Studia i Materiały Informatyki Stosowanej, Tom 5, Nr 13, 2013
str. 17-24

FOXBOARD EMBEDDED SYSTEM AND RECOGNITION OF ISOLATED WORDS FOR CONTROL OF MECHATRONIC SYSTEM

Radoslav Bučko

Technical University of Košice
Department of Theoretical and Experimental Electrical Engineering
Letná 9, 042 00 Košice, Slovak Republic
e-mail: radoslav.bucko@tuke.sk

Abstract: *This paper describes the Fox Board embedded system, Teach-Robot mechatronic system and spectrograms, which have been chosen as a method of speech recognition. Three-windows functions were used in order to improve the recognition rate of commands. This paper demonstrates the results of improved isolated words recognition by Foxboard embedded system that can be used for control of Teach-Robot mechatronic system.*

Keywords*: Teach-Robot, FoxBoard, isolated words recognition*

1. INTRODUCTION

The recognition of spoken speech is highly developed in modern-day. Most research is focused on recognition of the English language for such communication, however research on our department is aimed at recognition of Slovak language. Research of this subject on our department is oriented on recognition of simple instructions by spectrogram. Spectrograms have been chosen for their simplicity, because FOX Board System offers only a limited computing power and RAM capacity.Modelowanie procesu negocjacji

2. FOX BOARD

As said above, FOX Board had been chosen for the embedded system (Fig.1). The size of FOX Board, of its entire printed circuit board (PCB) is 66x72 mm. FOX Board runs a real Linux operating system (not an uC Linux) on an ETRAX 100LX microprocessor, a 100MIPS RISC CPU. Word length is 32 bit, clock frequency is set to 100 Hz. Linux operating system is written in 8MB FLASH memory. Capacity of RAM is 32MB. The main advantages this embedded systems are: I/O ports (Fig.1) and networking protocol HTTP, FTP, Telnet and SSH. [1] [2]

Due to the importance of USB for the speech recognition system, speech recording had been done on personal computer (PC), with sending the recorded data from PC to FOX Board via WI-FI (wireless network).

In order to connect to a Wi-Fi LAN, the FOX Board needs to be equipped with an USB wireless network interface controller. Second USB key was used only for program and data storage.

Fig.1 **I/O of FOX Board**

3. TEACH-ROBOT MECHATRONIC SYSTEM

Teach-Robot system had been chosen for control by simple spoken instructions. Teach-Robot mechatronic system is an angular arm with 5 axles and 6 DC-motors (Table 1). Teach-Box is also provided with manual controller, as well as with communication protocols for communication between Teach-Robot and personal computer (Fig.2). [3]

Fig.2 Teach-Robot

TABLE 1 - MOVING SPECIFICATIONS

Function	Motor	Number of pulses	Angle
grip of jaws	M1	70	60°
rotation of jaws	M2	130	200°
wrist up/down	M3	420	90°
rotation of upper arm	M4	420	90°
rotation of lower arm	M5	350	80°
rotation of body	M6	700	320°

4. SOLUTION PROPOSAL

From the description of Teach-Robot mechatronic system and from the requirements of voice control, following Slovak voice commands were selected:

- Vľavo (Left), Vpravo (Right) for step motors M2 and M6 horizontal rotation
- Hore (Up), Dole (Down) for pulling and retracting the arm by using the step motors M3,M4,M5
- Otvor (Open), Zatvor (Close) for opening and closure of the robotic hand by step motor M1
- Stop (Stop) for stopping the motion of specified step motor
- Štart (Start) for beginning of voice control session
- Koniec (End) for setting the robot to its default position (Home – all step motors are in their default position) and for ending of the voice control session
- numerals JEDEN (One), DVA (Two), TRI (Three), ŠTYRI (Four), PÄŤ (Five) a ŠESŤ (Six) for selection of the step motors or for setting the step
- KROK (Step) and numeral – setting the step of the selected motor from 1 to 6

These 16 short commands are enough for basic voice control of the Teach-Robot mechatronic system.

5. SIGNAL PROCESSING

Speech signal is recorded mostly by microphone, which means, that the analog signal is recorded. Analogue cycles are digitalized, so the continuous signal is represented by the sequence of numbers. This process is called pulse code modulation.[5]

Pulse code modulation consists of two operations:

- sampling in time,
- quantization.

Sampling in time – samples, which are taken from the continuous signal, are in periodic moments $t_n=n.T$ which sizes correspond to the immediate values of continuous signal in sampling time t_n. T is the sampling period and $n=0,1,...,\infty$. [3]

5.1. Processing by time

Most methods of short term analysis in time can be described by the following equation:

$$Q_n = \sum_{k=-\infty}^{\infty} \tau(s(k))\, w(n-k), \qquad (2)$$

where Q_n is the short time characteristics, $s(k)$ is the sample of acoustic signal got by pulse code modulation in time k, $\tau(s(k))$ is the transformation function and $w(n)$ is the weight sequence (or window) which chose the samples $s(k)$.

Hamming´s windows. Hamming´s windows are used when processing in time is realised. Hamming´s window is defined as:

- $w(n)=0,54-0,46\cos[2\pi n/(N-1)]$ for $0 \le n \le N-1$,
- **$w(n)=0$ for other n.**

Hann´s windows. Hann´s window is defined as:

- $w(n)=0,5\{1-\cos[2\pi n/(N-1)]\}$ for $0 \le n \le N-1$,
- $w(n)=0$ for other n.

Rectangular window. Rectangular window (sometimes known as Dirichlet´s window) is defined as:

- $w(n)=1$ for $0 \le n \le N-1$,

$w(n)=0$ for other n.

6. SPECTROGRAM

As said above, spectrograms had been chosen as a method of speech recognition by embedded systems. A spectrogram is a time-varying spectral representation (forming an image), that shows how the spectral density of a signal varies in time.
The vertical axis shows the positive time towards the top, the horizontal axis represents frequencies, and the colors represent the most important acoustic peaks for a given time frame, with red representing the highest energies. [6]
A spectrogram is calculated from the time signal using the short-time Fourier transform (STFT). In our research three windows for STFT – Hann, Rectangular and Hamming window had been chosen. The spectrogram had been divided to several sectors, in order to ease the computing process. The final value was made by arithmetic mean of these sectors.
Figures (Fig.5, Fig.6) are showing the same word (word “stop”) spoken by two people. Similarities can be observed there. Fig.7 is showing different word (word "vpravo"). As can be seen, the spectrographic picture is clearly different, which is very important for speech recognition system.
.

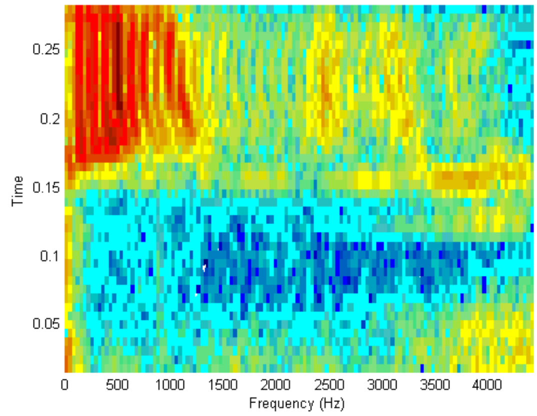

Fig.5 Spectrogram of word “stop” spoken by first person

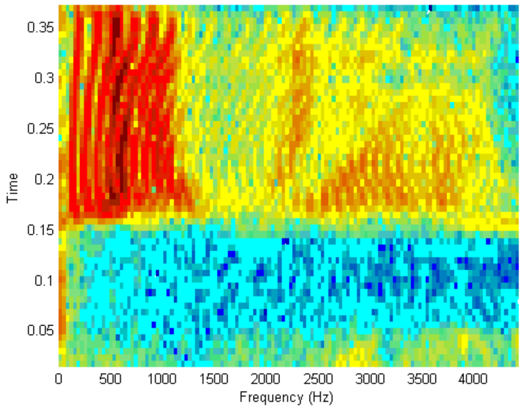

Fig.6 Spectrogram of word “stop” spoken by second person

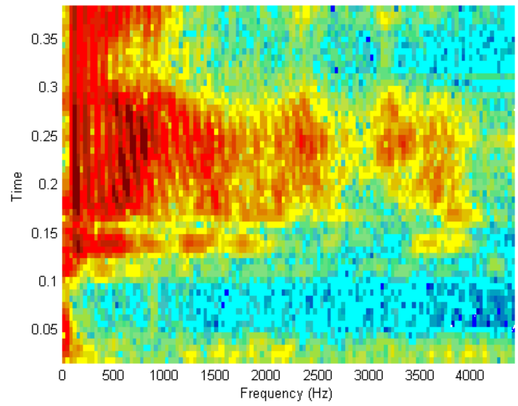

Fig.7 Spectrogram of word “vpravo” spoken by second person

The spectrogram had been divided to several sectors, in order to ease the computing process. The final value was made by arithmetic mean of these sectors. (Fig. 8).

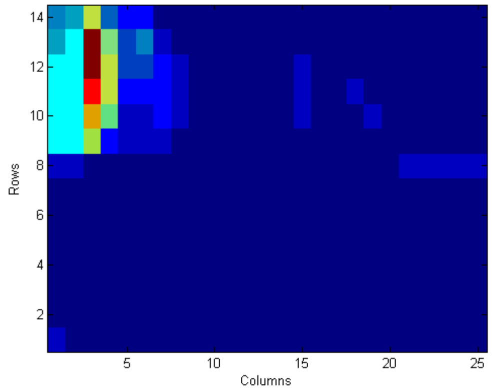

Fig.8 Spectrogram of word "stop" with arithmetic mean of sectors spoken by first person

7. SPECTROGRAM NORMALIZATION

Figures (Fig.8, Fig.9) show the different word (word "stop" and word "vpravo") spoken by one person. It can be seen, that number of rows is unequal-14 rows in spectrogram of word "stop" and 19 rows in spectrogram of word "vpravo" . In order to calculate the Euclidean distance of these two spectrograms (word "stop" and word "vpravo"), one of them needs to be normalized to the same number of rows. It is possible to normalize the higher amount of rows (19), to lower amount (14) using the crop or the missing rows can be filled to the higher amount of rows, when. The second method had been chosen, where the rows 15-19 had been filled with blank rows (fig.10).

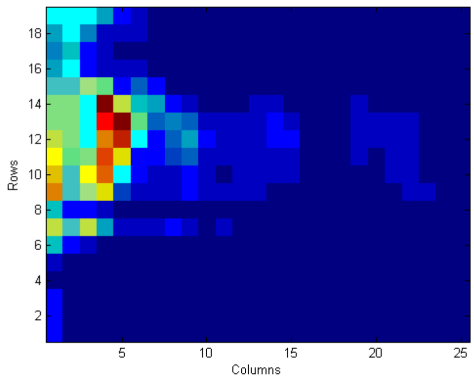

Fig.9 Spectrogram of word "vpravo" with arithmetic mean of sectors spoken by first person

Euclidean distance of spectrogram of words „vpravo" and „stop" is 10,7857. (Fig.10) This distance is higher than the Euclidean distance of the spectrogram of two same words „stop" = 4,27785.

In order to obtain an improved recognition of commands, 2 things need to be done:

1) Calculate the Euclidean distance from the spectrograms, which are calculated by STFT with 3 windows – Hann´s, Hamming´s and Rectangular.

2) Second arithmetic mean calculation.

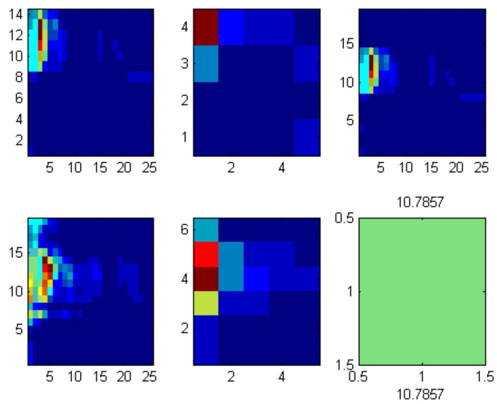

Fig.10 Comparison of spectrogram of word "stop" - up and word "vpravo" - down with arithmetic mean, second arithmetic mean, normalization of spectrogram and Euclidean distance

8. RESULTS - SUCCESS RATE OF SPEECH RECOGNITION

The following results were obtained when the method of spectrogram comparison for speech recognition was tested. The 20 spoken commands had been tested.

The table (Tab. 2) shows the success rate of command recognition by speakers whose spoken commands had been recorded into dictionary before, with speakers who didn´t spoke the commands before, with the utilization of first arithmetic average. The total average success rate for speakers who previously recorded the commands into dictionary had been 69,6875%. On the other hand, the success rate for speakers who didn´t record their commands into dictionary had been 66,5625%, so the success rate is lower by about 3,125%.

Figure (Fig.12) shows the success rate of isolated words recognition by using the first arithmetic average.

Table (Tab. 3) shows the success rate of spoken commands recognition by speakers who recorded their commands into dictionary when using the second arithmetic average. The total average success rate without using the 3 window functions had been 72,5% - about 2,8125% more than success rate when first arithmetic average was used.

TABLE 2 SUCCESS RATE OF ISOLATED WORDS RECOGNITION, WHEN FIRST ARITHMETIC AVERAGE WAS UTILIZED

Word	Success rate of word recognition by speakers who recorded the commands into the dictionary [%]	Wrong recognition [%]	Success rate of word recognition by speakers who didn´t record the commands into the dictionary [%]	Wrong recognition [%]
dole	**75**	**25**	**65**	**35**
hore	**70**	**30**	**70**	**30**
štart	**65**	**35**	**60**	**40**
stop	**75**	**25**	**70**	**30**
vľavo	**70**	**30**	**70**	**30**
vpravo	**75**	**25**	**75**	**25**
koniec	**70**	**30**	**65**	**35**
jeden	**70**	**30**	**70**	**30**
dva	**75**	**25**	**70**	**30**
tri	**75**	**25**	**75**	**25**
štyri	**55**	**45**	**55**	**45**
päť	**65**	**35**	**60**	**40**
šesť	**65**	**35**	**60**	**40**
krok	**70**	**30**	**65**	**35**
otvor	**70**	**30**	**75**	**25**
zatvor	**70**	**30**	**60**	**40**

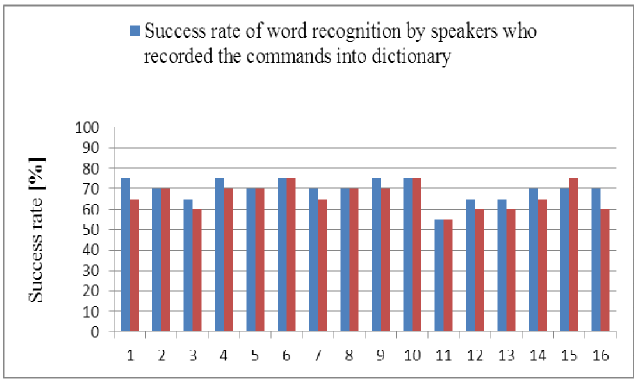

Fig. 12. Success rate of isolated words recognition by using the first arithmetic average.

TABLE 3 SUCCESS RATE OF ISOLATED WORDS RECOGNITION, WHEN SPOKEN BY SPEAKERS WHO RECORDED THE VOICE COMMAND PATTERNS.

Word	Successful recognition [%]	Wrong recognition [%]	Successful recognition with 3 windows [%]	Wrong recognition with 3 windows [%]
dole	**75**	**25**	**80**	**20**
hore	**75**	**25**	**80**	**20**
štart	**65**	**35**	**75**	**25**
stop	**80**	**20**	**85**	**15**
vľavo	**75**	**25**	**85**	**15**
vpravo	**80**	**20**	**80**	**20**
koniec	**70**	**30**	**75**	**25**
jeden	**75**	**25**	**80**	**20**
dva	**75**	**25**	**75**	**25**
tri	**80**	**20**	**85**	**15**
štyri	**60**	**40**	**65**	**35**
päť	**70**	**30**	**75**	**25**
šesť	**65**	**35**	**70**	**30**
krok	**70**	**30**	**75**	**25**
otvor	**75**	**25**	**75**	**25**
zatvor	**70**	**30**	**75**	**25**

When the 3 window functions were used, (which were added to increase the success rate) the average value of the recognized commands increased to 77,1875% - about 4,6875% more.

Figure (Fig. 13) shows the success rate of isolated words recognition, when spoken by speakers who recorded the voice command patterns.

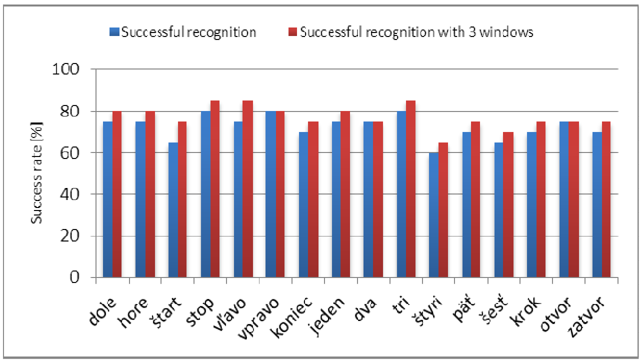

Fig. 13. Success rate of isolated words recognition when spoken by speakers who recorded the voice command patterns

The table (Tab. 4) shows the success rate of voice commands recognition, when spoken by speakers who didn´t record the patterns into the dictionary, when second arithmetic average was used. The total success rate without using the 3 window functions is 67,8125% - about 1,25% more than the success rate when first arithmetic average was used. With utilization of the 3 window functions the success rate of commands recognition had increased to 72,1875% - about 4,375% more.

TABLE 4 SUCCESS RATE OF VOICE COMMANDS RECOGNITION, WHEN SPOKEN BY SPEAKERS WHO DIDN´T RECORD THE PATTERNS

Word	Successful recognition [%]	Wrong recognition [%]	Successful recognition with 3 windows [%]	Wrong recognition with 3 windows [%]
dole	**65**	**35**	**70**	**30**
hore	**75**	**25**	**75**	**25**
štart	**60**	**40**	**65**	**35**
stop	**70**	**30**	**75**	**25**
vľavo	**75**	**25**	**80**	**20**
vpravo	**75**	**25**	**80**	**20**
koniec	**70**	**30**	**75**	**25**
jeden	**70**	**30**	**75**	**25**
dva	**75**	**25**	**80**	**20**
tri	**75**	**25**	**80**	**20**
štyri	**55**	**45**	**60**	**40**
päť	**60**	**40**	**65**	**35**
šesť	**60**	**40**	**65**	**35**
krok	**65**	**35**	**70**	**30**
otvor	**75**	**25**	**75**	**25**
zatvor	**60**	**40**	**65**	**35**

As can be seen from the results, the addition of 3 window functions had higher impact on speech recognition in cases, when the speaker recorded their command patterns into dictionary before. Figure (Fig.14) shows the success rate of isolated words recognition, when spoken by speakers who didn´t record the patterns.

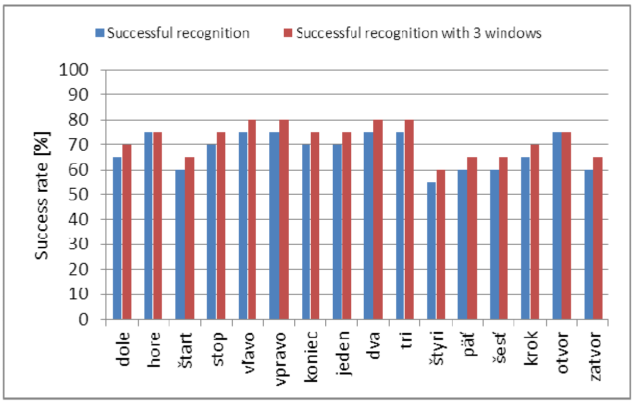

Fig. 14. Success rate of isolated words recognition, when spoken by speakers who didn´t record the patterns.

The average success rate of words recognition spoken by speakers who didn´t record their voice command pattern into the dictionary, together with the use of 3 window functions is about 5% lower than the success rate of the speakers with recorded voice command pattern. In these results, where the success rate is about 77%, when using the 3 window functions could be further improved by utilizing the classifier by utilizing the dynamic programming.

9. CONCLUSION.

With the limited computing power of embedded systems like the FoxBoard (we don´t consider the mobile phones) it is essential to lower the computing load. The success rate of isolated words recognition when using 3 windows is about 77%. Such success rate is not enough for professional use, therefore in the future, the dynamic programming will be tested in order to improve it.

10. ACKNOWLEDGEMENT.

The paper has been prepared under support of Slovak grant project KEGA No. 005TUKE-4/2012.

Nadácia Volkswagen Slovakia

Volkswagen Slovakia Foundation, grant no. 041/13_RT – Fast charger for electromobiles based upon the principle of directed energy transfer by air.

Literatura

1. Manual Teach-Robot with Studio version 01-06-2005.
2. FOXBoard, “documentation” ,http://foxlx.acmesystems.it/?id=4, 10.1.2012
3. Vince, T.: Teach-robot control system via Internet, OWD 2010 : 12 International PhD Workshop : 23–26 October 2010. - Gliwice : Silesian University of Technology, 2010 , ISBN 83-922242-7-2
4. T. Vince, J. Molnár, R. Bučko: Real-time regulation systems based on internet - optimization algorithm. In: Transactions of KMOSU. Vol. 4, no. 3 (2010), p. 150-153. - ISSN 2072-8263
5. Kováčová, I., Kováč, D. Industrial Electrical Engineering – Modeling of Converters. // In: 1st part, book, Published by TU of Košice, 2011, 66 pages, ISBN 978-80-553-0617-9
6. Hagiwara R., How to read a spectrogram, http://home.cc.umanitoba.ca/~robh/howto.html, 7.4.2011
7. Molnár J., Bučko, R., Bačko, M.: Automatic measurement system in automobile based on Internet. // In: Transactions of KMOSU. Vol. 4, no. 2 (2010), p. 166-169. - ISSN 2072-8263
8. Bučko, R., Success rate of isolated words recognition by embedded system, In: Proceeding of scientific and student's works in the field of Industrial Electrical Engineering, SSIEE - 2013, Vol. 2, Part 1, Technická univerzita v Košiciach, 2013, pp. 97-100, ISBN 978-80-553-1425-9