DESIGN AND IMPLEMENTATION OF WIRELESS VOICE CONTROLLED MOBILE ROBOT

Dr. Ali Ahmed Abed
College of Engineering- University of Basrah
aaad_bah@yahoo.com

Dr. Abbas A. Jasim
College of Engineering- University of Basrah
abbas.a.jasim@ieee.org

Received 26 August 2015 Accepted 21 January 2016

ABSTRACT

This paper presents a technique for a speech recognizer used to control the motion of an intelligent automated mobile robot. The aim is to interact with the mobile robot using natural and direct communication techniques. The voice is processed to get proper and safe movement of a mobile robot and satisfying high recognition rate. Features are extracted from speech signal using Mel Frequency Cepstral Coefficients (MFCC). To realize feature matching, an efficient Dynamic Time Warping (DTW)-based speech recognition system is presented which is applicable for isolated words of Arabic language. The tested words are compared to a trained database using this DTW algorithm. On the other side, the mobile robot is designed with two servo motors as driving actuators. These actuators are controlled by L298 motor driver circuit. The control algorithm is programmed and downloaded into a PIC18F45K22 microcontroller which is interfaced to a USB port of a 10" notebook computer. The robot proves a capability of understanding the full meaning of the five Arabic speech commands that steer it forward, backward, right, left, or stop.

Keywords: Arabic speech recognizer, Mel Frequency Cepstral Coefficients, dynamic time warping, Pattern Recognition , Mobile robot.

تصميم وتنفيذ روبوت متحرك لاسلكي مسيطر عليه بالصوت

د. علي احمد عبد
كلية الهندسة-جامعة البصرة

د. عباس عبد الامير جاسم
كلية الهندسة-جامعة البصرة

الخلاصة

يقدم البحث طريقة لبناء مميز كلام يستخدم للسيطرة على حركة روبوت متحرك آلي وذكي يستطيع التفاعل وفهم لغة الكلام الطبيعية بصورة مباشرة. يقدم البحث الخطوات التفصيلية اللازمة لمعالجة الاشارة الصوتية بما يضمن نسبة تمييز عالية تؤدي الى حركة آمنة وطبيعية للروبوت. الخوارزميات المستخدمة للمعالجة الصوتية هي: خوارزمية MFCC وخوارزمية معتمدة على DTW تستخدم لتمييز الكلمات العربية المنفصلة. تعتمد عملية التمييز على مقارنة الكلمات الاختبارية مع الكلمات المدربة مسبقاً والمخزونة في قاعدة بيانات مسبقة. من جانب آخر، فقد تم بناء روبوت متحرك ثنائي المحركات من نوع servo تساق بواسطة مسيطر نوع L298 وذلك لتحقيق خوارزمية السيطرة الصوتية. تمت موائمة الروبوت مع مميز الكلام من خلال معالج (مايكروكونترولر) نوع

PIC18F45K22 كدائرة بينية صممت بشكل كامل لهذا الغرض. واخيراً تم اختبار المنظومة بأكملها من خلال توجيه الروبوت باستخدام خمس كلمات عربية هي: امام، خلف، يمين، يسار، قف والتي بواسطتها يمكن توجيه الروبوت على المسار المطلوب.

الكلمات المفتاحية: مميز كلام عربي، معاملات تردد ميل، معاملات الوقت الديناميكي، تمييز الانماط ، الروبوت المتحرك

NOMENCLATURE

ANN	Artificial Neural- Network
Dist(x,y)	Euclidean distance between two points
DTW	Dynamic Time Warping
F	Tone frequency in Hz
FFT	Fast Fourier Transform
F_{mel}	Mel frequency in Hz
GD	Global Distance
HMM	Hidden Markov models
K	Number of frames
L	Number of samples in each frame
LD	Local Distance
LPC	Linear Predictive Coding
M	Number of samples that separated frames
MFCC	Mel Frequency Cepstral Coefficients
RR	Recognition Rate
V	Voice Activity Detection
X	Sequence feature vector in n dimensional space
Y	Another sequence feature vector in n dimensional space

1. INTRODUCTION

The Arabic language is the fifth widely used language world-wide since there are at least 200 million people speak Arabic, (**Khalid, 2013)**. There are little researches in speech recognition field that deal with Arabic as compared to English or Japanese. The Arabic language has monosyllabic and polysyllabic words with two categories of phonemes: pharyngeal and emphatic, which found in all Semitic languages, (**Al-Zabibi, 1990) and (Alkhouli , 1990)**. The automatic speech recognition, which got a good attention for many decades, allows a computer to recognize spoken words inputted by a mike. Speech recognizers are used in many applications such as: interacting with deaf people, healthcare, home automation, robotics, etc. There are a large number of approaches for speech recognition such as: Dynamic time warping (DTW), Artificial Neural- Network (ANN), Hidden Markov models (HMM), etc. In this work, an efficient DTW-based speech recognition system for isolated Arabic words is given as a feature matching algorithm and a Mel Frequency Cepstral

Coefficient (MFCC) approach is used as a feature extraction approach because of its robustness and effectiveness compared to other well-known methods like Linear Predictive Coding (LPC), (**Lindasalwa, 2010)**. After that a mobile robot is designed, as will be explained in the subsequent sections, and controlled by the designed speech recognizer to get a complete speech controlled system suitable for different applications. It is desired to command the mobile robot by voice via special interface that plays a significant role as a master control circuit for the servo motors of the robot. In voice control system, a difficulty may appear in the control circuit leading to a recognition error, which means that the recognized command is interpreted as opposite command. For example "Left" is interpreted as "Right" especially in languages with very high acoustic similarity like Polish. This problem is not significant in Arabic when using the direction words because they differ completely in pronunciation.

Unlike other languages, Arabic language is characterized by having tremendous dialectical variety, diacritic text material, morphological complexity which may lead to some challenges against having a highly accurate Arabic recognizer. In the work of (**Jean-Marc , 2007)**, the voice of the speaker depends on the distance and azimuth. The work satisfied a distance of 2m and azimuth range of 10^o to 90^o. In 2010, two voice recognition algorithms which are MFCC and DTW is built, evaluated and compared with other techniques to prove their effectiveness, (**Lindasalwa, 2010)**. In 2011, (**Ahmed,2011)** had proposed a technique called (multiredgilet transform) with neural network to control the motion of a wheelchair dedicated for handicapped people. The work presented by **(Rachna, 2011)** is to build a microcontroller-based mobile robot controlled with speech. He studied various factors such as noise and distance factor for his speech recognition system. **(Khalid, 2013)** suggested DTW, MFCC and voice activity detection (VAD) for isolated words of Arabic language but with unsufficient recognition rates. In our work, we built a speech recognizer using MFCC, DTW, and VAD for five Arabic words and a high recognition rates are satisfied without depending on azimuth and the distance is limited by the wireless transmission distance. This speech recognition system is used to control the motion planning of an autonomous mobile robot designed completely to get a voice controlled robotic system.

The rest of the paper is organized as follows: Section 2 is concerned with the explanation of our speech recognition system with all its stages. In section 3, the complete design of the mobile robot with the used components is explained. Section 4 provides the software structure of the overall system. In section 5, the obtained results and verification are given with some required discussion. Section 6 summarizes the main conclusions.

2. THE VOICE RECOGNITION SYSTEM

The presented Arabic speech recognition system consists of the following stages:

A. Preprocessing

This stage is important to enhance the recorded speech signal characteristics by removing noise leading to obtain a high quality recorded speech. The high frequency contents of the input signal are emphasized by a first order FIR filter (implemented in software) to flatten the signal spectrum. Also, this stage should overcome the problem of using different types of microphones and different speaking loudness. This stage is hidden in the first stage "Read the voice input" of **Figure1.**

B. Voice Activity Detection (VAD)

Another problem that affects the performance of the speech recognizer is detecting the start and end points of the voice signal, (**Khalid, 2013)**. The speech signal is segmented into spaced frames of 10ms

width. After that, short-term power and zero-crossing rate are used to detect the speech/non-speech regions. It is clear that short-term power is increased in speech regions while zero-crossing rate is increased in non-speech regions. Hence, these two techniques give a good indication of speech appearance.

C. Feature Extraction

- Framing: The speech signal is segmented into K frames of L samples for each one. The adjacent frames are separated by M samples (M<L). In this work, values of L=256 and M=87 samples are chosen.

- Hamming Window: It is applied to the above framed signal to reduce the discontinuity at both ends of each frame.

- Fast Fourier Transform (FFT): It is used to convert the above windowed signal from time domain to frequency domain to prepare it for the next stage.

- Mel Filter Bank: Since human hearing is less sensitive to frequencies higher than 1KHz, a Mel-scale is used so that for each tone with a frequency F (in Hz), a subjective pitch is measured on a Mel-scale according to the following equation, (**Plannerer, 2005)**:

$$F_{mel} = Log_{10}(1 + \frac{F}{700}) \quad (1)$$

  After calculating of the magnitude of the resulting FFT signal and using the Mel-scale filter bank, which consists of 24 triangular band pass filters having an equal spacing before 1 KHz and logarithmic scale after 1 KHz, the Mel spectrum coefficients are found as the summation of the filtered results.

- Inverse Discrete Cosine Transform (IDCT): It is used to return back to time domain. But before that, the logarithm of the magnitude of the output of Mel-filter bank is computed since logarithm compresses the dynamic range of values.

- Liftering: is a filtering in the spectrum domain that used to extract the vocal tract cepstrum. Some cepstrum coefficients at the end can be dropped. One can use the first 12 coefficients for each frame and neglect the others.

D. Database Collection

A feature database of spoken Arabic words for pattern matching process is created and stored in computer memory as a collection of template vectors. Feature database is built by recording several utterances for each one of the five target words collected from different speakers (male and female) during the training process. After each target word is recorded, preprocessing, VAD, and feature extraction processes are performed and the final feature vector pattern is obtained and stored in the database. For each of the adopted Arabic control target words, fifteen patterns, template vectors were created in this manner. Thus the training database of feature vectors contains 75 patterns arranged as five cluster one for each control word. All recording sessions are done under normal environment with 8 kHz sampling frequency and 8 bit per sample mono channel.

E. Feature (Pattern) Matching

Pattern matching process is performed in the runtime phase of the proposed system to find the template pattern that best fit to the spoken word. Pattern matching is done in online manner. The test spoken word is recorded in the same environment used for training phase. The test feature vector is then found after applying preprocessing, VAD, and feature extraction steps. Test feature vector is compared with all template vectors in the training database vector by vector. The best fit template vector is selected. The target word for which the best fit template belongs in is said to be recognized.

The time alignment of different utterances is the core problem for distance measurement in the speech recognizer presented in this paper. A small shift leads to incorrect identification. Dynamic Time Warping (DTW) is an efficient method to solve this problem, **(Shivanker, 2013)**. DTW algorithm aims to align two sequences of feature vectors by warping the time axis repetitively until an optimal match is found. In other words, it performs a piece wise linear mapping of the time axis to align both signals. If two sequences of feature vector in n-dimensional space is considered:

$$x = [x1, x2, \ldots . xn] \;\; and \;\; y = [y1, y2, \ldots ., yn] \tag{2}$$

Then they are aligned on the sides of a grid. The distance between two points is obtained by the Euclidean distance:

$$Dist(x, y) = |x - y| = \sqrt{[(x_1 - y_1)^2 + (x_2 - y_2)^2 + . \ldots (x_n - y_n)^2]} \tag{3}$$

The best match or alignment between these two sequences is the path through the grid, which minimizes the total distance between them. The overall distance (global distance) is calculated by finding and going through all the possible routes through the grid, each one compute the overall distance. Then global distance is obtained by the following recursive equation:

$$GD_{x,y} = LD_{x,y} + \min(GD_{x-1,y-1}, GD_{x-1,y}, GD_{x,y-1}) \tag{4}$$

GD=Global (Overall) distance, LD=Local (Euclidean) Distance.

3. THE MOBILE ROBOT SYSTEM

A. The Robot Vehicle

The used vehicle is a new breed of tracked robot chassis designed specifically for mobile robot projects, (**Rover 5, 2013)** as shown in **Figure 2**. It has two actuated wheels and two passive wheels used for balance. The clearance can be adjusted by rotating the gearboxes in 5-degree to maintain tension as required. An optical quadrature encoder that gives 1000 pulses over 3 revolutions of the output shaft is embedded. It has two motors and encoders and 4 noise suppression coils.

B. The Vehicle Motors and Motors' Driver

The two motors used in this car are servo motors with the following specifications: motor rated voltage is 7.2V, stall current is 2.5A, output shaft stall torque is 10Kg/cm, and speed is 1Km/hr. The Motor driver is based on the L298, which is a dual H-bridge driver designed to drive inductive loads such as

DC and stepping motors, (**User's Guide, 2012)**. It lets you drive two DC motors, controlling the speed and direction of each one independently. The connection diagram of the driver with the two motors and a real photo of it are shown in **Figure 3**.

C. The Robot Controller Circuit

The brain of the robot that is considered as a decision making element and controls the motion planning of the robot is the microcontroller. PIC18F45k22, (**Microchip Technology Inc., 2012),** is one of the most popular microcontrollers that could be adopted in robotic systems. In this project, we used the EasyPIC7 development board, **(User's Guide Mikroelektronika, 2013)**, which will save effort and time for building a complete controller and interface circuit. It has an RS232 port for UART serial communication with the computer as shown in **Figure 4**. Since there is no RS232 serial port in our notebook computer, a USB-to-RS232 converter is used. To facilitate the serial communication between the speech recognizer designed with MATLAB and the microcontroller-based interface circuit of the mobile robot, the following code is embedded within the speech recognizer:

```
command=?;
s=serial('com1');
fopen(s);
fwrite(s,command);
fclose(s);
delete (s);
```

Where: "command" is the value that corresponds to the desired direction given acoustically by a person. "com1" is the serial port number available, which may take other numbers depending on the computer.

D. The Microphone

A quality microphone is the key when utilizing automatic speech recognition (ASR). In most cases, a desktop microphone is not sufficient because they may pick up more ambient noise that gives ASR programs a hard time. Hand held microphones are also not the best choice as they can be cumbersome to pick up all the time. While they do limit the amount of ambient noise, they are most useful in applications that require changing speakers frequently. The best choice and the most common is the headset style. It allows the ambient noise to be minimized, while allowing us to have the microphone at the tip of our tongue all the time. Logitech Wireless Headset H600 (shown in **Figure 4)** is found experimentally to be the best choice, so it is used to control the robot while it is moving. The headset will send the voice orders to the robot as a response to a low sound tone that tell us when one can start speaking. This headset consists of the following parts: Nano USB receiver plugin in the notebook computer, Noise-canceling microphone, 2.4 GHz wireless transmitter, Simple on-ear control, and six-hour rechargeable battery.

E. The NoteBook Computer

It is a 10 inches laptop computer with USB ports for serial communication with the microcontroller of the robot and for the nano receiver of the headset. A MATLAB R2010 is installed to be the environment of the speech recognizer that takes the input voice from our wireless mike and produces the corresponding commands outputted via the USB port to the interface circuit reaching to the microcontroller which in turn interprets these commands and produces the suitable action for the motors to begin the planned motion. The block diagram of the overall computer control hardware system is shown in **Figure 5**. Of course, this laptop has a relatively large size the matter that may restrict the robot usage to some applications but it is desirable to other applications that need a computer controlled robots such as moving robots in space **(Que, 2011),** and industrial robots (**Yoshioka , 2014)**.

4. THE OVERALL SOFTWARE STRUCTURE

Voice recognition works based on the premise that a person voice exhibits characteristics that are unique among different speakers. The signal during training and testing sessions can be greatly different due to many factors such as: people voice change with time, health condition, speaking rate and also acoustical noise and variation recording microphone. To explain the software stages starting from voice input to commands output, a flowchart is shown in **Figure 1**. The microcontroller should be programmed in order to receive the commands from the serial port and use them to actuate (drive) motors accordingly. MicroC is a powerful, feature rich development tool for PIC micros designed to provide an easy solution for developing applications in embedded systems and control, **(User's Manual of Mikroelektronika, 2012)**. The C program downloaded inside our PIC is given below:

```
char uart_rd;
void main() {
ANSELC = 0;
UART1_Init(9600);
Delay_ms(100);
TRISB = 0x00;
LATB = 0x00;
while (1) {
if (UART1_Data_Ready()) {
uart_rd = UART1_Read();
 LATB =uart_rd; }}}
```

5. RESULTS AND DISCUSSION

In order to evaluate the performance of the presented speech recognizer, recorded samples are separated into training and testing sets. The recognition rate of each Arabic word can be calculated by the following equation, **(Khalid, 2013)**:

$$RR = \frac{No.of\ correctly\ recognized\ words}{No.of\ tested\ words} * 100\% \quad (5)$$

Each word of the five Arabic words is trained fifteen times to provide sufficient number of template vectors for four different cases of experimental work. Hence, the number of stored templates is 75. In the test phase, each of the five words is tested at run time of the mobile robot that operates 30 times. Mobile robot motion with respect to the corresponding voice command is monitored and recognition accuracy (recognition rate) is computed. **Table 1** shows the RR for a sample of tested words in the database which have been already recorded. While **Table 2** shows the elapsed time required to perform the matching process for four different cases use 3, 5, 10, or 15 template vectors for each word. From **Table 1**, three template feature vectors per target word give poor recognition rate. RR is significantly increased with 5 template vectors. Using 10 template vectors give very good RR. Increasing the number of template vectors to 15 led to little increasing in the recognition rate as shown in **Figure 6**. On the other side, the time needed for DTW to perform the matching process for the four defined cases seems to vary in approximately linear form as shown in **Figure 7**. From the held experiment with results given in **Table 1** and **Table 2**, using ten template feature vectors per word gives very good RR with moderate time. The input voice waveforms of the five spoken words are shown in **Figures 8-12**.

Now, the presented speech system is applied to the motion of the designed mobile robot which is tested with the five Arabic words in **Table 1**. It is possible with a series of voice commands to steer

the robot along a desired trajectory to its target. Also, controlling the voice commands leads to control motion and avoid obstacles in the way of the robot. To satisfy a linear path for the robot, only the isolated words "امام" or "خلف" is used (both wheels are rotated clockwise or anti-clockwise) as shown in **Figure 13**. Circular path of the robot is obtained by a single isolated voice command "يسار" or "يمين" according to the desired direction. **Figure 14** shows circular motion of the robot in clockwise direction using voice command "يمين" in which right motor is "off" and left motor is "on".

6. CONCLUSIONS

Using a simple and efficient automatic speech recognition technique for isolated Arabic words to satisfy the motion planning of mobile robots is the interest of this paper. The extracted features are stored in a .mat file using MFCC algorithm. The experiment results are analyzed with the aid of MATLAB and proved good efficiency especially when applied with the designed mobile robot. This process can be extended for a number of speakers. Also, the work proved that the DTW is a sufficient nonlinear feature matching technique in speech identification with minimal error rates and fast computing speed the matters that are very important in robotics. Generating database stores feature vectors to be matched with run time test words rather than storing the whole speech signal saved memory space and made matching process very fast. The processing units (the notebook and the microcontroller) are directly attached to the mobile robot in one package that made the design representing a complete autonomous robot. Voice commands are given using wireless microphone results in flexible movement of the robot. The range of movement of the implemented mobile robot is only restricted by the range of the wireless microphone. The physical range of movement can be extended simply by speaker movement because there is no fixed station is used for any processing purpose.

REFERENCES

Ahmed Q. A., 2011, Controlled of mobile robots by using speech recognition, Journal of Babylon University, Applied Science, Vol. 3, issue 19.

Alkhouli M., 1990, Alaswaat Alaghawaiyah, Dar Alfalah, Amman, Jordan, (Arabic reference).

Al-Zabibi M., 1990, An acoustic-phonetic approach in automatic Arabic speech recognition, The British Library in Association with UMI, UK.

Jean-Marc V., Shun'ichi Y., Jean R., Francois M., Kazuhiro N., and Hiroshi G., 2007, Robust recognition of simultaneous speech by a mobile robot, IEEE Transactions on Robotics, Vol.23, No.4. P. 742-752.

Khalid A.D., Ala F., Iyad F., Baraa A., and Saed W., 2013, Efficient DTW-Based speech recognition system for isolated words of Arabic language, World Academy of Science, Engineering and Technology, Vol. 7, P. 106-113, Iraq.

Lindasalwa M., Mumtaj Begam and I. Elamvazuthi, 2010, Voice recognition algorithms using Mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques, Journal of Computing, Vol.2, issue 3, ISSN 2151-9617.

Microchip Technology Inc., 2012, Low power, high performance microcontrollers.

Plannerer B., 2005, An Introduction to speech recognition, Bernd Plannerer, Munich, Germany.

Que D., Jian Yang ; Bo Wei ; Hui Li ; Zhihong Jiang ; Danfeng Li ; Hongjie Li ; Qiang Huang, 2011, A method on trajectory plan for humanoid space robot , Robotics and Biomimetic (ROBIO), IEEE International Conference, P. 281-286. Thailand.

Rachna J., and Saxena S., 2011, Voice automated mobile robot, International Journal of Computer Applications (0975-8887), Vol. 16, No. 2, India.

Rover 5, available at: https://www.sparkfun.com/products/10336

Shivanker D., Geeta N., and Poonam P., 2007, Isolated speech recognition using MFCC and DTW, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol.2, Issue 8, 2013, ISO 3297: certified organization, India.

User's Guide, 2012, L298 dual H-bridge motor driver.

User's Manual, 2012, MicroC: C compiler for Microchip PIC microcontrollers, Mikroelektronika, Belgrade.

User's Guide, 2013, EasyPIC7", Mikroelectronika, Belgrade.

Yoshioka, T. ; Shimada, N. ; Ohishi, K. ; Miyazaki, T. ; Yokokura, Y., 2014, Link-coupled vibration suppression control considering product of inertia for industrial robots, Advanced Motion Control (AMC), IEEE 13th International Workshop, P. 675 – 680, Yokohama, Japan.

Table (1): Recognition rates for different feature sets

Tested Arabic words	Transcription	English writing	RR 3 5 10 15
امام	Amam	Forward	43 67 90 93
خلف	Khalf	Backward	40 63 80 80
يمين	Yemeen	Right	46 73 86 90
يسار	Yesar	Left	53 83 90 93
قف	Qif	Stop	46 67 83 86

Table (2): Feature matching time for different cases

No. templates	Comparison Time (msec)
3	5.8
5	9.4
10	15.2
15	23.7

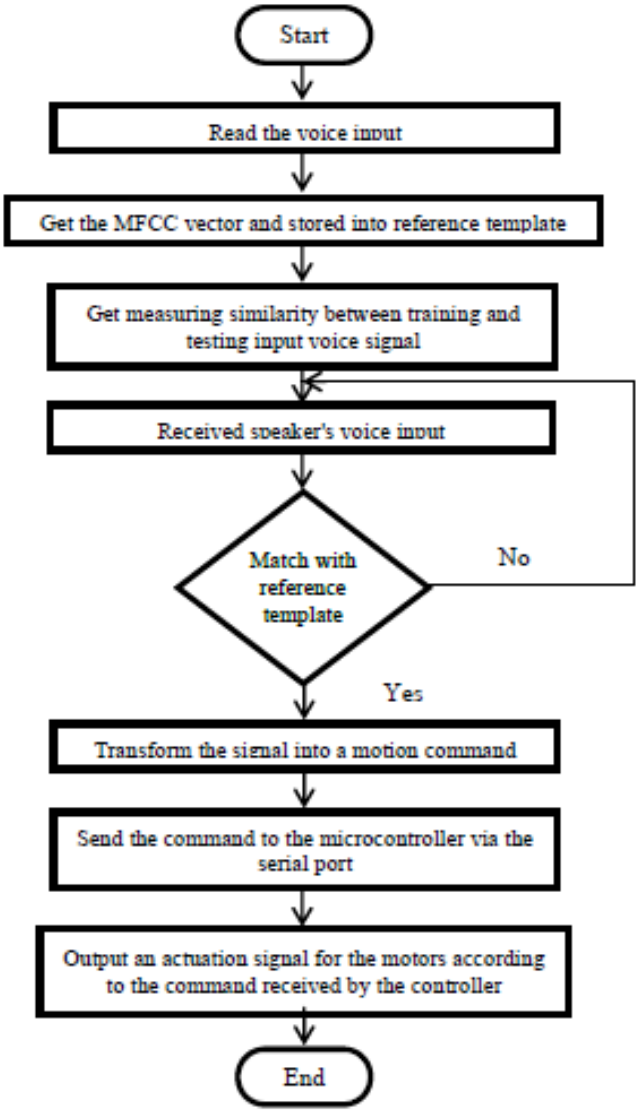

Figure (1): The overall software system

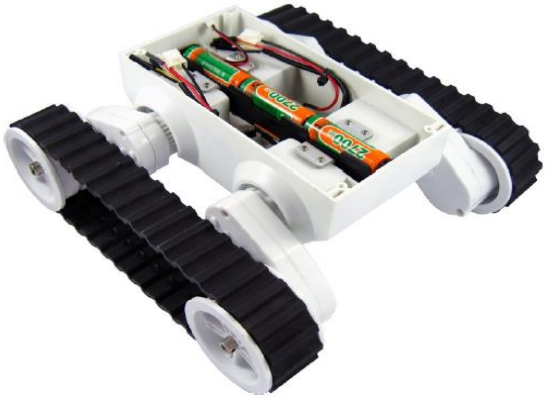

Figure (2): The robot chassis

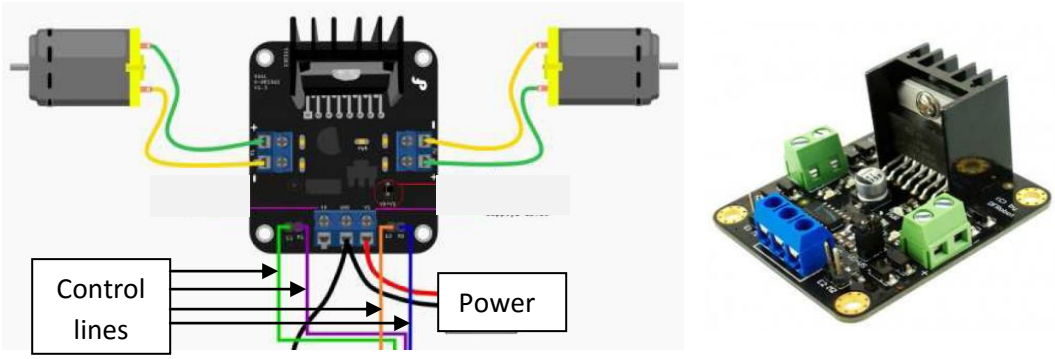

Figure (3): Motors driver connection diagram and its real photo

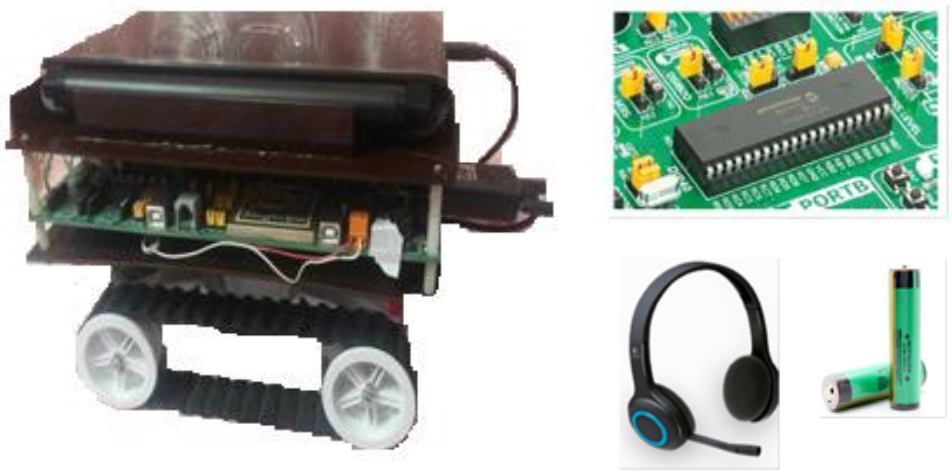

Figure (4): The overall designed listening robot with its main parts

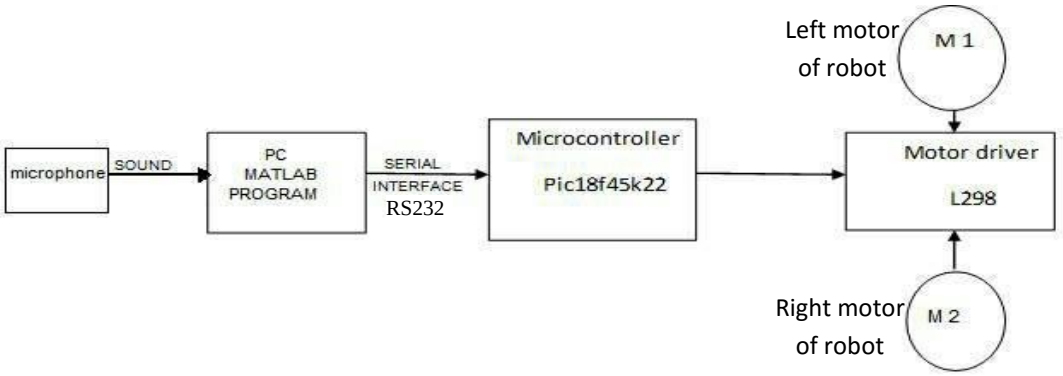

Figure (5): The overall hardware system

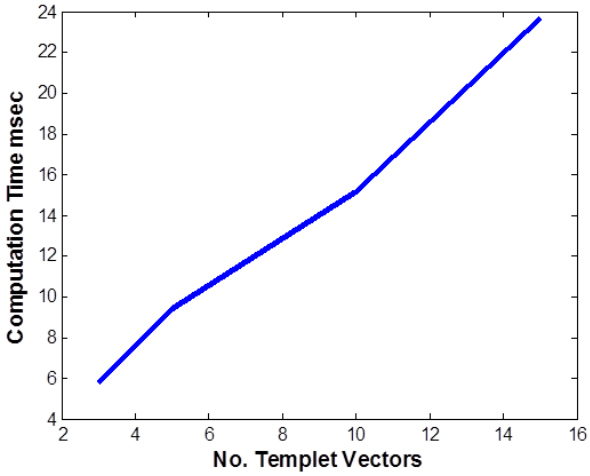

Figure (6): Relationship between RR and no. of template vectors

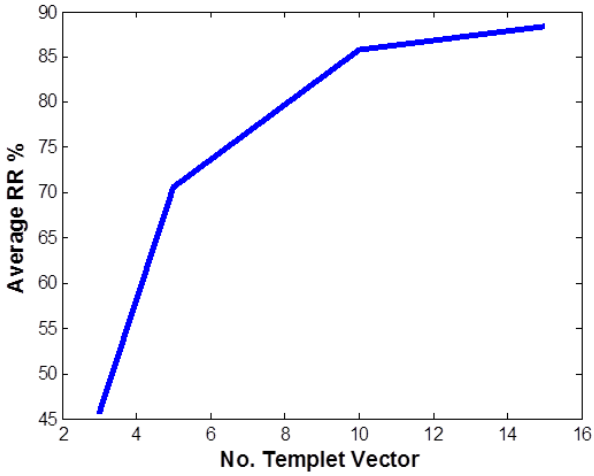

Figure (7): Relationship between time and no. of template vector

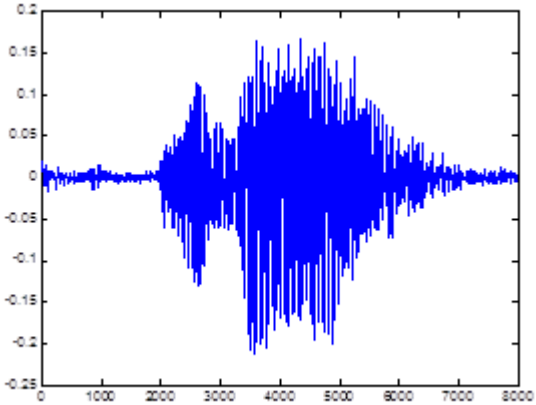

Figure (8): The input voice signal of "امام" word

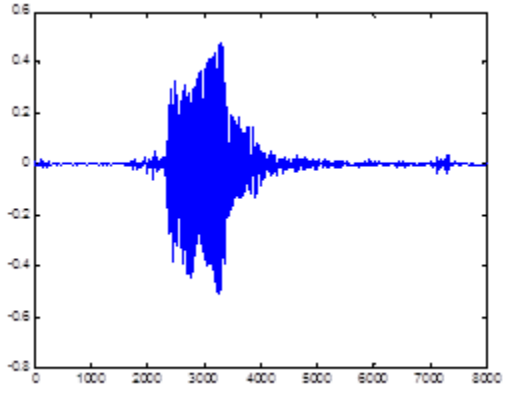

Figure (9). The input voice signal of "خلف" word

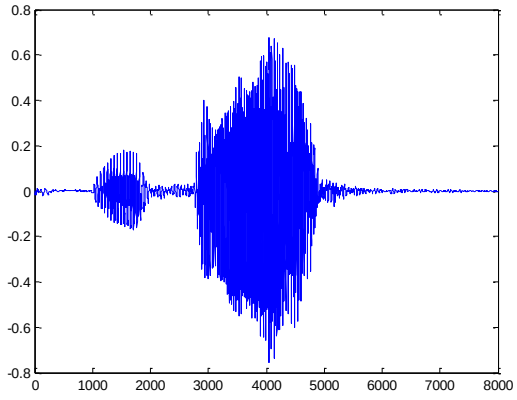

Figure (10): The input voice signal of "يسار" **word**

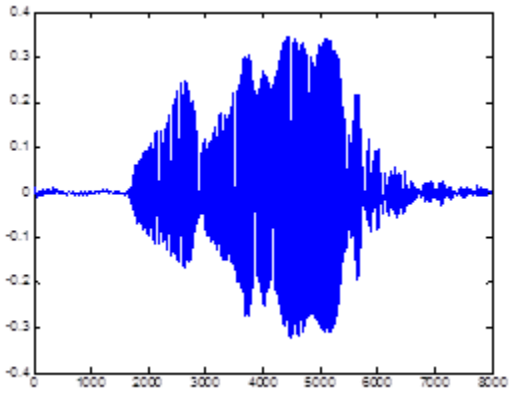

Figure (11): The input voice signal of "يمين" **word**

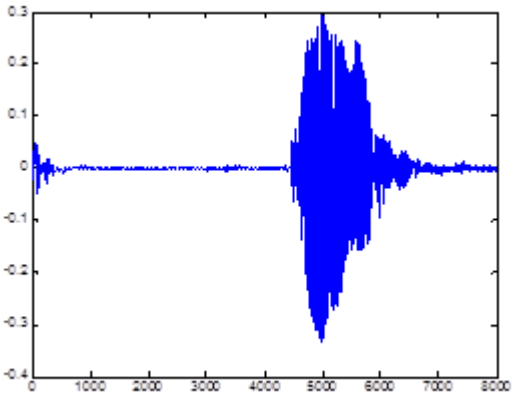

Figure (12): The input voice signal of "قف" **word**

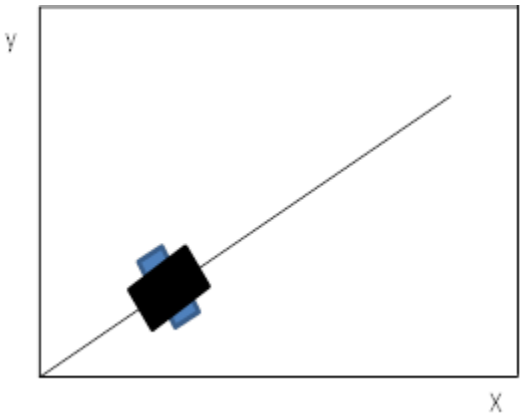

Figure (13): Linear path navigation of the mobile robot

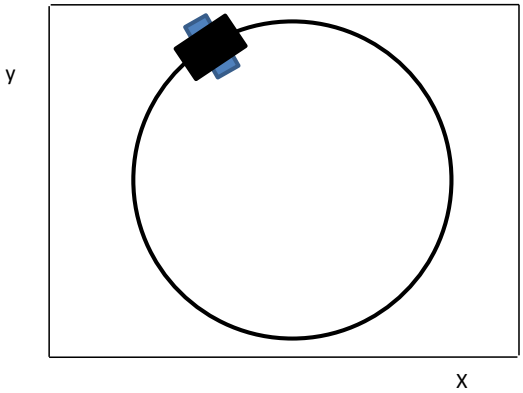

Figure (14): The circular path navigation of the mobile robot