



# IMPLEMENTATION OF SPEECH RECOGNITION HOME CONTROL SYSTEM USING ARDUINO

Nurul Fadzilah Hasan<sup>1</sup>, Mohd Ruzaimi Mat Rejab<sup>2</sup> and Nurul Hidayah Sapor<sup>1</sup>

<sup>1</sup>Fakulti Kejuruteraan Elektrik & Elektronik, Universiti Malaysia Pahang, Kuantan, Pahang, Malaysia

<sup>2</sup>Fakulti Kejuruteraan Mekanikal, Universiti Malaysia Pahang, Kuantan, Pahang, Malaysia

E-mail: [nurulfadzilah@ump.edu.my](mailto:nurulfadzilah@ump.edu.my)

## ABSTRACT

Electronically control of household activities has long been explored in various ways. From electronic remote control using infra-red sensors, to voice-controlled application, we are continuously trying to find a better way to control electrical and electronic devices to ease our daily life. This paper presents the development of a low cost remote home control system using speech recognition. The system focused on controlling fan and lamp wirelessly by applying speech recognition into the system, with Arduino Uno as the controller. Two different units were developed: the main control unit which also act as the transmitter and the receiver unit which controls the fan and lamp. The main control unit accept voice command from the user and convert it into text by using Easy VR shield. The signal is then transmitted to the receiver using RF signal. With the use of RF technology, the system is able to work wirelessly. This system is hoped to be able to help people to use the electronics devices effectively besides provide improved convenience and comfort to user especially for elderly and disabled who live alone and help them to be more independent.

**Keywords:** arduino uno, easyVR, home control.

## INTRODUCTION

Smart home technology has long been explored, since the early 1980s when the “intelligent building” concept was used. The concept proposed an intelligent implementation of consumer electronic devices, electrical equipment, and security devices. It aimed for the automation of domestic tasks, easy communication, and human-friendly control, as well as safety [1]. Some smart house systems allow home control via LAN (Local Area Network), and WAN (Wide Area Network). This type of home control system allows devices to be controlled through computers and Android smart phones at the same time [2]. There are many methods available to control electric appliances at home. The most common way is by using an electronic remote control. Home appliances can also be controlled via voice control, or it can also implement home automation based on predefined user profiles or independent sensors. In this paper, voice control method is used to control home electric devices.

Voice control method offers a more user interactive approach in delivering control commands [3]. By applying speech recognition system, a system can be developed to help user control devices remotely. Voice control system for ZigBee based home automation has been introduced in journal of “ZigBee based voice Controlled Wireless Smart Home System”. Speaker independent automatic speech recognition technique has been used. In this system Zig Bee network receives voice command as input to an ARM9 controller, which converts the data into a required format to be used in the microcontroller. Finally, the system generates some control characters to switch ON/OFF the home appliances [4].

There are two types of speech recognition system. They are speaker-dependent and speaker-independent system. Speaker-dependent system is designed for a

specific speaker that works by learning the unique characteristics of a single person's voice [5]. It is also known as voice recognition. New users must first “train” the software by speaking to it, so the computer will analysed how the person talks. This system is useful as the security system. Speaker-independent systems on the other hand, require no training phase with data of users, and are desirable to many applications where training is difficult to conduct [6].

Arduino UNO is a multi-purpose microcontroller board based on the ATmega328P. It has 14 digital input/output pins and 6 analog inputs. Each of the 14 digital pins on the Uno can be used as an input or output. An Arduino Uno board can either be powered via USB connection or with an external power supply (AC-to-DC adapter or battery). Leads from a battery can also be inserted in the Gnd and Vin pin headers of the power connector. The board can operate on an external supply of 6 to 20 volts [7] [8]. Arduino Uno can communicate with other devices such as a computer, another Arduino board, or any other types of microcontrollers. Its software serial library allows serial communication on any of the Uno's digital pins [9].

In the journal of title “Improved Authentication Using Arduino Based Voice & Iris Recognition Technology”, a voice recognition system is proposed to build as security function. The Arduino board as important role integrate with EasyVR Shield [10]. The proposed system is using speaker-dependent system to train the password command. In this project also using the password command rolled as the security of the system. Therefore, the usage of both boards can be used to develop a speech recognition system. The model of the system in EasyVR Module is HMM Model.



**Table-1.** Types of commands in EasyVR commander software.

| Types of Group | Functions  |
|----------------|--|
| Trigger        | A special group that can add user-defines trigger word. The trigger words are used to start the recognition process.   |
| Group          | A speaker-defined group where recognized users' voice that trained in this group only. This group did not use in this project.   |
| Sound Table    | A special group to play the sound for system. There is only one built-in sound in this group which is beep sound. User can import any sound in format of .WAV to use it for the system.          |
| Password       | This is a special group for "vocal passwords" which can be used for up to five words by using Speaker Verification (SV) technology. It is suitable for security system that needs fully guarded. |
| Word set       | Word set consists of sets of built-in speaker independent command. Total number of command is 32 where built-in in three different groups.   |

In this paper, trigger and word set group are used for the voice command. All the voice command and the output of the system are listed in Table-2. In trigger command, the word of "System On" is trained. It is used to activate the system. For the usage of Word set Command in this project, all the three word set group are used as command word to control the output device.

**Table-2.** Voice command and output of the system.

| Trigger /Word set group | Voice Command | Output   |
|-------------------------|---------------|--|
| Trigger                 | System On     | System is activated                                    |
| Word set 1              | Action        | Trigger the lamp                                       |
|                         | Turn          | Trigger the fan  |
|                         | Run           | The lamp and fan are ON<br>*The fan is ON with speed 1 |
|                         | Stop          | System is deactivated                                  |
| Word set 2              | Up            | The lamp is ON   |
|                         | Down          | The lamp is OFF  |
| Word set 3              | Zero          | The fan is OFF   |
|                         | One           | The fan is ON with speed 1                             |
|                         | Two           | The fan is ON with speed 2                             |
|                         | Three         | The fan is ON with speed 3                             |

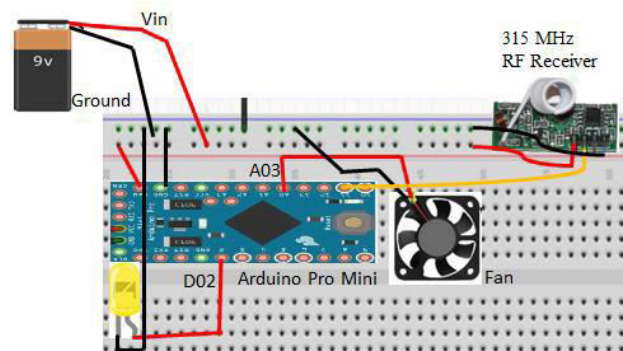
Output from EasyVR module is transmitted to receiver module via a simple wireless data link using 315 MHz RF transmitter. It is a low cost RF transmitter that can transmit signal up to 100 meters which provide an adequate one-way data communication. The transmitting frequency is 315MHz and an external antenna use in this project is 13cm single-core line. These features are enough for this system because the system is intended to be implemented for a standard room size. However, the antenna design, working environment and supply voltage seriously impact the effective distance.

The operating voltage is 3.5V to 12V and the operating current is 4mA @ 5V while 15mA @ 9V. It is using ASK modulation which is very simple modulation

and demodulation. The transfer rate of the signal data is 4KB/S with transmitting power 10mW.

### Receiver module

The heart of the receiver module is Arduino Pro Mini, which is a microcontroller board based on the ATmega328. It runs at 5V and 16 MHz and has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, an on-board resonator, a reset button, and holes for mounting pin headers. A six pin header can be connected to an FTDI cable or Spark fun breakout board to provide USB power and communication to the board. This microcontroller controls the RF receiver and both outputs: the lamp and fan. The receiver module can work well for low performance, non-critical applications. The operating voltage is 5V DC with receiving frequency of 315MHz. The receiver sensitivity is 105dB with the help of external antenna 32cm single core wire, wound into a spiral. Figure-3 below shows the circuit diagram for the receiver module.

**Figure-3.** Circuit diagram for the receiver module.

There are four pins on the RF Receiver which are source pin, ground pin, and two pins of data. The source and ground pin is connected with the Vin and Gnd pin of Arduino Pro Mini. Pin A03 on Arduino Pro Mini is used to connect the data pin of RF Receiver. This connection is very important to make the communication between these two devices functioning well. Figure-3 shows the receiver module of the system.

### Virtual wire library

Virtual Wire is a library for Arduino that provides features to send short messages in one way transmission signal over wireless, using ASK (amplitude shift keying). It supports a number of inexpensive radio transmitters and receivers including 315MHz RF Transmitter Receiver. Therefore, it is very useful in the communication between the RF transmitter and receiver to produce the output of the system. Messages are sent with 4-to-6 bit encoding for good DC balance. ASK receiver require a burst of training pulses to synchronize the transmitter and receiver and also requires good balance between 0s and 1s in the message stream in order to maintain the DC balance of the message.





In this system, a motor and LEDs are used to represent the fan and lamp. In addition, an  $8\Omega$  speaker is also used to guide the user for the timing to speak the correct command. Other than that, LEDs circuit is developed to give the information to user about the current condition of the system. The motor used in this system is RF-300CA DC motor which only needs 0.5V to 4.0V to power it up.

## RESULTS AND ANALYSIS

Figure-4 shows the full system working which consist of transmitter part, receiver part and the output of the system.



**Figure-4.** Whole system running.

### Experiment 1: The effective working distance between transmitter and receiver

An experiment was conducted to test the distance between transmitter and receiver modules in order to ensure the effectiveness of the wireless system. It is done in a room with dimensions 5 x 12 meters. In this experiment, the distance between user and the microphone is fixed to 60 cm. The distance between transmitter and receiver is varied from 1 meter until 8 meters apart. User then spoke the command control words at the transmitter module. Each command word was repeated five times. At the receiver's end, the correctness of the output based on command given is observed and recorded.

Table- 3 shows the results of the experiment. Based on result in Table- 3, the wireless system works well for the first 6 meters. When the receiver was moved further than 6 meters, the receptions began to drop. Commands had to be repeated in order for the system to work well. So based on the results, it can be seen that the maximum working distance for the wireless system is 6 meters.

**Table-3.** Results of experiment 1.

| Distance between Tx and Rx (meters) | Output Observed at RX             |
|-------------------------------------|-----------------------------------|
| 1.0                                 | OK                                |
| 2.0                                 | OK                                |
| 3.0                                 | OK                                |
| 4.0                                 | OK                                |
| 5.0                                 | OK                                |
| 6.0                                 | OK                                |
| 7.0                                 | 3 correct output, 2 false output, |
| 8.0                                 | 1 correct output, 4 false output  |

### Experiment 2: Accuracy of the command

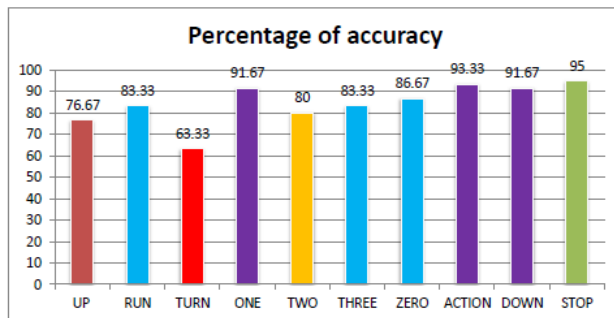
The purpose of this experiment is to test the accuracy of the voice command received, at the receiver's end. Accuracy of voice command is the most important outcomes in this system's speech recognition module. The aim is for the system to be able to recognize each command correctly and produce the correct output at the receiver's end. This experiment was conducted in a room with dimensions 4 x 5 meters. The distance between transmitter and receiver is fixed at 2 meters. The distance between user and the microphone is also fixed; at 60 cm. 20 persons were chosen as samples. The samples are aged between 20 to 23 years old and can speak English fairly well.

A set of 10 word commands are then given to each participant. Then each of them was asked to speak each command three times to the microphone at the transmitter module. In total, there are 60 samples for each command word spoken by all participants. The output at the receiver's end is observed. For each command, the number of incorrect output produced (the number of error) is recorded.

In Table- 4, the result of the experiment is shown. Figure-6 shows the percentage of accuracy for each command. The accuracy depends on the pronunciation spoken by the user. Based on the overall observation from the result above, there are four command words that have more than 90% of accuracy. They are "Action", "Down", "One", and "Stop". This means that all of them are easy to be pronounced by the user. The hardest word to recognize by the system is "Turn" which scored about 63% of accuracy. This system is using US English and the pronunciation of the word should have been in that accent. For the "Up" command, it is the second lowest of accuracy which is 76.67%. Some of the error occurs because the command is spoken with slow speed. So the system cannot detect the correct command. When the user improved the "Up" word pronunciation with the correct tone and loudness, the system detect the "Up" command and it works well.

**Table- 4.** Results of experiment 2.

| Word Command | No. of correct command | No. of errors |
|--------------|------------------------|---------------|
| UP           | 40                     | 14            |
| RUN          | 50                     | 10            |
| TURN         | 38                     | 22            |
| ONE          | 55                     | 5             |
| TWO          | 48                     | 12            |
| THREE        | 50                     | 10            |
| ZERO         | 52                     | 8             |
| ACTION       | 56                     | 4             |
| DOWN         | 55                     | 5             |
| STOP         | 57                     | 3             |



**Figure-5.** Percentage of command accuracy.

### Experiment 3: Effect of accent/dialect command to the system

Each individual sounds different from each other because of the different in our vocal tract shapes, larynx sizes, and other parts of our voice production organ. We also have different manner of speaking, with variety of accent, rhythm, intonation style, pronunciation pattern, and choice of vocabulary [12]. The aim of this experiment is to analyse the effect of accent/dialect command to the system. Accent is defined as words and the musicality of their speech. Dialect on the other hand is a person's accent and the grammatical features of the way that person talks. Both of them are alike and effect the pronunciation of their speech. In this experiment, only three commands were used, which are "Zero", "Three" and "Run". These words are chosen because it they are clearly different in pronunciation.

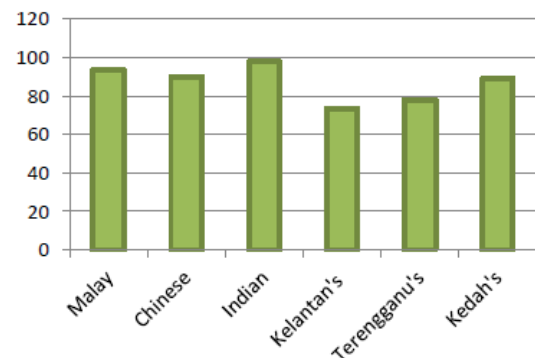
The distance between transmitter and receiver is fixed to 2 meters and the distance between user and the microphone is also fixed to 60 cm. For each accent, three samples were chosen. The same numbers of sample is also chosen for each dialect. Each command word was repeated three times by each speaker. Thus, each group of accent and dialect have a total 27 samples of command words. Then the output at the receiver's end was observed and recorded.

Table-5 shows the result of this experiment. Three different groups are listed in accent category. They are: Malay accent, Chinese accent and Indian accent. Then there are three different groups for dialect categories, which are Kelantan's Terengganu's and Kedah's dialect. The results show the highest number of error occurs is 9 out of 27 which is for sample of Kelantan's dialect. It was followed by Terengganu's and Chinese with 7 and 5 error occurs out of 27.

**Table-5** Experiment 3 result.

| ACCENT     | CORRECT OUTPUT | INCORRECT OUTPUT |
|------------|----------------|------------------|
| Malay      | 26             | 1                |
| Chinese    | 22             | 5                |
| Indian     | 27             | 0                |
| DIALECT    | CORRECT OUTPUT | INCORRECT OUTPUT |
| Kelantan   | 18             | 9                |
| Terengganu | 20             | 7                |
| Kedah      | 24             | 3                |

Figure-6 shows the percentage of correct command based on different accent and dialect. The percentage for Kelantan's is 73.33%, Terengganu's is 77.78% and Kedah's is 88.89%. But for Malay, Chinese and Indian, three of them are around 90% and above. From the results, it can be seen that different accent and dialects affects the way a person pronounce command word. As the result shows, wrong pronunciation did affect the performances of the system.



**Figure-6.** Percentage of correct output based on different dialect/accents.

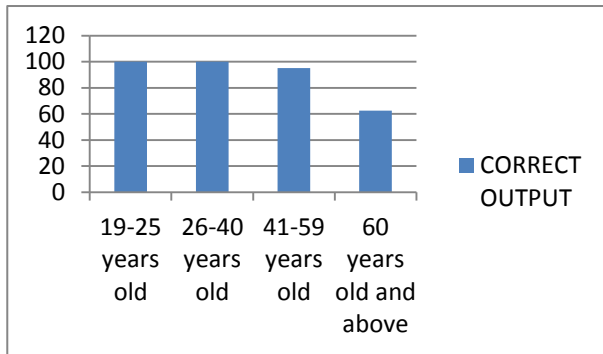
### Experiment 4: Effects of user's age to the system

Generally, speech performance is significantly worse in the case of elderly speaker compared to young adult speaker. One of the reason for this is because some parameters of the speech signal (e.g. speech rate, F0, jitter, shimmer) change with age [13]. Thus the aim of this experiment is to study about the effectiveness of the system for different age group. Four different groups of age were chosen, with ten speakers for each group. Each speaker was then asked to repeat the command word "One" four times. Therefore the total number of samples for each group of age is 40. The distance between transmitter and receiver is fixed to 2 meters and the distance between user and the microphone is also fixed to 60 cm.

Table- 6 and Figure- 8 show the results of this experiment. From the results, it can be seen that elderly user tend to produced more incorrect output compared to younger users. This could be because of the pronunciation of the command word is incorrect.

**Table-6.** Experiment 4 result.

| AGE GROUP       | CORRECT OUTPUT | INCORRECT OUTPUT |
|-----------------|----------------|------------------|
| 19-25 years old | 40             | 0                |
| 26-40 years old | 40             | 0                |
| 41-59 years old | 38             | 2                |
| 60 and above    | 25             | 15               |

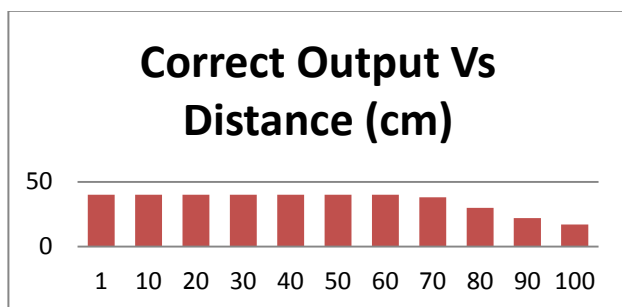


**Figure-7.** Percentage of correct output based on different age group.

#### Experiment 5: Speaker to microphone distance

The purpose of this experiment is to study the effect of the distance of the transmitter's microphone from the user. In this experiment, the distance between transmitter and receiver is fixed to 2 meters and the distance between user and the microphone is varied between 1 cm and 100cm, with 10cm increment value. Ten speakers were involved in this experiment, with each speaker spoke the command word four times. Thus the total sample at each distance marker is 40.

Figure-9 show the results from this experiment. The results show that the number of correct output deteriorates when the distance between the user and microphone is increased. It also shows that the ideal range of distance between user and microphone is in between 1cm to 60 cm.

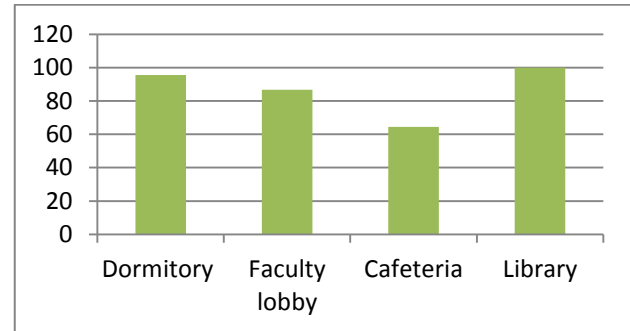


**Figure- 8.** Number of correct output for different speaker to microphone distance.

#### Experiment 6: Environment challenge

The speech recognition system is quite sensitive to the environment noise. This experiment is done to study the impact of environment challenges for this system. Therefore, the speech recognition system has been set up at different places around UMP Pekan Campus. It was tested by five speakers who were required to speak three command words which are "Down", "Stop" and "Action". Each speaker was then asked to repeat each command word three times. Therefore the total number of samples at each location is 45. The distance between transmitter and receiver is fixed to 2 meters and the distance between user and the microphone is also fixed to 60 cm.

The result of the experiment is shown in Figure-10. The result of this experiment shows that, the library have the maximum number of correct command which is 100 accuracy and the Cafeteria have the lowest number of correct command which is only around 60%. Therefore, the conclusion of this experiment is that the level of noise affected the system's performance. It is shown that quieter environment give the best performance of the systems.



**Figure-9.** Percentage of system based on different places.

#### CONCLUSIONS

Implementing speech recognition technique for home control system can make our life easier. This type of control systems can be applied to situations where it is not feasible to manually control home appliances. It can help disabled and elderly living at home.

The effectiveness of the system depends on several factors, which are the user pronunciation, the level of noise in the room where the system is set up and the distances between the transmitter and receiver modules. The system could be further improve by using better performance RF transmitter and receiver.

#### REFERENCES

- [1] D. H. Stefanov, Z. Bien and W. C. Bang 2004, "The smart house for older persons and persons with physical disabilities: Structure, technology arrangements, and perspectives," In: IEEE Trans. Neural Syst. Rehabil. Eng., Vol. 12, No. 2, pp. 228–250.
- [2] L. Ningqing, Y. Haiyang and G. Chunmeng 2013, "Design and implementation of a smart home control system," . In: Proc. - 3<sup>rd</sup> Int. Conf. Instrum. Meas. Comput. Commun. Control. IMCCC 2013, pp. 1535–1538.
- [3] G. Muthuselvi and B. Saravanan 2014. "real time speech recognition based building automation system," In: Vol. 9, No. 12, pp. 2831–2839.
- [4] A. A. Galadima. 2014. "Arduino as a learning tool," In: 2014 11<sup>th</sup> Int. Conf. Electron. Comput. Comput., pp. 1–4.



- [5] T. Kinnunen and H. Li 2010, "An overview of text-independent speaker recognition: From features to supervectors," In: *Speech Commun.*, Vol. 52, No. 1, pp. 12–40.
- [6] T. Pellegrini, V. Hedayati, I. Trancoso, a Hämäläinen, and M. S. Dias 2014, "Speaker age estimation for elderly speech recognition in European Portuguese," In: *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, No. September, pp. 2962–2966.
- [4] T. Obaid, H. Rashed, A. A. El Nour, M. Rehan, M. M. Saleh, and M. Tarique 2014, "Zig Bee Based Voice Controlled Wireless Smart Home System", In: *International Journal of Wireless & Mobile Network (IJWN)* Vol. 6, No.1.
- [5] O.Prabhakar, & N. Sahu 2013. A Survey On: Voice Command Recognition Technique. *International Journal of Advanced Research in*, Vol. 3, No. 5, pp. 576–585.
- [6] McLoughlin I. 2009. *Applied Speech and Audio Processing with MATLAB examples*", Cambridge University Press.
- [7] A. A. Galadima 2014. "Arduino as a learning tool," In: 2014 11<sup>th</sup> Int. Conf. Electron. Comput. Comput., pp. 1–4.
- [8] A. Kioumars and L. Tang 2011. "Atmega and xbee-based wireless sensing," in *Automation, Robotics and Applications (ICARA)*, In: 2011 5<sup>th</sup> International Conference on, pp. 351-356.
- [9] Arduino 2014. "Introduction: Arduino Uno Overview." Retrieved from [arduino.cc/en/Main/ArduinoBoardUno](http://arduino.cc/en/Main/ArduinoBoardUno).
- [10] Rani, M. U., Goutham, J., & Parthiban, M. 2014. "Improved Authentication Using Arduino Based Voice and Iris Recognition Technology", pp. 2319–2322.
- [11] An M., Yu Z., Guo J., Gao S. and Xian Y. 2014. The Teaching Experiment of Speech Recognition based on HMM, pp. 2416–2420.
- [12] T. Kinnunen and H. Li 2010. "An overview of text-independent speaker recognition: From features to supervectors," In: *Speech Commun.*, Vol. 52, No. 1, pp. 12–40.
- [13] T. Pellegrini, V. Hedayati, I. Trancoso, a Hämäläinen, and M. S. Dias 2014, "Speaker age estimation for elderly speech recognition in European Portuguese," In: *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, no. September, pp. 2962–2966.