Controlling a Robot Using Small Database Speech Recognition

Tanasak Phanprasit
Department of Electrical and Electronics Engineering, Bangkok University, Rangsit,
Bangkok 12120, Thailand, E-Mail: tanasak.p@bu.ac.th

Abstract

Control the movement of the robot in order to obtain maximum efficiency must be considered response functions and the highest level of security as set designer. This paper presents results of testing industrial robots by using small database speech recognition. The small database was created by applying the Fast Fourier Transforms (FFT) to convert speech signal in time domain to speech signal in frequency domain. Then, the speech signal in the frequency domain (data size) was reduced by using the Principal Component Analysis (PCA). The commands were represented by Eigen value and the Eigen vector. The ten speakers consist of 5 males and 5 females. Each speaker was required to utter all seven commands. The results showed the accuracy of the speech recognition, in case the commands came from the ten speakers, was 67.00 percent. Otherwise, the accuracy was 61.29 percent.

Keywords: *Speech Recognition, Fast Fourier, Transform, Principal Component Analysis*

1. Introduction

Technology has become a part of our daily lives. Humans have the automatic control system to facilitate convenience. Computer systems and robots play a big role in daily lives of more people, especially in places where there are environmental hazards, such as a bomb place, finding the information on the military battlefield, the building fires, etc. The robotic motion control can be achieved in two groups. The first group is speech recognition technique such as [1]. The second group is the image processing technique such as researcher [2]. In this research, use the speech recognition technique. This paper presents the small

size of database that using speech recognition technique for controlling the robotic motion called robot speech recognition (RSR). Controlling the robotic motion consist of 2 parts, hardware system and software system. At the hardware system, the wireless communications system is used to control RSR. In a simulation, MATLAB is used. By applying the present method, the effective of this paper discusses the average accuracy and the adequacy of small size of database.

This paper is organized as follows: Section 2 describes theory and principles. Sections 3 and 4 are the structure of the RSR model and how to create the RSR. A Section 5 is the experimental and result and the conclusion is provided in section 6.

2. Theory and Principals

Conversion speech signal in time domain to speech signal in frequency domain requires the following steps. The first is calculating the magnitude value using Fast Fourier Transform (FFT) [3] then calculating the difference size value between of each speech signal with a mathematical (PCA) [4]. It is represented by the Eigen value and the Eigen vector. All the Eigen will be used as a database of each speech signal. In the next stage is to define the beginning and end of speech signal.

2.1 Define the begin and end of speech signal

An example of a speech signal is the word "Khaw" as shown in Figure 1. The highest peak value is the peak Pitch. And the length between the value of peak Pitch up to position the next peak Pitch called the Pitch period. To detect the peak period, the Rectangular Window is applied. The peak Pitch must be located in the center of the window. It must be on condition that, the width of the Rectangular Window must be greater than the Pitch period and less than twice the Pitch period. Nyquist's Theorem says that sampling rate must be twice of bandwidth (3.1 KHz). In this research, using a sample rate is 8,000 points. The speech signal will be adjusted by using cutting the starting point - the end point. The speech signal went through the process of adjusting to the norms and to make every speech signal with the same peak pitch.

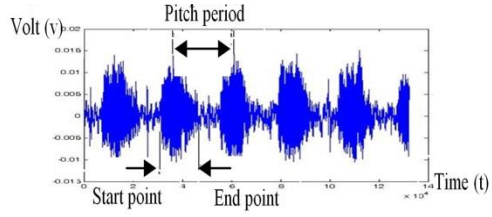

Figure 1. An example of a speech signal, the word “Khaw”

2.2 Block Diagram of RSR System

In the design and construct the RSR, the main part of RSR system consists of the following: 1) the initial processing of speech signals (Preprocess) 2) The Fast Fourier Transform 3) The Principal Component analysis, as shown in Figure 2.

Figure 2. Block diagram of the RSR system

2.3 Preprocessing Speech Signal

In the process of speech signal processing, basic steps, as shown in Figure 3.

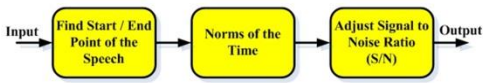

Figure 3. Block diagram of processing speech signal

Find start / end point of the speech signal. In this paper, we calculate the zero crossing ratio and energy value of the speech signal. This method can found the beginning / end point of speech signal is correct more than the other way around.

Normalize of the time. Because the length of speech signal (Sai, Khaw, Na, Lung, Kun, Long and Yoot) have a different length. Thus, the normalization of the each speech signal must be the same length. In this paper, using estimates in a linear function. It takes less time and memory.

Adjust Signal to Noise Ratio. This process will affect the performance of the speech signal. If the voltage of signal is adjusted to the higher level than level of noise voltage (S/N), the recognition efficiency will be higher. The proposed method uses first order digital filter as follows.

$$H(z) = 1 - az^{-1} \qquad (1)$$

a is the coefficient of the filter. In general, the values range from 0.92 to 0.97. In this research, we choose a = 0.94.

2.4 Fast Fourier Transform (FFT)

From the figure 2 is the block diagram of RSR system. The speech signal is converted into the frequencies domain using the Fast Fourier Transform. The Fast Fourier Transform process is used Discrete Fourier Transform, (DFT) as follows.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-jk(2\pi/N)n} \qquad (2)$$

where *x (n)* is a real number of speech input.

In calculating the FFT, only real number of the speech signal is used. The imaginary part is ignored. Speech signal using a number of points equal 8,000 points; they were rearranged as a two-dimensional data (2-D, 2x4000). The 2-D will be scaled down by half. The Principal Component Analysis is used. The result is equal 4,000 points, in the next stage is to analyze the speech signal with the PCA in the next section. The example of the frequency spectrum and speed signal "Yoot" is show in figure 4.

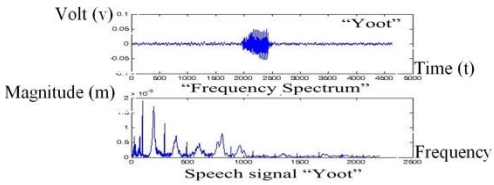

Figure 4. Speech signal and Frequency spectrum

2.5 Principal Component Analysis (PCA)

PCA is a statistical method, which is applied to various applications, such as data compression. The mathematical principles of PCA reduce the dimension of spatial data. Therefore the PCA is applied to reduce dimension of speech signal output of DFT. The results of this process are Eigen vector (v_i) and Eigen value (λ_i). The equation of PCA is as follow.

$$\psi_i = \frac{1}{M}\sum_{n=1}^{M}\Gamma_i ; (i = 1,2 : M = 1,2,...,4000) \quad (3)$$

ψ_i is the average value of speech signal.

M is the number word of the speech signal (5 words).

Γ_i is the group of speech signal (Group 4,000).

$$\phi_i = \Gamma_i - \psi_i \quad (4)$$

ϕ_i is the difference value between the group and the average value of speech signal.

ψ_i is the average value of speech signal.

$$A = \left[\phi_1 \quad \phi_2 \quad ,..., \quad \phi_{4,096}\right] \quad (5)$$

$$C = A^T A \quad (6)$$

C is the Covariance matrix.

A is the group of Covariance matrix.

From Covariance matrix, the Eigen vector and the Eigen value represent the Thai voice commands (Sai, Khaw, Na, Lung, Kun, Long, Yoot) in the data base**.** In the selecting Eigen vector, in this paper the second Eigen vector (*v2*) is chosen because it is the positive pair.

3. The structure of the model

The structure of the RSR is consisted of three main parts. First part is

input unit, second part is control unit and third part is output unit. A simplified block diagram depicts the overall of control system the RSR is shown in Figure 5.

Input unit is consisted of three parts. First part is the TX/RX Wireless module for speech signal and it was passed to the CPU2 that can be processed in arithmetic FFT and PCA to separated Thai voice commands types. Second part is the sensors for monitoring the rotation of wheels and calculating the distance for the RSR to move. Third part is the digital compass to find a path for the RSR.

Control unit is composed of three processors. The first central processing unit (CPU 1) was served to control the direction of the moving RSR. The second central processing unit (CPU 2) used for mathematical calculations. The last central processing unit (CPU 3) used for communication between Microphone and Wireless Router.

Output unit is consisted of three parts. First part is the LCD display for monitor the status of the RSR. Second part is the DC motor used to drive the arm. Third part is a DC motor used for moving the RSR follow to Thai voice commands. The block diagram of RSR structure is shown in Figure 6.

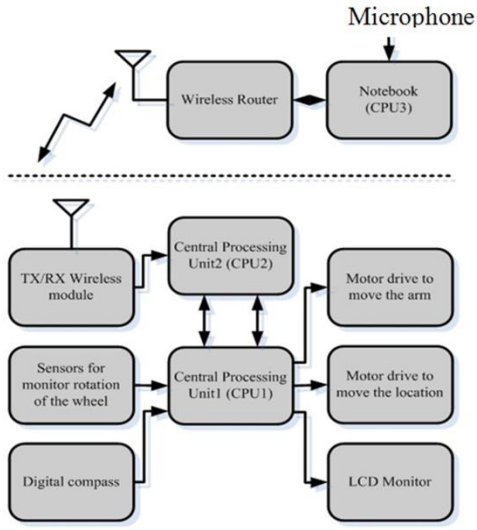

Figure 5. Block diagram of control system the RSR

4. How to create the RSR

The RSR will be divided in two parts.

First part is the structure of the RSR (mechanics structure). In this research, the diagram of RSR is structured as shown in Figure 6. In designing structure of an RSR, the designer has to consider elements of RSR, such as DC motor control, batteries, etc. to provide a balanced structure and movement. Structure of the RSR is consisted of the following materials.

- Side view: all four sides are made of an aluminum metal sheet. The front and rear sides are 420cm. (wide) and 50cm. (long). The other two sides are 550cm. (long) and 8.6cm (wide).

- Front view: the front part of the RSR is an aluminum metal sheet (170cm x 120cm) for the sensors to monitor rotation of the wheel.

- Top view: the top view is an aluminum sheet (420cm x 170cm) to accommodate board circuit.

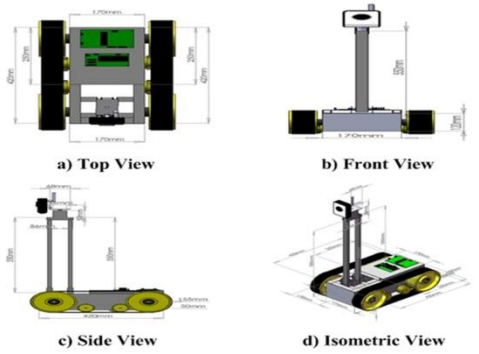

Figure 6. The block diagradiagram of RSR structure

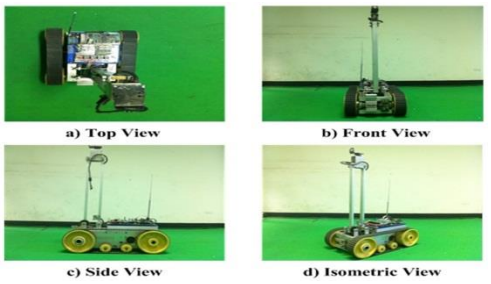

Figure 7. The structure of RSR prototype

- Sensors are installed at the bottom of RSR, for monitor the rotation of the wheels, to check motion distance of the RSR.

Second part is control circuit which consists of two circuits. The first control circuit is microcontroller (PIC18F8722). The second control circuit is Pulse width modulation (PWM). The rest of the control circuits are installed at inside area of the RSR. The structure of RSR prototype is shown in Figure 7.

5. Experiment and Result

To evaluate the effectiveness of the presented method Thai Language commands, we used seven different source both male and female for experimentation, such as A="Sai (means left)", B="Khwa (means right)", C="Na (means forward)", D="Lung (means backward)", K="Kun (means upward)", F="Long (means downward)", and G="Yoot (means stop). The symbol X is the Thai voice commands, the symbol Y (%) is the average value of the accuracy rate, the symbol Av is the average value of the accuracy rate at each time, as shown in table 1 to 4. Programming, C Language is used.

TABLE 1. SPEECH IN DATABASE (FEMALE)

X	Order of Speech										Y (%)
	1	2	3	4	5	6	7	8	9	10	
A	C	A	A	A	C	A	A	B	A	A	80.00
B	B	D	D	B	B	B	B	B	B	D	70.00
C	A	C	C	C	C	A	D	C	C	C	80.00
D	D	D	A	D	D	D	D	D	C	D	80.00
E	E	G	E	E	E	E	E	E	E	E	90.00
F	G	F	F	G	F	F	F	F	F	F	80.00
G	G	F	G	G	G	G	G	G	G	G	90.00
Av	57.14	57.14	71.43	85.71	85.71	85.71	85.71	85.71	85.71	85.71	70.00

Table 1, the speech in database (female) the results revealed that the average accuracy was 70% of all. The speech that the result highest average accuracy are the “E =Kun”, and “G = Yoot”. The speech is not in database the results in Table 2 that was 70% of all, that is “G = Yoot”.

TABLE 2. PEECH IS NOT IN DATABASE (FEMALE)

X	Order of Speech										Y (%)
	1	2	3	4	5	6	7	8	9	10	
A	A	C	A	A	C	C	A	A	A	A	70.00
B	B	B	D	B	B	D	B	B	D	B	70.00
C	C	A	C	C	A	D	C	D	D	C	70.00
D	A	A	D	D	A	D	D	A	D	D	60.00
E	E	A	D	E	E	E	D	E	E	E	70.00
F	F	D	F	F	G	G	F	F	F	F	70.00
G	G	G	G	E	G	G	F	G	G	G	80.00
Av	85.71	28.57	71.43	99.99	42.86	42.86	71.43	71.43	71.43	99.99	70.00

TABLE 3. SPEECH IN DATABASE (MALE)

X	Order of Speech										Y (%)
	1	2	3	4	5	6	7	8	9	10	
A	C	A	A	A	C	D	A	B	A	A	70.00
B	B	D	D	B	B	B	B	B	B	D	70.00
C	A	C	C	C	C	A	D	C	C	C	70.00
D	D	D	A	D	D	F	F	D	C	D	60.00
E	E	G	E	E	B	E	E	E	E	E	80.00
F	G	F	F	G	F	F	F	F	F	F	80.00
G	G	F	G	G	G	G	G	G	G	F	80.00
Av	57.14	57.14	71.43	85.71	71.43	57.14	71.43	85.71	85.71	71.43	72.86

TABLE 4. SPEECH IS NOT IN DATABASE (MALE)

X	Order of Speech										Y (%)
	1	2	3	4	5	6	7	8	9	10	
A	A	C	C	C	A	A	A	A	B	A	60.00
B	B	B	D	B	A	D	B	B	D	A	50.00
C	C	A	B	C	A	C	C	C	A	C	60.00
D	A	A	D	B	B	D	D	A	D	D	50.00
E	E	A	D	F	E	E	D	E	E	B	50.00
F	F	D	F	F	G	G	F	F	F	F	70.00
G	G	G	G	E	G	D	F	G	G	G	70.00
Av	85.71	28.57	42.86	42.86	42.86	57.14	71.43	85.71	57.14	72.43	58.57

Table 3, the speech in database (male) the results revealed that the average accuracy was 72.86% of all. The speech that the result highest average accuracy are the “E =Kun”, “F = Long” and “G = Yoot”. The speech is not in database the results in Table 4 that was 58.57% of all, that are “F = Long” and “G = Yoot”.

TABLE 5. SPEECH IS NOT IN DATABASE AND IN DATABASE (MALE AND FEMALE)

Speech (Sex)	Average accuracy value (%)	
In database	70.00	**Av** = 67.00%

(Female)		
In database (Male)	64.00	
Not In database (Female)	64.00	**Av** = 61.29%

The comparison between speech in database and speech is not in database (male and female) as show in Table 5 revealed that the average accuracy was 67.00% of in database (female, male), 61.29% of not in database (female, male).

6. Conclusion

This paper presents results of testing industrial robots by using small database speech recognition. The small database was created by applying the FFT to convert speech signal in time domain to speech signal in frequency domain. Then, the speech signal in the frequency domain was reduced by using the PCA. The commands were represented by Eigen value and the Eigen vector. The ten speakers consist of 5 males and 5 females. Each speaker was required to utter all seven commands. The results showed the accuracy of the speech recognition, in case the commands came from the ten speakers, was 67.00 percent. Otherwise, the accuracy was 61.29 percents.

References

[1] Hong Liu, Xiaofei Li, "A Selection Method of Speech Vocabulary for Human-Robot Speech Interaction," Systems Man and Cybernetics (SMC), IEEE Internation Conference on, pp. 2243-2248, 2010.

[2] Sylvain Calinon, Julien Epiney and Aude Billard, "A Humanoid-Robot Drawing Human Portraits," The 5th IEEE-RAS International Conference on Humanoid Tsukuba, Japan, Dec. 5-5, pp. 161-166, 2005.

[3] Raji Sukumar. A, Firoz Shah. A, Sarin sukumar. A, Babu. P., "Key-Word Based Query Recognition In a Speech By Using Artificial Neural Network," Second International Conference Computation Intelligence, Communication System and Networks, 2010.

[4] Jan-Yee Lee, Jeih-weih Hung, "Exploiting Principal Component Analysis in modulation spectrum enhancement for robust speech recognition," IEEE Trans on audio, 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), vol. no.3, pp. 1947-1951, 2011.