Journal of Engineering Technology Vol. 1: 19-26, 2011
ISSN 2231-8798

Development Of Automatic Text System Using Speech Recognition

A.M. Zaid, M.M. Abdul Jamil & Y.L. Loo
Faculty of Electrical & Electronic Engineering
Universitiy Tun Hussein Onn Malaysia

Corresponding email: amranz@uthm.edu.my

Abstract: Nowadays, communication is an importance element in our daily life. The Short Message Service (SMS) and email which were implemented via typing are the most popular way in conveying messages or ideas. This existing technology requires a keyboard or keypad to compose the message and considered unfriendly to disabled people. With the advancement of technology, the typing process can be replaced via voice command using speech recognition. Most of the voice recognition system available in market today is the software programs that require a host computer and sound card to operate. These programs typically run in the background of a DOS or Windows environment, stealing themselves a portion of memory and Central Processing Unit processing power while allowing other programs like Excel or Word to run concurrently. The concurrent operation of the speech-recognition program slows the operation of any program that runs using voice recognition. In this study, speech recognition was created to replace the typing process to voice typing process which creates the external hardware to perform the tasks. The main part of this project is Analog to Digital Converter (ADC), PIC microcontroller, IC Databases, and Liquid Crystal Display (LCD). The ADC is use to convert the analog input signal from microphone to digital signal. IC databases are the library of text. The LCD display is to display the text message. When the user speaks to the microphone, the ADC will convert the sound signal to digital signal. Then the PIC microcontroller will compare the digital signal to IC Databases. The similar signal found will be displayed at the LCD display as a text message. The entire study present about a microcontroller with a voice control function which converts the alphabets tone to alphabets text.

Keywords: : Voice command, speech recognition, IC database, sound signal input, text message output.

1.0 INTRODUCTION

Typing message using keypad or keyboard will cause the user can't perform parallel tasks, and unfriendly to disable and blind peoples. These challenges can be solve by implementing automatic text message using speech recognition. Speech recognition is the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words [1].

Richardson et al. [2] have reported that the accuracy of speech recognition depends mainly on speaking rate, background noise, mismatch in microphone, language models and variations in speaker accent. Speech-recognition systems also have constraint concerning the style of speech they can recognize which are isolated, connected, and continuous [3].

Many classifiers have been proposed in the areas of speech recognition and speaker recognition including Gaussian Mixture Models (GMMs), Artificial Neural Networks (ANNs), Hidden Markov Models (HMMs) and Support Vector Machines (SVMs) [4].

Previous work on Speech Recognition has been concentrated on developing recognizers using Romanized characters. For example Speech Recognition will recognize the Arabic word as an English one, and then maps the Arabic word to its Romanized pronunciation [5]. H. Hyassat and R.A Zitar introduce the first SPHINX-IV-based Arabic recognizer and propose an automatic toolkit, which is capable of producing Pronunciation Dictionary [5]. Nora Barroso et al. [6] develop a language identification system oriented to robust Multilingual Speech Recognition in the Basque context where coexist three languages: Basque, Spanish and French. Mehryar Mohri et al. [7] survey the use of weighted finite-state transducers (WFSTs) in speech Recognition and show that WFSTs provide a common and natural representation for hidden Markov models (HMMs), context-dependency, pronunciation dictionaries, grammars, and alternative recognition outputs.

In this study, speech recognition is use to type the alphabets, therefore user only using voice to type message rather than using hand. The input for the system is enunciation to microphone, and the output will be the alphabets words in CPU. The speech recognition is interface with CPU via microcontroller with Human interface device (HID). The recognition system design in this project is dependant recognition system, therefore the user need to train the system before using. After the system been trained, when the user pronouns the alphabets enunciation, the speech recognition will compare the record in RAM. If the speech recognition found the match words in the RAM, it will display target words memory number in the speech recognition circuit. Concurrently, it also sends the signal to microcontroller via buffer circuit. Microcontroller will process the signal and send the packages signal to CPU for typing the words. The overall system diagram is shown in Fig. 1.

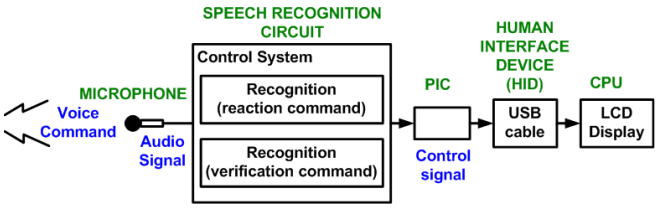

Fig. 1 Overall system diagram

2.0 METHODOLOGY

2.1 Hardware Development

The development of system hardware begins with the initial design of the hardware as shown in Fig. 2. Upon completion of the general idea on the design of the system, the entire related components need to be in order. The project begins with the construction of the speech recognition circuit using HM2007 integrated circuit. The speech recognition circuit design is speaker dependent and isolated recognition style.

The next step is to design the buffer circuit as the connecting medium between microcontroller and speech recognition. The Hex inverter is applied in buffer circuit. Hex inverter will invert the output signal from speech recognition to be a microcontroller input. The microcontroller will process the signal and send to CPU via USB cable.

The final step is the development process on the resisting circuit. Upon completion, the hardware would be decorated in order to simulate the real life situation.

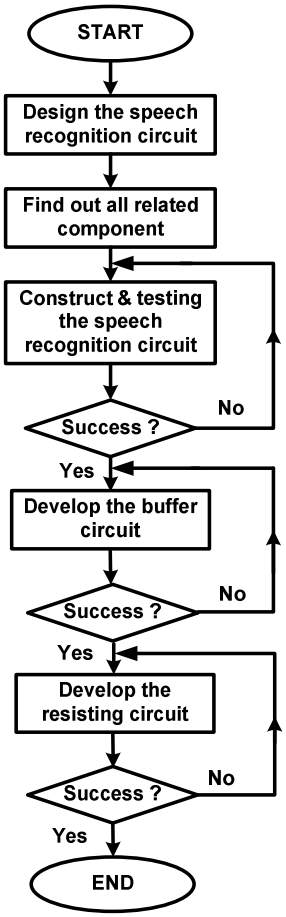

Fig. 2 Hardware Development Flow Chart

2.2 Software Development

In this process, the first step is generally to acquire all the relevant software to develop the algorithm for the system. The flow for the system is design and decided through flow chart in Fig. 3. Upon completion of the source code, the microcontroller will be programmed and tested for its functionality.

Programming language use to program the PIC is C programming and the compiler is MCC18. The program is design into two types, which are alphabets input and words input. The microcontroller use is PIC18F4550 which provide the USB connection function that use for HID.

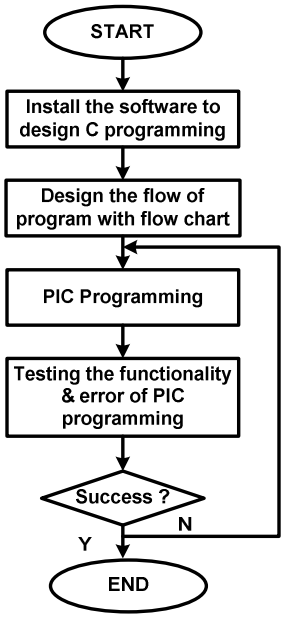

Fig. 3 Flow Chart of Software Development

2.3 Hardware and software integration

The step involves the combination of software and hardware of the system in which the source code of the system is tested. The microcontroller would be reprogrammed if the algorithm of the software fails to perform as desired. If succeeds it will proceed to the next process in which any idea of improvement on the system is discussed and implemented.

The initial task is only convert alphabets enunciation to alphabets text. After completed the initial task, the project is improved to convert simple words enunciation to words text. The simple words convert in this project is "good morning", "i am", "amran".

2.4 Program flow chart

This project consists of 2 types of program which is alphabets input and words input. Fig. 4 illustrate the program Flow Chart for alphabet input while Fig. 5 shows the program Flow Chart for words input.

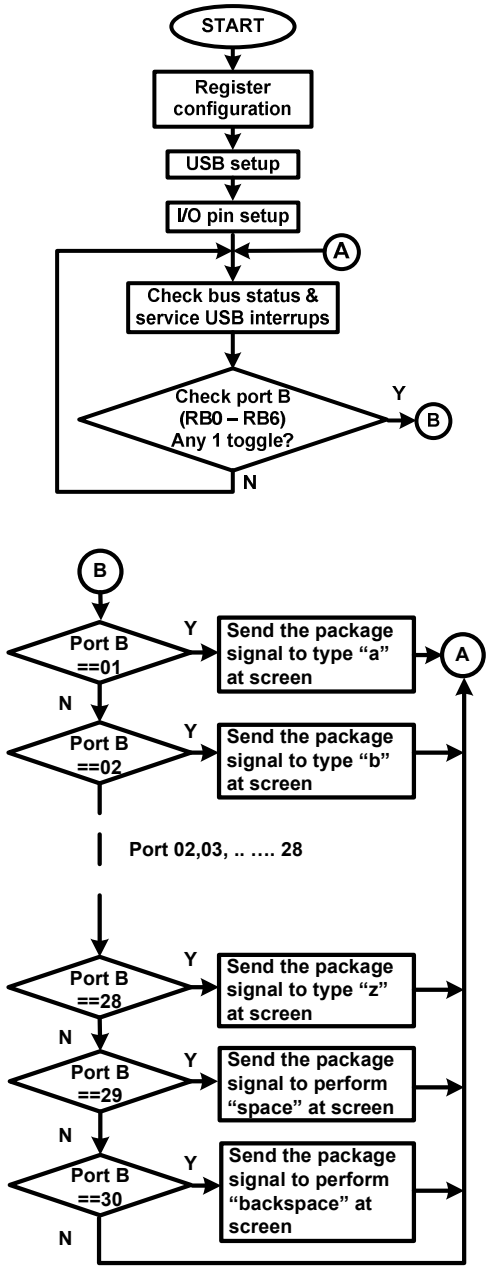

Fig. 4 Program Flow Chart for alphabet input

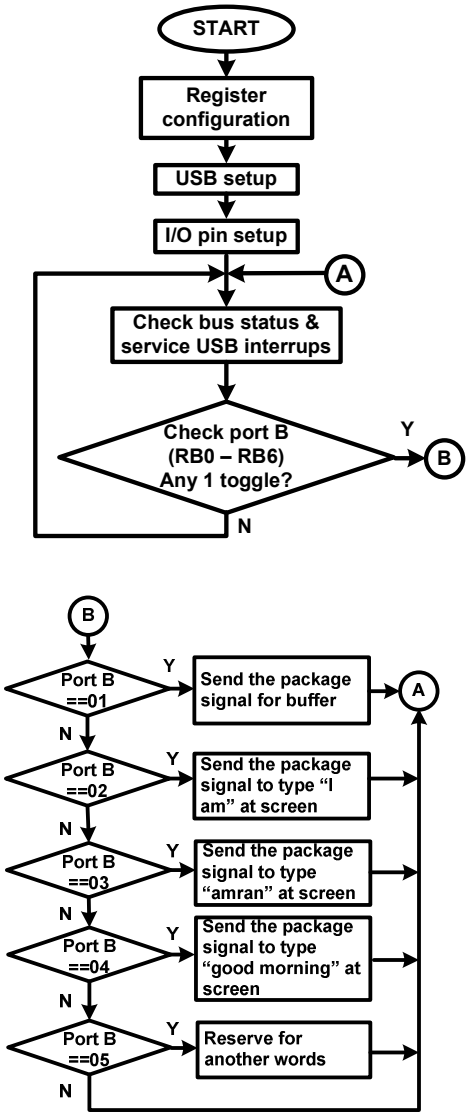

Fig. 5 Program Flow Chart for words input

3.0 RESULT AND ANALYSIS

3.1 Speech Recognition System

The speech recognition system developed in this study is dependent speech recognition system. It is developing using speech recognition IC HM2007 and SRAM. HM2007 is the specified speech recognition IC with build in ADC and comparator, but without internal memory, therefore HM2007 need the external memory to record the input signal. The external memory used is SRAM 8Kx8.

The output from speech recognition was connected to input of the buffer circuit. From Fig. 6, the output for speech recognition is 0.1V represent LOW state while Fig. 7 show the speech recognition output 3.6V represent HIGH state. The output from speech recognition is output from OCTAL D-TYPE FLIP-FLOP WITH 3-STATE OUTPUT (SN74LS373). By comparing the output voltage from speech recognition (in Fig. 6 and Fig. 7) to the data specification SN74LS373 from table 1 showing that the hardware speech recognition is function perfectly.

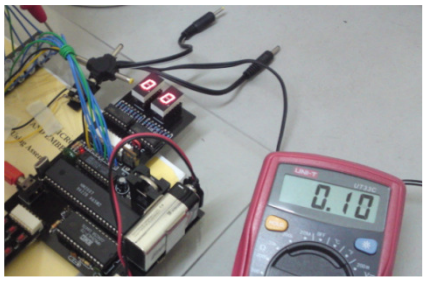

Fig. 6 Speech recognition initial condition

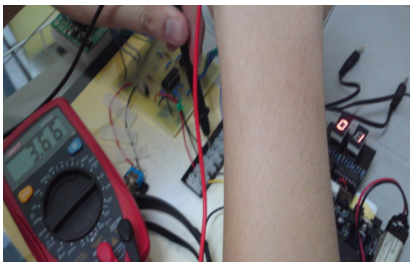

Fig. 7 Speech recognition input detect

Table 1: Data Specification SN74LS373

Parameter	Limit			Unit
	Min	Typ	Max	
Input HIGH voltage	2.0			V
Input LOW voltage			0.7	V
			0.8	V
Imput Clamp Diode Voltage		-0.65	-1.5	V
Output HIGH voltage	2.4	3.4		V
	2.4	3.1		V
Output LOW voltage		0.25	0.4	V
		0.35	0.5	V

The speech recognition system is speaker dependant, therefore the system need to be trained by the user before using. After trained the system, the system will operated on enunciation from the user. For example, if the enunciation "a" was trained as word number "1", saying the "a" into the microphone will cause the number 1 to be display.

The Homonyms enunciation will cause the systems error. For instance the enunciation "b", "d" and "x", "s" are

sound alike. Enunciation alike will confuse the speech recognition circuit. Therefore when choosing the target words for the system, need to avoid for using the words Homonyms enunciation.

3.2 Buffer circuit

The buffer circuit is implementing using 2 CMOS 4069 Hex Inverter .The buffer circuit is the connected medium between speech recognition and microcontroller. From Fig. 6 and Fig. 7 the output signal from speech recognition is 0.1V and 3.6V, but the microcontroller will only recognize "0" (low) with the voltage below 4.5V, and "1" (high) with the voltage over 4.5V. In order to interface with microcontroller, the signal output from speech recognition need to be differentiate as LOW with voltage less than 4.5V and HIGH with voltage over 4.5V.

To achieve this condition, the buffer circuit is created using CMOS 4069 Hex inverter IC. Then Hex Inverter IC will invert the signal 0V to 4.5V and 2.8V to 0V. Fig. 8 and Fig. 9 is operated together, hex inverter was invert the input signal LOW (0.12V≈0V) from Fig. 8 to HIGH (4.75V) from Fig. 9. In other sides, hex inverter also invert input signal HIGH (3.4V) to LOW (0V). In order to ensure the buffer circuit function successfully, comparing the data from real hardware to data specification is needed. Fig. 10 shows that Hex inverter will operate to invert Low input voltage from 0V – 2.5V to 5V, and invert High input voltage from 2.5V - 5V to 0V.

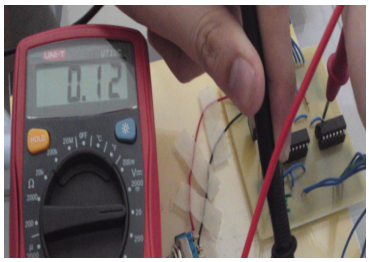

Fig. 8 LOW input buffer circuit

This data comparison is show that buffer circuit is function perfectly. By using the buffer circuit, microcontroller was able to differentiate between low and high signal.

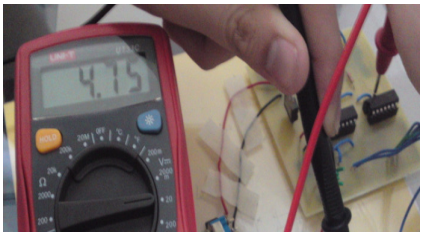

Fig. 9. HIGH output buffer circuit

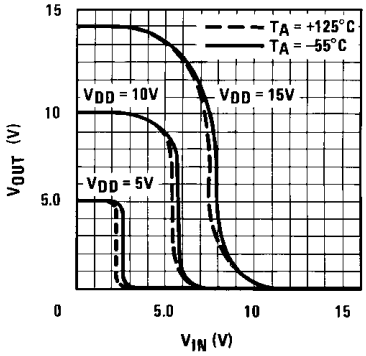

Fig. 10 Typical Performance Characteristics CD4069 [8]

3.3 Human Interface Device (HID)

The microcontroller is connecting with CPU via USB cable with HID method. This method is to create the keyboard driver from firmware via programming, because we need the CPU assume the device connected as the keyboard. Therefore the signal send to CPU via USB cable will be process as the keyboard signal. In order to typing words at CPU, the packages signal need to be sent from microcontroller to CPU is all refer to Table 2.

3.4 Overall System

The system having two option which are option 1 and option 2. Option 1 is dealing with alphabet input while option 2 dealing with word input.

The speech recognition using option 1 is 40 words memory (number 1 through 40) with the length of 0.96 seconds. Fig. 11 shows the initial condition when the system was activated. Output signal from speech recognition is all at 0V, and then it will invert by the buffer circuit to 4.5V and connect to microcontroller. From Fig. 12, it show that the CPU not typing any words.

Table 2: USB HID to PS/2 Scan code translation table

Key Name	HID Usage Page	HID Usage ID
System Power	01	81
System Sleep	01	82
System Wake	01	83
No Event	07	00
Overrun Error	07	01
Post Fail	07	02
Error Undefined	07	03
aA	07	04
bB	07	05
cC	07	06
dD	07	07
eE	07	08
fF	07	09
gG	07	0A
hH	07	0B
iI	07	0C
jJ	07	0D
kK	07	0E
lL	07	0F
mM	07	10
nN	07	11
oO	07	12
pP	07	13
qQ	07	14
rR	07	15
sS	07	16
tT	07	17
uU	07	18
vV	07	19
wW	07	1A
xX	07	1B
yY	07	1C
zZ	07	1D

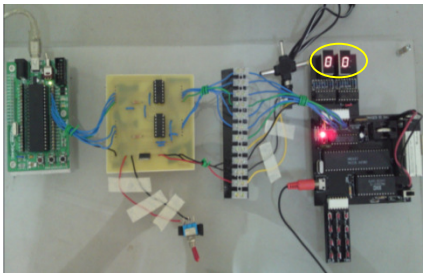

Fig. 11 Initial condition system

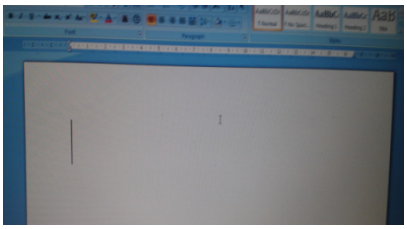

Fig. 12 Microsoft word display in CPU

After system be trained, when enunciation 'a' is detect from speech recognition, the "01" will be displayed as shown in Fig. 13. This will produce High (3.66V) at LSB speech recognition output as shown in Fig. 14.

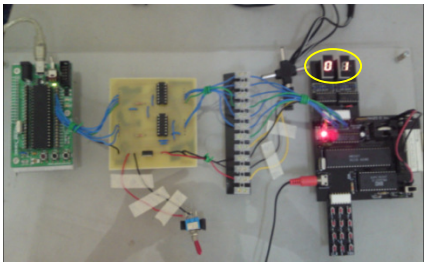

Fig. 13 'a' Input detected

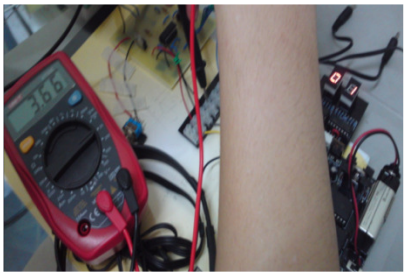

Fig. 14 Measurement on LSB pin

The buffer circuit will invert the 3.66V (HIGH) to 0.25V≈0V (LOW), and this was show in Fig. 15. This will activated the microcontroller LSB pin with the LOW (0.27V≈0.25V) signal as in Fig. 16.

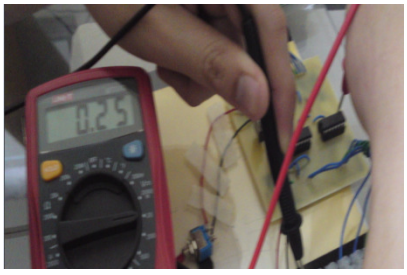

Fig. 15 Output pin buffer circuit

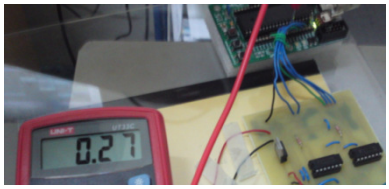

Fig. 16 Measurement on LSB input pin microcontroller

Finally the microcontroller will send the package signal to CPU for typing "a" in the Microsoft word as in Fig. 17.

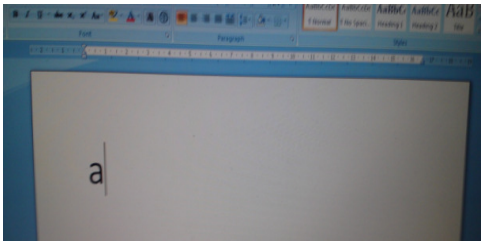

Fig. 17 'a' display in Microsoft word

The speech recognition using option 2 is a 20 words memory (numbers 1 through 20) with the length 1.92 seconds. The initial condition is the same as in Fig. 11. As usual the speech recognition also needs to be trained. In this case, "good morning" is trained in target word "04". When enunciation 'good morning' is detect from speech recognition, the "04" will be displayed as shown in Fig. 18 and then it will send signal to microcontroller via buffer circuit. After microcontroller receive the "04" signals, it will send the package signal to CPU for typing "good morning" via USB cable and this is shown in Fig. 19. All the package signal will be sent according to sequence 'g', 'o', 'buffer', 'o', 'd', 'space', 'm', 'o', 'r', 'n', 'i', 'n', 'g'. 'buffer' is represent buffer signal, because the CPU cannot accept the same signal continuously.

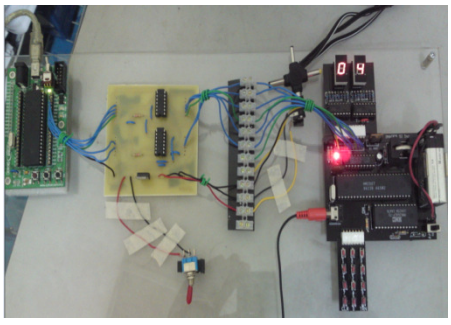

Fig. 18 "good morning" Input detected

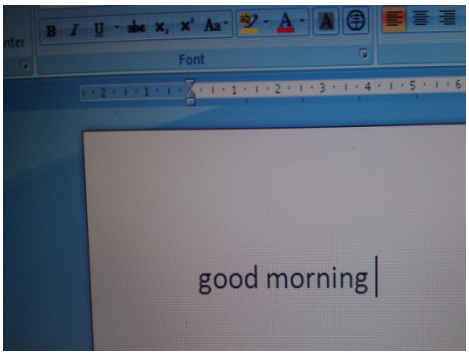

Fig. 19 "good morning" display in Microsoft word

The system is dependant speech recognition system, therefore the user need to train the circuit before using. Since the programming part already fix "01" is "a", "02" is "b" and so on. Therefore the users need to train the speech recognition circuit according to the data fix. The error will occur in the system when it detect Homonyms enunciation, to avoid this error the user need to trained another enunciation to replace it. For example, the user can trained "02" is "boy" and "04" is "dog", when user pronouns "boy" the system will type "b".

Besides, this system is also able to type words. The user only need to trained the input in one single target words memory, and set the package signal will need to be sent to CPU according to the sequenced via programming.

3.0 CONCLUSION

Speech recognition technology is to control some device by using voice. Speech recognition is classified into two categories, speaker dependent and speaker independent. The speech recognition build in this project is based on the speaker dependent and isolated recognition style. This means that the system needs to be train by user before use. The implementation of the project was done by using speech recognition IC HM2007. The purpose of this study is to improve the efficiency of work and be an assistance device for disabled people. The device allow the user perform parallel task, because the user just need to use mouth to type message rather than using eyes and both hand on keyboard and it also helping the disabled people to type in computer. Besides, this project also develops the usage of speech recognition. The speech recognition can be applied in many area such as robotic, electronic and communication to control input device. For example, the mobile robot and the robot arms also can be controlled by using speech recognition.

4.0 RECOMMENDATION

In this study, the memory slot for speech recognition is using RAM. RAM is the memory for temporary data and it will automatically erase the data inside when power off or reset. Therefore the user need to retrain the system every time when switch on the power. This can be improving by implementing ROM external memory to the speech recognition system for permanent data recording. The system design in this project is using too many power source, for example the speech recognition need one 3V and 9V battery, buffer circuit need adaptor to supply 9V, and microcontroller need 5V from USB cable. This can be improving by designing the external circuit for power supply to the system. The external circuit will take one single source for input and supply the voltage to speech recognition, buffer circuit and microcontroller.

The speech recognition in this system designed using HM2007 speech recognition IC. This IC provide high efficiency recognition but it error occur when sound alike enunciation, for example 'b' and 'd', 'x' and 's' are sound alike. Hidden Markov Model (HMM) method can be applied for design speech recognition. A Hidden Markov Model (HMM) is a statistical Markov model in which the system being modeled and is assumed to be a Markov process with unobserved (hidden) states. An HMM can be considered as the simplest dynamic Bayesian network.

REFERENCES

[1] "Automatic Speech Recognition and Understanding", Available: http://ewh.ieee.org/r10/bombay/news6/AutoSpeechRecog/ASR.htm

[2] M.Richardson, M.Y. Hwang, A. Acero, & X. Huang, "Improvements on speech recognition for fast talkers", In *Eurospeech conference*, September 1999.

[3] John Iovine, "Speech Controlled Mobile Robot", *Robot Androids and Animation,* 2nd ed., New York: McGraw Hill, 2002, pp143-165.

[4] I. Shahin, "Identifying speakers using their emotion cues", *International Journal of Speech Technology*, Vol. 14, pp 89–98, 2011

[5] H. Hyassat, R.A. Zitar, "Arabic Speech Recognition Using SPHINX engine" , *International Journal of Speech Technology*, Vol. 9, pp 133-150, 2006

[6] N. Barroso, K. L. de Ipiña, C. Hernández, A. Ezeiza, M. Graña, "Semantic speech recognition in the Basque context Part II: language identification for under-resourced languages", *International Journal of Speech Technology*, August 2011. Available: http://www.springerlink.com/content/x80551176g431387/fulltext.pdf

[7] M. Mohri, F. Pereira and M. Riley, "Weighted finite-state transducers in speech Recognition", *Journal of Computer Speech and Language,* vol 16, Issue 1, pp 69-88, 2002

[8] CD4069 Hex Inverter Datasheet, Futurlac. Available: http://www.futurlec.com/4000Series/CD4069.shtml