device plugged to PC via serial port and this device contains the microcontroller and LCD which shows the name of person whom authorized to access or any other issues could be implemented of this work [9].

As shown in figure (1) the block diagram of the proposed system, the first step is the capture the voice signal via microphone and this signal is an analogue form which need to be converted to a digital form by using Fourier Transform used to denoising a signal[5].

The frequency domain is used for noise reduction and echo cancellation of the voice signal. The system implemented in frequency domain and fourier transform to split frequency sampling[6].

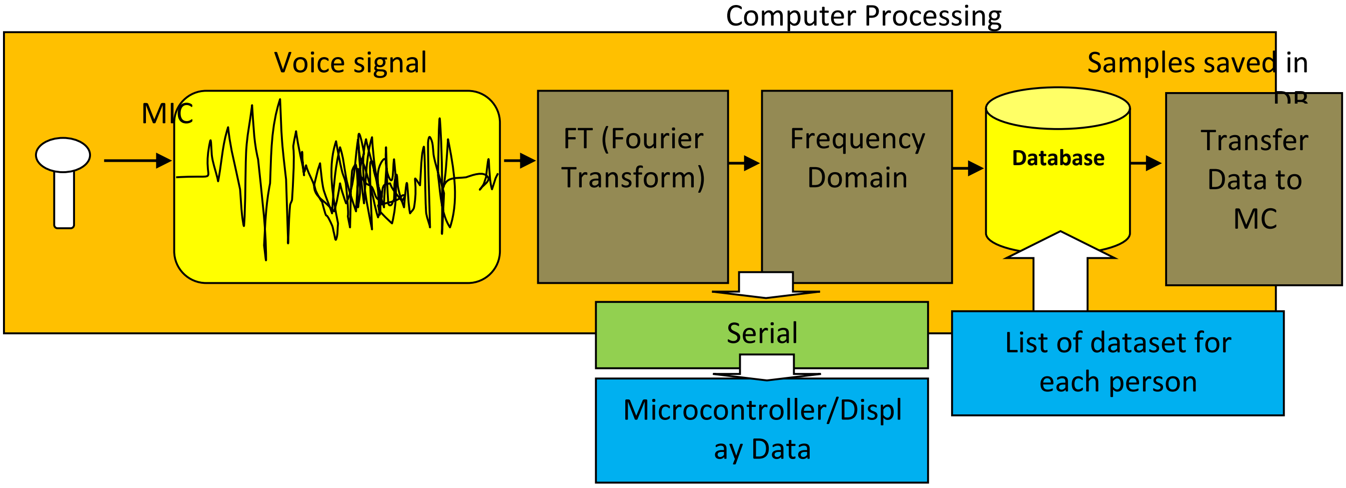


Figure 1. Block diagram of proposed work

2.Feature Extraction

The basis of most feature extraction approaches is the observation, that the speech spectrum shape encodes significant information about the speaker's vocal tract. This appears to be mainly the case due to two reasons[2].

First, the human vocal tract forms a resonator and second, the glottis is a source of pitch harmonics. Consequently, most feature extraction approaches aim at extracting spectrum-based features and typical method for extracting first-order features is Fourier Transformation (FT). Based on these first-order spectra, many approaches compute higher order cepstral features, such as Mel-Frequency Cepstral Coefficients (MFCC). Cepstral analysis is based on the idea of inverse transforming the logarithmic spectrum from frequency to temporal domain, thus the cepstrum of a signal can be defined as the Inverse FT (IFT) of the logarithm of the FT of the original signal[2].

Spectral or cepstrum features are extracted for temporal subsections of the entire audio signal, called windows. These windows are usually shifted sequentially over the entire signal and are characterized by a window size (in milliseconds, ms) and a window-to-window offset in milliseconds. Figure 1., illustrates this typical feature extraction process for voice biometrics which proposed work, starting from the output signal on the left and leading to the person displayed name on LCD representation on the right.

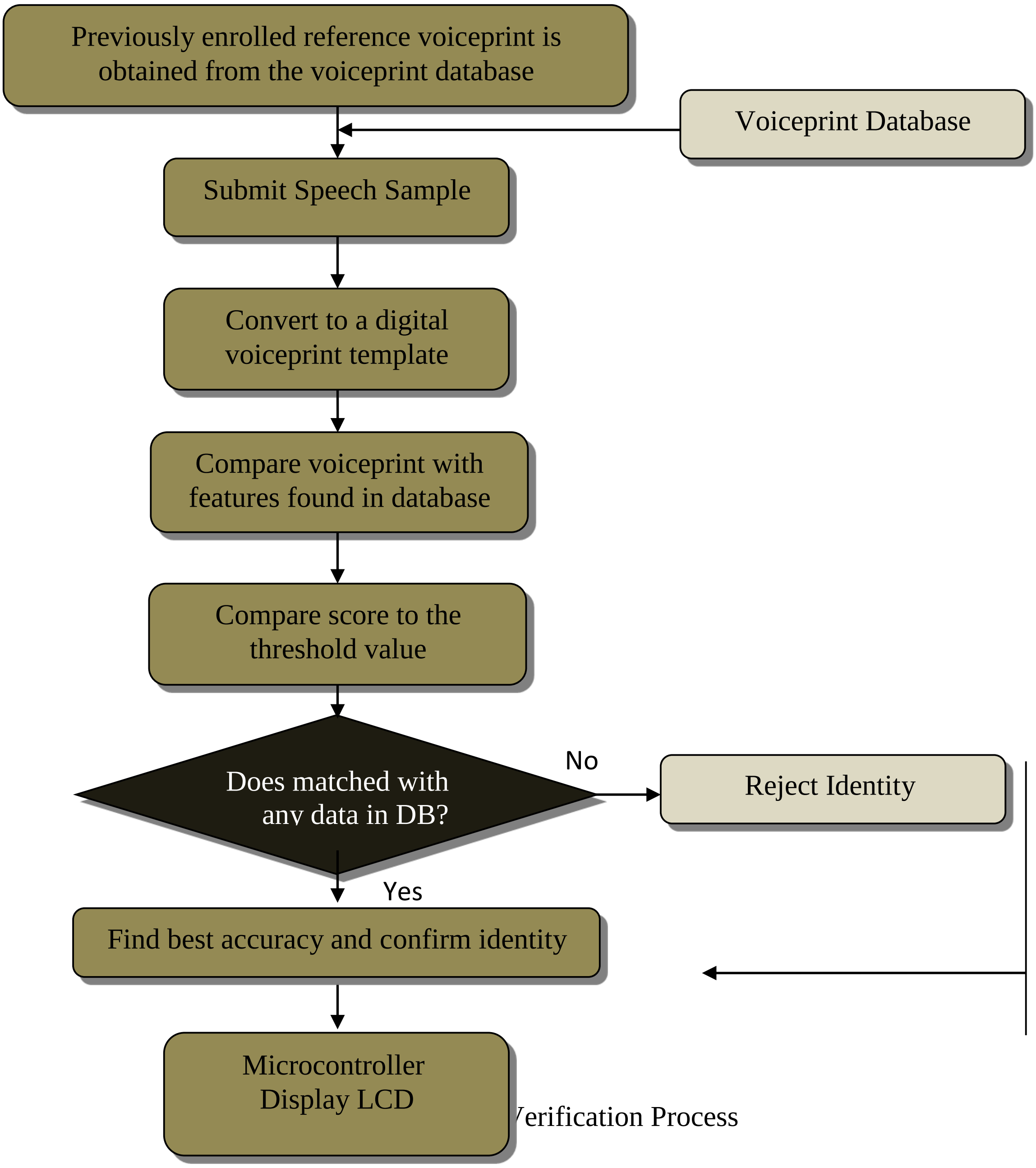
Note that in this illustration, no preprocessing is included. However, in many practical systems, such preprocessing includes techniques like filters for reducing background noises or other methods for emphasizing those signals that are relevant for biometric features. These filters are typically implemented as transfer functions in frequency domain[2].

3.Comparison and Classifications

The feature factor has extracted from phase of feature extraction of both stages (trained and recognition), mainly two different models are used in speech biometrics: template models and stochastic models. In the first category, cumulated distances are calculated between the features of the actual test sample and the stored reference features in database as stream with the name of person. Based on such distance values, the verification process can be implemented, for example, simply as a threshold-based decision function whereby the distance between reference and sample may not exceed a given threshold value. For that in this research the threshold value range is +/-10 has been chosen[2].

4.Voice Authentication

The voice authentication recognized the user's particular word or phrase. The voice scanners can have a problem with false non-matching, failing to authenticate a known user, but they tend to be resistant to false matching, so impostors are unlikely to be authenticated. Voice biometrics are a numerical model of the sound, pattern and rhythm of an individual's voice. A voice biometric or 'voice print' is unique to an individual [3].



erification Process

5.Frequency Domain Interpretation of the Short-Time Fourier Transform

The short-time Fourier transform of the sequence s(m) is defined as

$$S_n(e^{jw_i}) = \sum_m s(m)w(n-m)e^{-jw_i m} \ldots\ldots (1)$$

if take the point of view that are evaluating$S_n(e^{jw_i})$ for a fixed n=n_0, then we can interpret eq.1

$$S_n(e^{jw_i}) = FT[s(m)w(n_0-m)]|_{w=w_i} \ldots\ldots (2)$$

where FT denotes the Fourier Transform. Thus $S_n(e^{jw_i})$ is the conventional Fourier transform of the windowed signal, $s(m)w(n_0-m)$, evaluated at the frequency w=w_i. Figure (3).illustrates the singals s(m) and w(n-m), at times n=n_0=50, 100, and 200 to show which parts of s(m) are used in the computation of the short-time Fourier transform. Since w(m) is an FIR filter, if we demote that size by L, then using the conventional Fourier transform interpretation of $S_n(e^{jw_i})$, then can state the following[4]:

1. If L is large, relative to the signal periodicity (pitch), then $S_n(e^{jw_i})$ gives good frequency resolution. That is, then can resolve individual pitch harmonics but only roughly see the overall spectral envelope of the section of speech within the window[4].
2. If L is small relative to the signal periodicity, then $S_n(e^{jw_i})$gives poor freuqency resolution (i.e., no pitch harmonics are resolved), but a good estimate of the gross spectral shape is obtained[4].

To illustrate these points, Figure (4) show examples of windows signals, $s(m)w(n_0-m)$, (part a of each figure) and the resulting log magnitude short time spectra, 20 $\log_{10}|S_n(e^{jw_i})|$ (part b of each figure). Figure 4 shows result for an L =500 point Hamming window applied to a section of voiced speech. The periodicity of the singal is clearly seen in the windows time waveform, as well as in the short-time spectrum in which the fundamental frequency and its harmonics show up as narrow peaks at equally spaced frequencies. Figure 5 shows a similar set of comparisons for an L=50 point Hamming window. For each short windows, the time sequence$s(m)w(n_0-m)$ does not show the signal peridicity, nor does the signal spectrum. In fact, what we see in the short-time Fourier transform log magnitude is a few rather broad peaks in freuqency corresponding roughly to the speech formants[4].

Figures (3) and (4) show the effect of using windows on a section of unvoiced speech (corresponding to the fricative /sh/) for an L=500 sample window (Figure 3) and L=50 sample window (figure 4). Since there is no periodicity in the signal, the resulting short-time spectral magnitude of Figure 4, for the L=500 sample window shows a ragged series of local peaks and valleys due to the random nature of the unvoiced speech. Using the shorter window smoothes out the random fluctuations in the short-time spectral magnitude and again shows the broad spectral envelope very well.

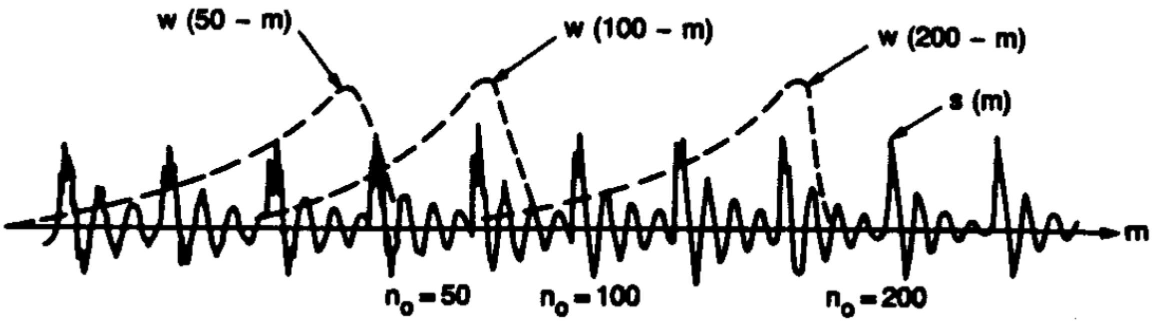


Figure (3) The signals s(m) and w(n-m) used in evaluation of the short-time Fourier transform [4]

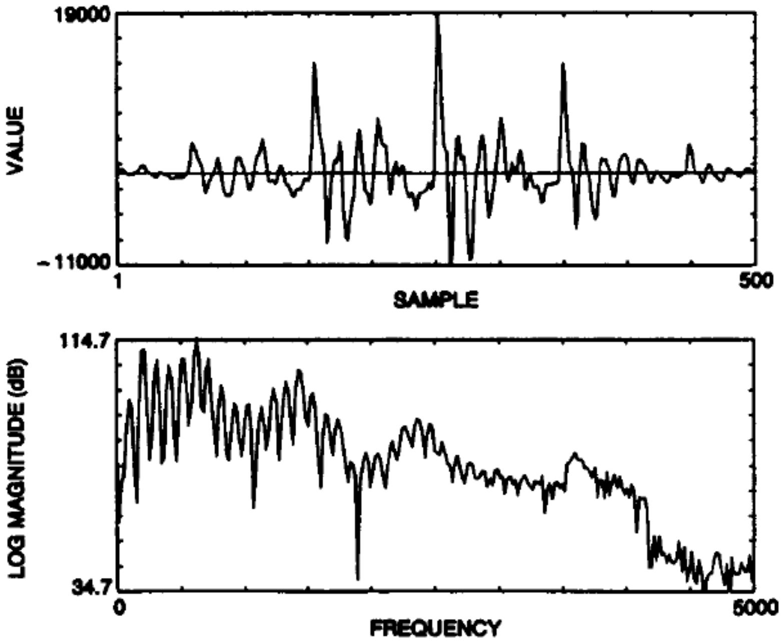


Figure (4) Short-time Fourier transform using a long (500 points or 50 msec) Hamming window on a section of voiced speech[4]

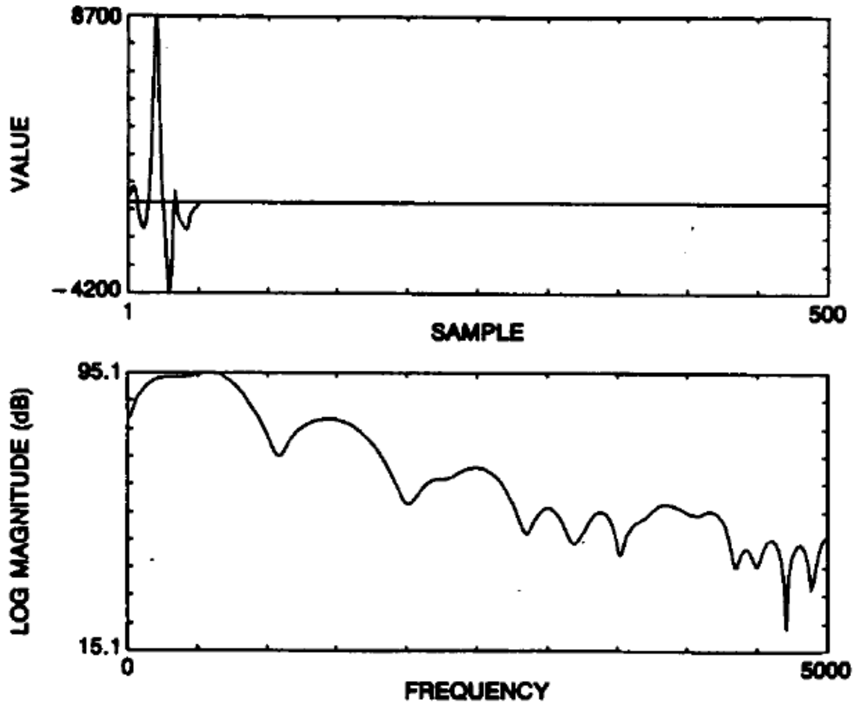


Figure (5) Short-time Fourier transform using a long (50 points or 5 msec) Hamming window on a section of voiced speech[4]

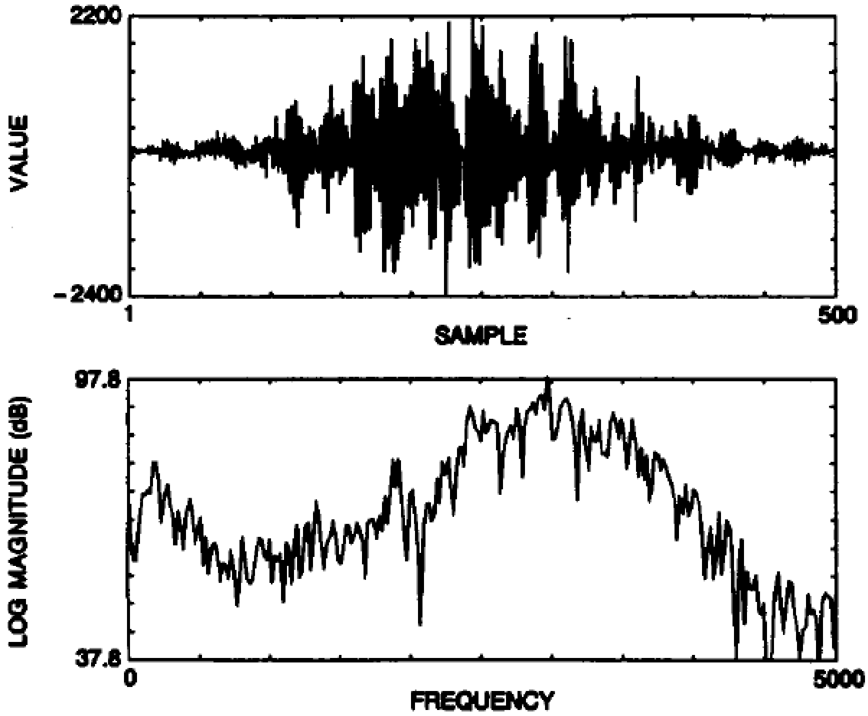


Figure (6) Short-time Fourier transform using a long (500 points or 50 msec) Hamming window on a section of unvoiced speech[4]

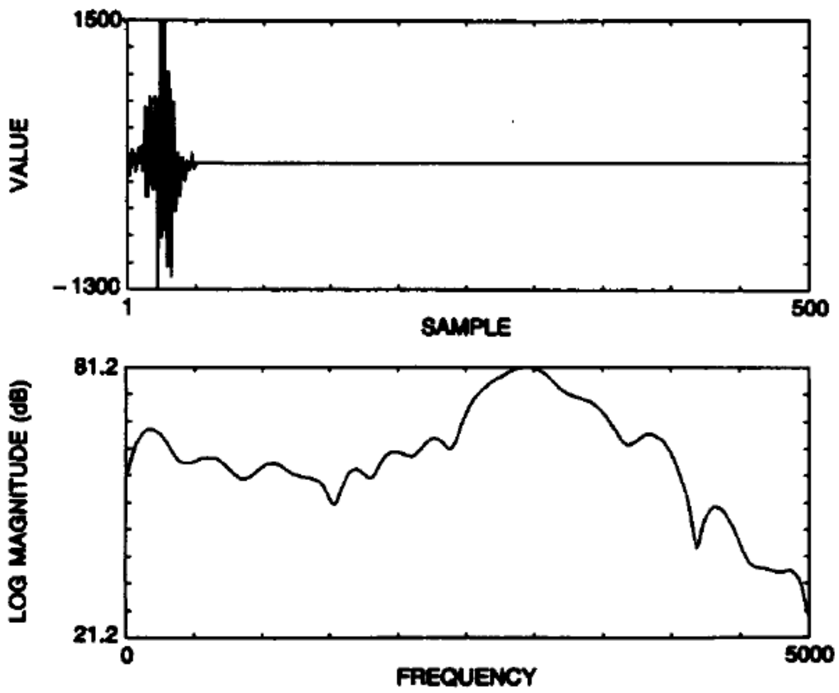


Figure (7) Short-time Fourier transform using a long (50 points or 5 msec) Hamming window on a section of unvoiced speech[4]

6.Hardware

In this work, after the voice recognized and matched with one found in database, the data will get from this database (Microsoft Access file) the name and send this data to microcontroller via RS232 port, and will display the information on LCD. The design of hardware is contains the main part which is microcontroller 8051 family (AT89s52), and the LCD 16*2 to display the name of person whom just record his/her voice, when then the system will allowed to an authenticator whom just sign in by his/her voice to be allowed to access. The purposed of using microcontroller gives benefits in many fields such security access in government and private sectors, controlling devices based on authorized person, person identification system, and in communication purposes. The database which has been used in this system is MS Office Access has limited of records reach to 32,600. The schematic circuit of the design hardware is shown in Figure (8).

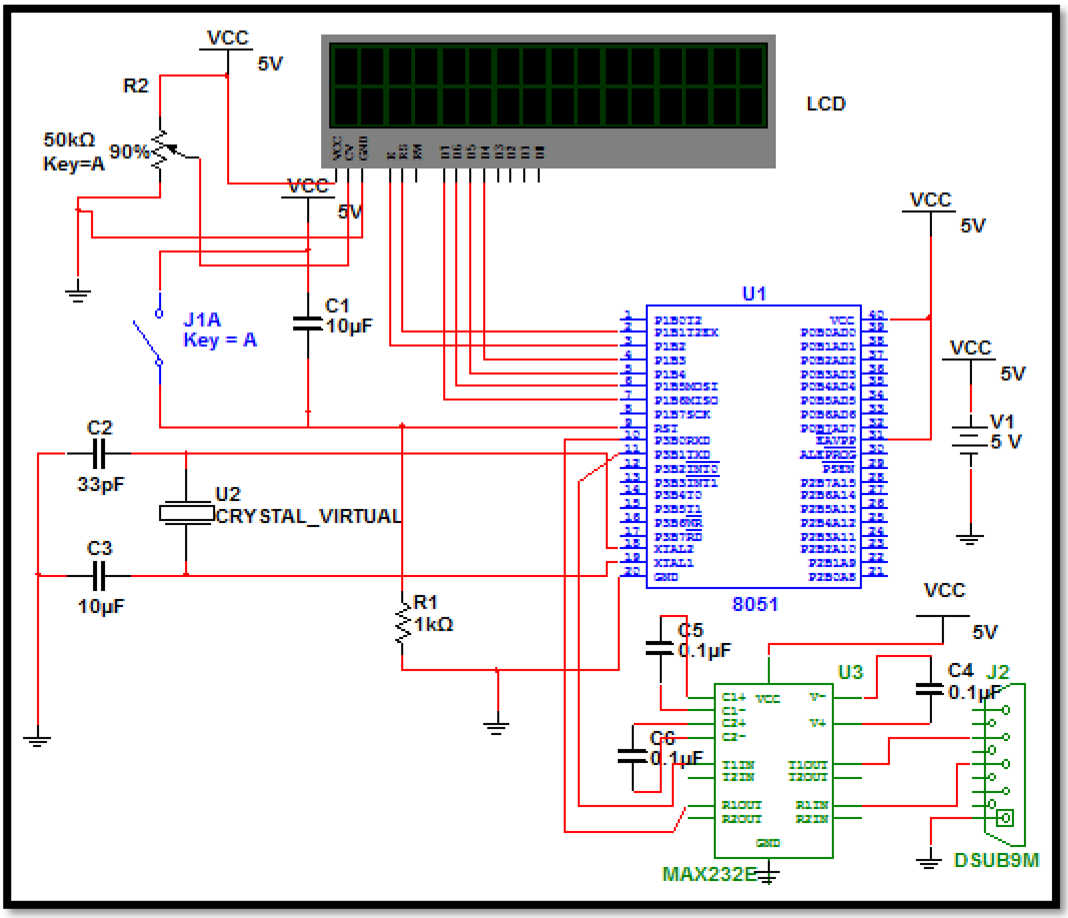


Figure (8) Schematic Circuit Design of Proposed System

7.Results

The proposed work includes software and hardware, and in the section below will be discussed in the procedure of the work.

Each person can be registered with his voice, and for each person has a number of samples found in database which gives more accuracy in recognition phase. In our system we've suggested a number of samples are (5) for each person. The word used to recognized is “alsalamelekum”. The training and recognition of voice signal has using Microsoft Visual Studio Visual Basic 6, besides that, using the SDK (Software Development Kit) of ASR (Automatic Speech Recognition) from Microsoft and we’ve developed the database of this SDK by applying Arabic isolated words. This work, is give an authority for persons whom has a registration then give an access after recognition get done.

For register step in voice recognition, the person needs to record his/her voice for 5 times, and save same name, and in case of login, the user need to record his/her voice to find the out the data in database for matched one, as shown in Figure (9).

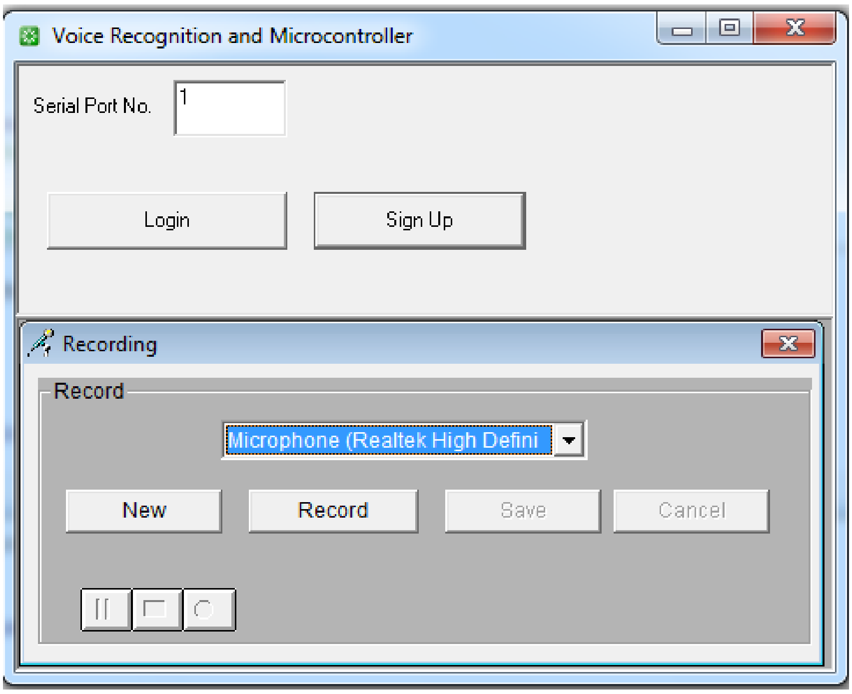


Figure (9)Login/Register form

After the matched of pitched of voice with one found in database, the message will send via serial port to circuit contain a microcontroller (AT89S52) which has been programmed to display the data that received from serial port of PC, as shown in Figure (10).

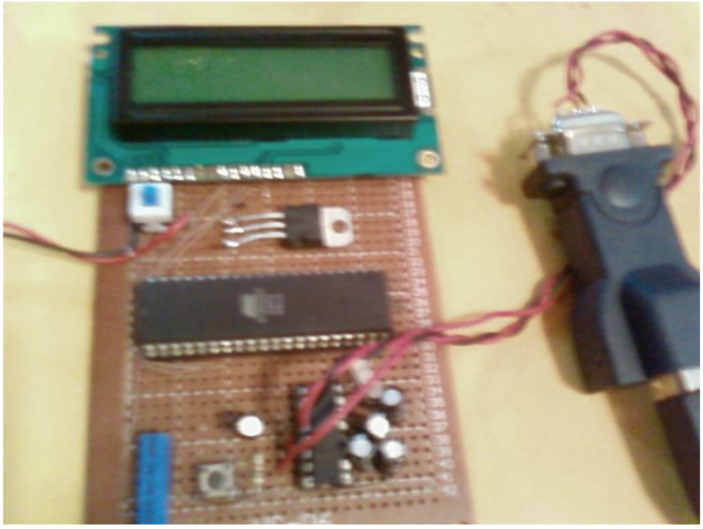


Figure (10) Proposed Circuit

Once the software starting, and the circuit powered up, LCD will show a statement “Voice Recognition” and kept show for a while, and this mean the circuit working well, and the connection to PC also working in right way as shown in Figure (11).

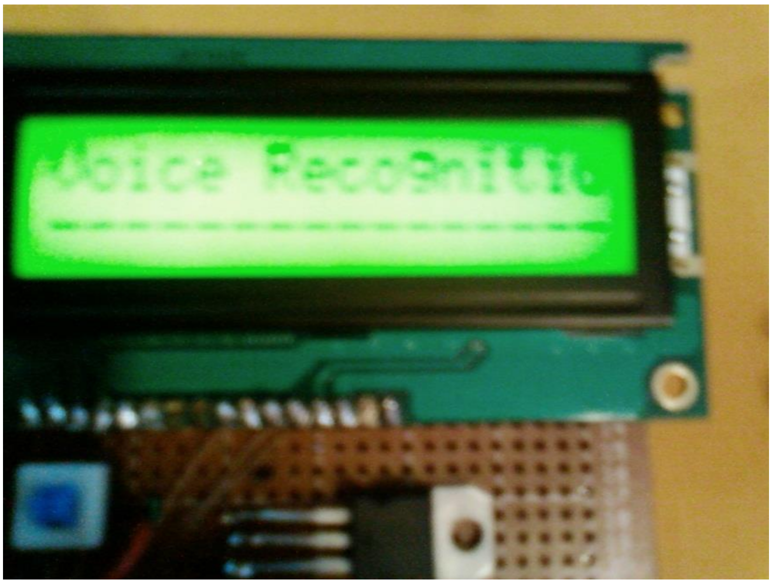


Figure (11) starting statement

When user/person starts recording his/her voice in application, if not recognized found, and matched not done, then the application will tell microcontroller that the person/user not recognized which in same time mean the person/user not authorized (not found his/her data of his/her voice in database), then microcontroller will show a message on LCD for this case as in Figure (12), and the person whom has been signed by his voice will be allowed to access which is method is the main goal of this research.

Figure (12) “not recognized” message on LCD

Otherwise, the recognized of user’s voice matched with once found in database, in this point, the stream of voice will capture by taking the pitches of voice, and then looking for best accuracy found in database, after this data found, the application will send a stream a name of person to microcontroller via serial port, and microcontroller will send the information to LCD, as shown in Figures (13,14 and 15). And these figures showing the names, which also proposed in our work, beside that, the person whom just signed in by the voice recognition software, has a authentication to be right to access the private or secured places and sectors.

Figure (13) “safaa” name show on LCD

Figure (14) “NOOR” name show on LCD

Figure (15) “mohammed” name show on LCD

The samples was captured are 5-6 samples per person, and give good accuracy for each one be using feature factor with one has been trained. Elapsed time measurement is suspended for typically 300 ms during serial via baud rate 9600b/s and data bit is 8.

8.Conclusion

This proposed work discussed the development of the voice recognition authentication system which can be used to display an authorized person information (name) on microcontroller circuit which is here the LCD, including the system could be used to unlock the door access or any locked system and allowed the authenticator to used his/her voice as an authentication system as an advanced security especially in biometric system.

In addition to display the name of authenticator and controlling the devices connected to output port of microcontroller, the design can be developed to send the information of persons (registered) to other security or location and department via different medium transfer (Bluetooth, Ethernet, web, and network), This work could be benefits for many security department which needs an authentication devices depend on biometric systems and one of them is a voice recognition. This system consists of voice recognition software and a microcontroller circuit that used AT89s52 with LCD 16*2, and the memory of the microcontroller is 8k byte.

Besides the achieving of the main objective, by using this system, it can help reduce any occurrence of getting access to whom not authorized to sensitive places and it offer a more safety way to access. Moreover this system could be fully equipped in a government offices especially in offices not anyone just an authorized need to get in.

9. References

1. Kenneth Revett; "Behavioral Biometrics, A Remote Access Approach"; John Wiley & Sons Ltd, 2008.
2. Sushil Jajodia; "Biometric User Authemtication for IT Security"; 2006.
3. Charles A. Shoniregun, Stephen Crosier;"Securing Biometrics Applications"; Springer Science+Busines Media,LLC. 2008.
4. Lawrence Rabiner, Biing-Hwang Juang; "Fundamentals of Speech Recognition"; 1993.
5. Slavy G. Mihov, Ratcho M. Ivanov, Angel N. Popov; "Denoising Speech Signals by Wavelet Transform"; ANNUAL JOURNAL OF ELECTRONICS, 2009.
6. Kuo, S.M.; Gan, W.S.; Asthana, P.; "Integrated noise reduction and acoustic echo cancellation in hands-free systems"; Intelligent Signal Processing and

Communication Systems, 2005. ISPACS 2005. Proceedings of 2005 International Symposium..; 2005

7. Chris Braithwaite, Fred Cowan, Hassan parchizadeh; "8051 Microcontroller, An Application Based Introduction"; Elsevier Publication, 2004.
8. Tobias Scheidat, Marcus Leich, Mark Alexander, and Claus Vielhauer; "Support Vector Machines for Dynamic Biometric Handwriting Classification"; University of Magdeburg, Department of Computer Science, Advanced Multimedia and Security Lab, Universit¨ atsplatz 2, D-39106 Magdeburg, Germany, 2009.
9. Hamid Jahankhani, David Lilburn Watson, Gianluigi Me; "Handbook of Electronic Security and Digital Forensics"; World Scientific; 2010.