SPEECH RECOGNITION USING SOM AND ACTUATION VIA NETWORK IN MATLAB

SUBASH CHANDAR. A[1], SURIYANARAYANAN. S[2] & MANIKANDAN. M[3]

[1]Student, Faculty of Engineering, National University of Singapore, Kent Ridge, Singapore
[2]Student, Department of Electrical Engineering, Indian Institute of Technology, Chennai, India
[3]Employee, Converteam EDC Private Ltd., Chennai, India
E-mail: subash_394pec@yahoo.co.in, ussuriya@gmail.com, m.manibala@gmail.com

Abstract- This paper proposes a method of Speech recognition using Self Organizing Maps (SOM) and actuation through network in Matlab. The different words spoken by the user at client end are captured and filtered using Least Mean Square (LMS) algorithm to remove the acoustic noise. FFT is taken for the filtered voice signal. The voice spectrum is recognized using trained SOM and appropriate label is sent to server PC. The client and the server communication are established using User Datagram Protocol (UDP). Microcontroller (AT89S52) is used to control the speed of the actuator depending upon the input it receives from the client. Real-time working of the prototype system has been verified with successful speech recognition, transmission, reception and actuation via network.

Keywords- *Speech recognition, Self organizing map, Fast Fourier transforms, Actuation, Least mean square, User Datagram Protocol.*

I. INTRODUCTION

The area of speech processing is developing and shows tremendous potentialities for widespread use in the future. A speech recognition system makes human interaction with computers possible through a voice/speech to initiate an automated service or process [1]. Controlling a machine by simply talking to it gives the advantage of hands-free, eyes-free interaction. Several literatures have been published for Speech recognition using neural networks [3]-[6]. Constructing an effective Speech recognition system requires an in-depth understanding of both the tasks to be performed, as well as the target audience who will use the final system. Actuation based on network offers unique advantage over traditional local control. Combining speech recognition with network actuation can be used to control the actuator from a remote place.

The system (Speech Recognition and Actuation via Network) is divided into two modules. The first module is the client module with speech recognition system which provides the interaction between the user and the PC. The words spoken by the user are captured by the client computer. The server module with serial communication is another one which receives data from the client computer using UDP. The server module provides the PC communication with the actuator through a micro controller. The client module is developed in M-script which in turn calls a simulink file responsible for sending data to client. The server module is a simulink file which receives data and sends it to serial port.

UDP is used for the network communication between host and remote computers. The server performs two - functions such as data reception and data delivery. Data is received by the server from the client, which delivers a label value for each word captured. The data is used by the server to actuate any device using a microcontroller. The micro controller produces the necessary control signal with respect to the data received and it is sent to the actuator. The data from the server is transmitted via a serial port to the micro controller.

II. SELF ORGANISING MAP

A self-organizing map (SOM) is a type of artificial neural network which is trained using unsupervised learning. Self-organizing maps are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space. The principle goal of the self-organizing map is to transform an incoming signal pattern of arbitrary dimension into a one- or two dimensional discrete feature maps, and to perform this transformation adaptively in a topologically ordered fashion.

During training phase, the input vector from training set is fed to the network, its Euclidean distance to all weight vectors is computed. The neuron with weight vector most similar to the input is called the winning neuron **w**. The weights of the winning neuron and neurons close to it in the SOM lattice are adjusted towards the input vector. The size of the topological neighborhood function shrinks with time because of time varying width, hence the distance from the winning neuron. The update formula for a neuron with weight vector is given by (1), (2), (3) and (4).

$$w_j(n+1) = w_j(n) + \eta(n) h_{j,i(x)}(n) \left(x - w_j(n)\right) \quad (1)$$

$$h_{j,i(x)}(n) = e - \left(\frac{d^2{}_{j,i}}{2\sigma(n)^2}\right) \quad (2)$$

$$\sigma(n) = \sigma_0 e - \left(\frac{n}{\tau 1}\right) \quad (3)$$

Where **h** is the topological neighborhood function, **d** is the Euclidian distance from the neuron *j* to the winning neuron *i*, $\boldsymbol{\sigma_0}$ and **σ** is the initial and time varying effective width of the topological neighborhood and **η** is the learning weight, **n** is number of iteration count value. $\boldsymbol{\tau_1}$ is the time-constant to control the decay rate of the learning rate which is given by,

$$\tau 1 = \frac{T}{log\ (\sigma_0)} \quad (4)$$

The new weight is the average of the input and the current weight. The synaptic weight vector $\boldsymbol{w_j}$ of winning neuron *i* move toward the input vector ***x***. All the neurons in the neighborhood of the winning neuron also move toward the input vector. The farther away neurons have less change. Upon repeated presentations of the training data, the synaptic weight vector tends to follow the distribution of the input vectors due to the neighborhood updating. Thus leading to a topological ordering of the feature map in the sense that neurons that are adjacent in the lattice as shown in Fig.1, will tend to have similar synaptic weight vectors.

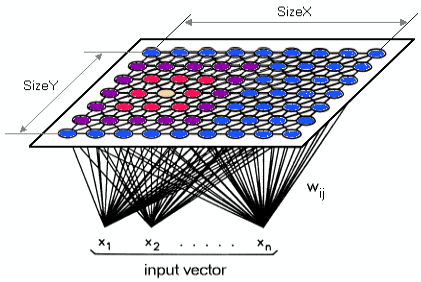


Fig.1. Lattice structure of SOM

The architecture is very simple and it comprises the following systems working in a combined manner, Speech Recognition system, Client PC, Network, Server PC, Serial port Interface and an Actuator (Motor).

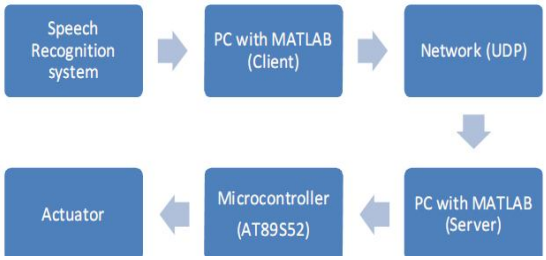


Fig.2. System architecture of Network based Actuation

A. Speech Recognition System

The Speech Recognition system consists of speech acquisition module and recognition module. The speech acquisition module records the user voice for two seconds preceded by a beep sound which indicates the acquisition of data and saves it as a row vector as shown in Fig.3. The raw voice data is filtered using LMS algorithm to remove the acoustic noise Fig.4. FFT is taken for filtered voice signal to reduce the vector length hence reducing the execution time. The speech recognition module is a trained SOM, which gives a label to the acquired data as per the trained value. The training data for SOM has four words from the user which are, 'START','STOP','SPEED1','SPEED2' and 'BREAK'. The training dataset is formed by taking FFT [2] of the input voice as shown in Fig.5, to avoid the difference due to delay of speech. Another advantage of taking FFT is, the spectrum remains same for region of interest, irrespective of cross interference and noise. The dataset consists of voice signals of different individuals both male and female. The trained network is then used to identify four words which are specified above. The cluster formation for different words spoke by different individuals by SOM is shown in Fig.6.

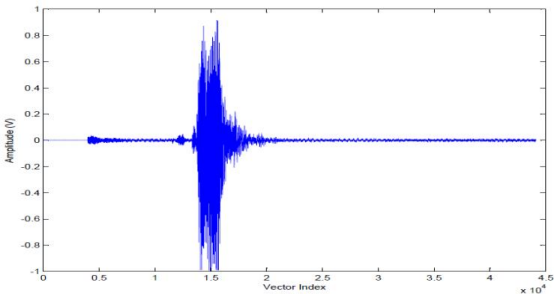


Fig.3. Raw voice signal

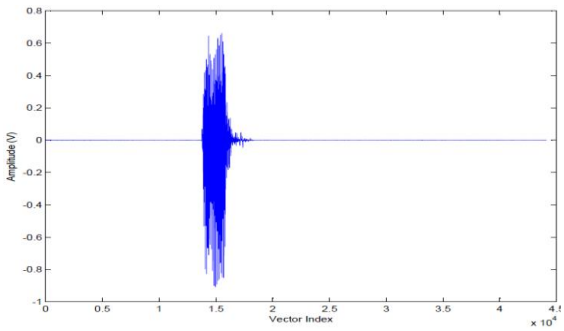


Fig.4. LMS filtered voice signal

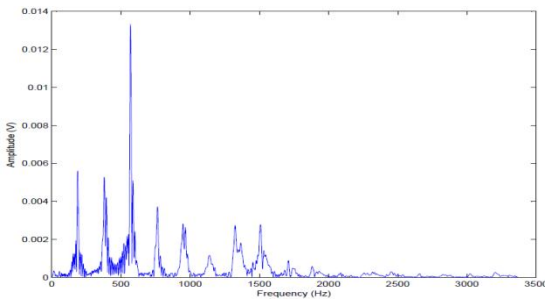


Fig.5. FFT of filtered of voice signal

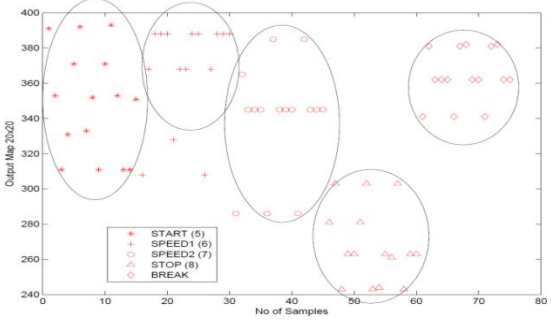


Fig.6. SOM mapping for different words

III. SYSTEM ARCHITECTURE

B. Client (UDP send)

The client module is designed in simulink. It is developed to receive the data from Speech Recognition module and to move the received data to the server via network using UDP in real time, with continuous polling. The port address of server, dynamic IP address and local port for binding is given as input to the client module. The UDP communication tool is taken from Instrument Control Toolbox in simulink. The data transmitted to server is a 1 byte character, label value for each of four words. The Word and their corresponding label values are shown in Table 1.

Table 1 Word and its Corresponding Values

WORD	START	SPEED1	SPEED2	STOP	BREAK
SOM LABEL	5	6	7	8	Exit
PULSE FREQUENCY	48Hz	24Hz	12Hz	0Hz	Exit

C. Network

The protocol used for network communication is User Datagram Protocol. The UDP is one of the core protocols of the Internet protocol suite. UDP provides communication service at an intermediate level between the client PC and the Internet protocol. The unassigned port in UDP ranges from 0 to 65535. The communication port and local binding port address used in our prototype module is 45000.

D. Server (UDP receive and Serial send)

The server module is a simulink file. It performs two different functions in real time.

1) *UDP receive:* The server simulink should be run before executing the client module. If this condition fails, the client will automatically quit, prompting "timeout error". The server will be in continuous polling and will check for any data in the specified port. If the data appears in the port of declared IP address of client, it passes the data to data transmission module.

2) *Serial send:* This block configures the serial port with the specified parameter (baud rate: 9600, Data bits: 8, Stop bit: 1, No parity) and sends the data once it gets from the UDP receive module.

E. Serial Port Interface

As most of the current day PCs are lacking serial port, A USB to RS232 converter is used for serial port communication which can be configured to any unused port to avoid hardware resource sharing.

F. Microcontroller-AT89S52

Microcontroller is a microprocessor designed specifically for control applications, and is equipped with ROM, RAM and facilities I / O on a single chip. AT89S52 is one of the family MCS-51/52 equipped with an internal 8 Kbyte Flash EPROM (Erasable and Programmable Read Only Memory), which allows memory to be reprogrammed. There are four input and output ports each consist of 8 bits. The clock frequency used is 12 MHz. The program is done in assembly language and burnt using KEIL C.

The data from the serial port is fed into SCON (Serial Communication) pin of microcontroller (AT89S52). The microcontroller decodes the value and saves it in the accumulator. Depending on the accumulator value different frequency is generated. The generated pulse is fed via port 2, which is connected to a low power DC motor. Thus for different values received from the client the speed of the motor will vary. The actuator can be directly driven by a microcontroller, if the power rating is less else a driver circuit has to be used to avoid loading from actuator.

IV. WORKING PRINCIPLE

The flow of data is explained in Fig.2. The speech recognition system captures the user voice in real time and recognizes the word spoken by the user after subtracting the ambient noise, with the help of trained SOM. The UDP send block is called by the Speech Recognition module which passes the appropriate label value mentioned in Table 1 to UDP receive block. The UDP receive block receives the data from UDP send block and sends it to Serial send block. The serial configuration block configures the serial port. The serial send block pushes the data received from the client to the Micro-controller. A look up table is developed in the microcontroller which generates frequency appropriate to the data received from the server. The speed of the actuator is controlled by the words, 'SPEED1' and 'SPEED2'.

V. HARDWARE SET UP AND EXPERIMENTAL RESULTS

A. *Hardware Setup*

To verify the architecture and working principle, a hardware prototype as show in Fig.7, is built using Microcontroller AT89S52. The code is developed in M-Script and a model is formed using simulink in -

Matlab.

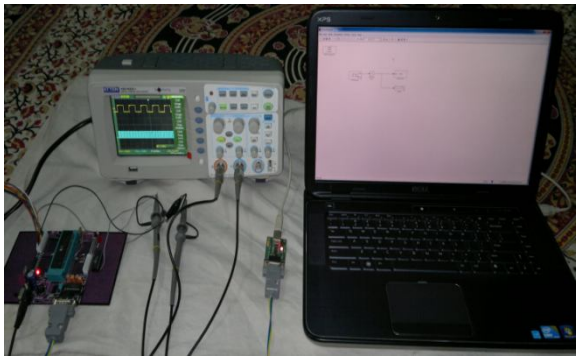
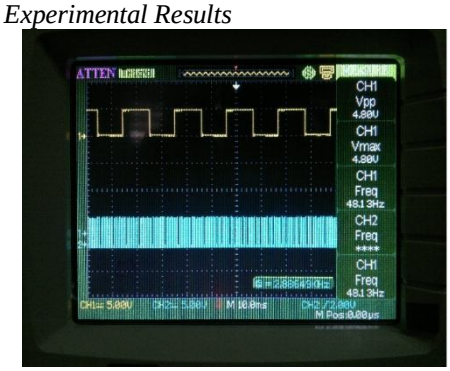


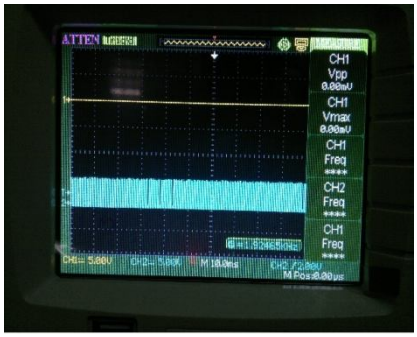
Fig.7. Hardware set up showing Server PC running Simulink model with Microcontroller and DSO waveforms

The communication via network is done using wireless router which assigns IP address dynamically to the PCs connected to the network. USB to RS232 is used in the hardware set up for serial communication with the microcontroller. A microcontroller with serial port interface is used for decoding of serial data from server. The microcontroller is programmed to receive the data from server at a baud rate of 9600bps. The pulse is generated at port 2 of the microcontroller. A DSO is used to check the pulse pattern. Pulse of frequency 48, 24, 12 and 0 HZ is generated for the data received from the client in real time. With help of these pulses, a motor can be actuated for different speed. The pulse waveform for 'START' and 'STOP' word are shown in the Fig.8 (a) and Fig.8 (b) respectively.

B. *Experimental Results*



a. Received Value (8 Bit) =5, Frequency =48Hz



b. Received Value (8 Bit) =8, Frequency =0Hz

Fig.8. Pulses from server for START and STOP word.

VI. CONCLUSION

Speech is captured and recognized in real-time with help of SOM developed in M-script and transmitted by UDP send block from Instrumentation and control toolbox of simulink. The obtained real time raw data is converted into frequency spectrum before analysis to increase the accuracy of SOM. transmitted. The server module is able to receive the data by UDP receive block and send it to Microcontroller (AT89S52) with serial send block. 'Enable Blocking Mode' in serial sends block of simulink model is disabled to reduce time delays between data transmission. Data size parameter in UDP receive block should be declared appropriately to avoid fluctuation in the pulse. Pulse of appropriate frequencies is generated for unique label value received from the trained SOM, which can be used for actuation. A lab prototype is built and all the experiments are conducted in real time. The results are obtained and found to be satisfactory.

VII. FUTURE WORK

The simplex communication method used in this paper can be done using Duplex communication through which real human-computer interaction can be visualized. To achieve more smoothing actuation, the no of words can be increased in the proposed topology can be increased. The SOM can be trained for different ascent and its performance can be evaluated. With some hardware modification, the proposed method can be used for Voice controlled home automation.

REFERENCES

[1] Punit Kumar Sharma, Dr. B.R. Lakshmikantha and K. Shanmukha Sundar, "Real Time Control of DC Motor Drive using Speech Recognition" , Power Electronics (IICPE), 2010 India International Conference.

[2] Shlomo Engelberg, Yishai Saidoff, and Yehezkel Israeli, "Voice Identification Through Spectral Analysis", Instrumentation & Measurement Magazine, IEEE Volume: 9, Issue: 5.

[3] Chenghui Yang, Weixin Yang and Shuwen Wang, "Based on Artificial Neural Networks for Voice Recognition Word Segment", Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference.

[4] Akshay Gupta, "Synthesis and Performance Analysis of a Recurrent Fuzzy Multilayer Perceptron for Speech Recognition", Methods and Models in Computer Science (ICM2CS), 2010 International Conference.

[5] TAbushariah, A.A.M.,Gunawan, T.S., Khalifa, O.O. and Abushariah, M.A.M., "English Digits Speech Recognition System Based on Hidden Markov Models", Computer and Communication Engineering (ICCCE), 2010 International Conference.

[6] Hasnain, S.K and Awan, M.S., "Recognizing Spoken Urdu Numbers Using Fourier Descriptor and Neural Networks with Matlab", Electrical Engineering. ICEE 2008, Second International Conference.