Micro-controller based Remote Monitoring using Mobile through Spoken Commands

Naresh P Jawarkar
Electronics Engg. Department,
B. N. College of Engineering,
Pusad – 445 215, Maharashtra, India
Email: npjwarkar@rediffmail.com

Vasif Ahmed
Electronics Engg. Department
B. N. College of Engineering,
Pusad – 445 215, Maharashtra, India
Email: ahmedvasif@gmail.com

Siddharth A Ladhake
Principal,
Sipna's College of Engg. & Technology,
Amravati – 444602, Maharashtra, India
Email: sladhake@yahoo.co.in

Rajesh D Thakare
Electronics Engg. Department,
Priyadarshini College of Engineering,
Nagpur – 440 001, Maharashtra, India
Email: rdt1231@indiatimes.com

Abstract— **Mobile phone can serve as powerful tool for world-wide communication. A system is developed to remotely monitor process through spoken commands using mobile. Mel cepstrum features are extracted from spoken words. Learning Vector Quantization Neural Network is used for recognition of various words used in the command. The accuracy of spoken commands is about 98%. A text message is generated and sent to control system mobile in form of SMS. On receipt of SMS, control system mobile informs AVR micro-controller based card, which performs specified task. The system alerts user in case of occurrence of any abnormal conditions like power failure, loss of control, etc. Other applications where this approach can be extended are also discussed.**

Index Terms— **Mobile phone, Short Messaging Services (SMS), AT commands, Mel coefficients, LVQ, AVR microcontroller**

I. Introduction

There has been tremendous rise in number of mobile customers in world (> 2 billion). Due to widespread growth of cellular network and drastic reduction in call rates and lower-end handsets, mobile usage has percolated all sections of society. Latest mobiles can not only allow you to click pictures, play music, store your address book, hook you to Internet, download e-mails, guide you through maze of city streets but also enable you to watch latest blockbusters, favorite TV programs, book tickets and transact business. As a result, large section of population is continuously upgrading their mobile models to avail these exquisite features. The dumped older models having just voice call and messaging facility are available at throwaway prices.

Any mobile having messaging facility and capability to support common AT commands can be used in this system. Nokia model 6610 is chosen because it supports AT commands. Most mobile manufacturers like Siemens, Motorola, LG, Samsung, Sony-Ericsson also provide AT command compatibility.

SMS is store and forward way of transmitting messages to and from mobiles. Each short message should not be longer than 160 characters (text / binary). Since SMS uses signaling channels for its transmission/ reception instead of dedicated data channels, these messages can be sent/ received simultaneously with voice/fax/ data services over GSM network. The major advantage of using SMS is provision of intimation to the sender when SMS is delivered at the destination and ability of SMSC to continue efforts for delivery of message for the specified validity period if network is presently busy or called user is outside the coverage area.

It is observed that many mobile users especially older generation find it inconvenient to use mobile keypad for text entry as it involves continuous pressing of many keys for alphabets and is time consuming. In order to relieve the users from this burden, spoken words are used to send commands for control.

A system is developed for remote monitoring and control of devices using mobile through spoken commands. The system offer several attractive features like—

- control from anywhere in world if cellular coverage is available
- acknowledgement about execution of command from system to user
- uses spoken commands from user for control
- alerts user on occurrence of any abnormal conditions like power failure, parameters exceeding prescribed limits, etc.
- ease of implementation and cost-effective approach.

The Block Diagram of the scheme is shown in Fig. 1. On the user side, microphone is used to translate the voice signal to electrical signal. The microphone is connected to MIC interface of sound section on motherboard of Pentium IV based PC. User mobile is connected through DKU-5 cable using USB port. In this

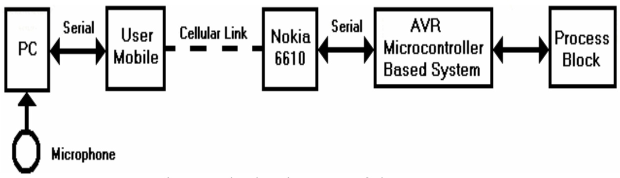

Fig. 1. Block Diagram of the System

approach, predetermined phrases of words are selected for various commands. The Mel cepstrum features are extracted from the spoken words for recognition. Mel cepstrum exploits auditory principles as well as discriminating property of the cepstrum and is proven to be one of the most successful feature representations in speech related recognition tasks [1, 2]. The spoken words are isolated and recognized after extraction of features. Learning Vector Quantization Neural Network is used for recognition of various words used in the command. A text message is generated if all spoken words are identified as per specified format. This message is transmitted in form of SMS to control system mobile using AT commands [3].

On control side, system mobile is connected to AVR micro-controller based system through RS-232C cable. Process block consists of 8 digital output ports, 8 digital input ports and one analog input port. The configuration of number of inputs, output and analog input ports can be varied as per the needs of the applications. Presently, LEDs are used to indicate status of output digital ports, dip switches to change the status of input digital ports, and potential divider provided to vary analog input voltage.

II. AT Commands

Now-a-days extensive list of mobile related AT commands are available for carrying out various activities like sending SMS, using GPRS services, sending fax, controlling speaker volume, battery status indication, etc [4-6]. AT commands require sending of text strings 'A', 'T', along with specified command strings through serial port to mobile and are executed on receipt of carriage return. The result codes are sent by mobile to Terminal Equipment (TE) to indicate the response after execution of command.

The text message is sent to mobile using CMGS commands. CNMI command is used to indicate to TE about the receipt of incoming SMS message from the network. On receipt of the SMS message, text words are checked with predetermined format, which includes password, desired device ON/OFF commands or status query. After interpretation of valid control message, microcontroller carries out the specified tasks and then sends SMS to pre-specified mobile number as acknowledgement of fulfillment of command or reporting of error during execution of command. There are varieties of commands available at our disposal like directly storing various predefined messages in phone memory, sending messages at appropriate time by calling the relevant message number depending on present conditions, storing incoming SMS in phone memory, deleting the message after execution of command, etc. But it was decided to discard these features to ensure easy adaptation to any mobile model having limited AT commands interpretation capability. So in our case, any incoming SMS message is directly routed to micro-controller (TE) and any outgoing text message is directly sent by micro-controller to designated mobile number without being stored in control system mobile phone memory.

III. Microcontroller System

AT Mega32 microcontroller is part of AVR series of microcontroller manufactured by Atmel with RISC architecture, 32 KB of in-system programmable Flash, 1K E^2PROM, 2K SRAM, 32-bit multi-function General purpose I/O, TWI, USART, SPI, JTAG interface support, etc [7]. Atmel also provides free software support in form of AVR 'C' compiler, AVR Studio for software debugging and simulation in assembly language and Ponyprog software is also available for flash programming [8].

The interface diagram of micro-controller system is shown in Fig 2. 8-bits of Port C are configured as digital inputs ports and in our case, their statuses depend on the position of corresponding dip switch. 6 bits of Port D (PD2 – PD7) and two higher bits of Port B (PB6 –PB7) are configured as digital outputs and corresponding LEDs indicate their status (ON for logic '0'). Lower 6 bits of Port B are connected to 2 × 16 characters LCD display in 4-bit data length mode. In this application, mobile 6610 is connected to micro-controller based system through RS-232C data link cable. Internally RxD and TxD are connected to 9-pin RS232 female connector through MAX 232 IC for TTL-RS232C signal translation. PA0 bit is connected to potential divider circuit to provide variable analog voltage at its input. The crystal of 1.8432 MHz is chosen for generating baud rate of 115.2 kbps for serial communication with mobile.

AT Mega32 is initialized to configure port C as input and Port D (lower 6-bits) and Port B as outputs. Universal Baud Rate Register (12 bit) UBRRH & UBRRL are initialized to 000H for baud rate of 115.2 kbps. The frame format of 8-bit, no parity and 1 stop bit is set using Universal Control & Status Register C (UCSRC). Receiver and transmitter are enabled through appropriate bits in UCSRB. For Transmission, it is recommended to check the empty status of transmit buffer with the help of UDRE flag in UCSRA. If the transmit buffer is empty then data byte to be transmitted to TxD pin is sent through Universal Data Register (UDR). For serial data reception, RxC bit in UCSCRA is checked at regular intervals and if it is active, incoming data byte from RxD pin is read through UDR.

For enabling SMS reception from control system mobile, microcontroller sends

1. 'ATZ' command to reset modem of control system mobile
2. 'AT' command to check modem status
3. 'AT + CGMF' command to select text mode
4. 'AT + CNMI' command to set incoming SMS message indications

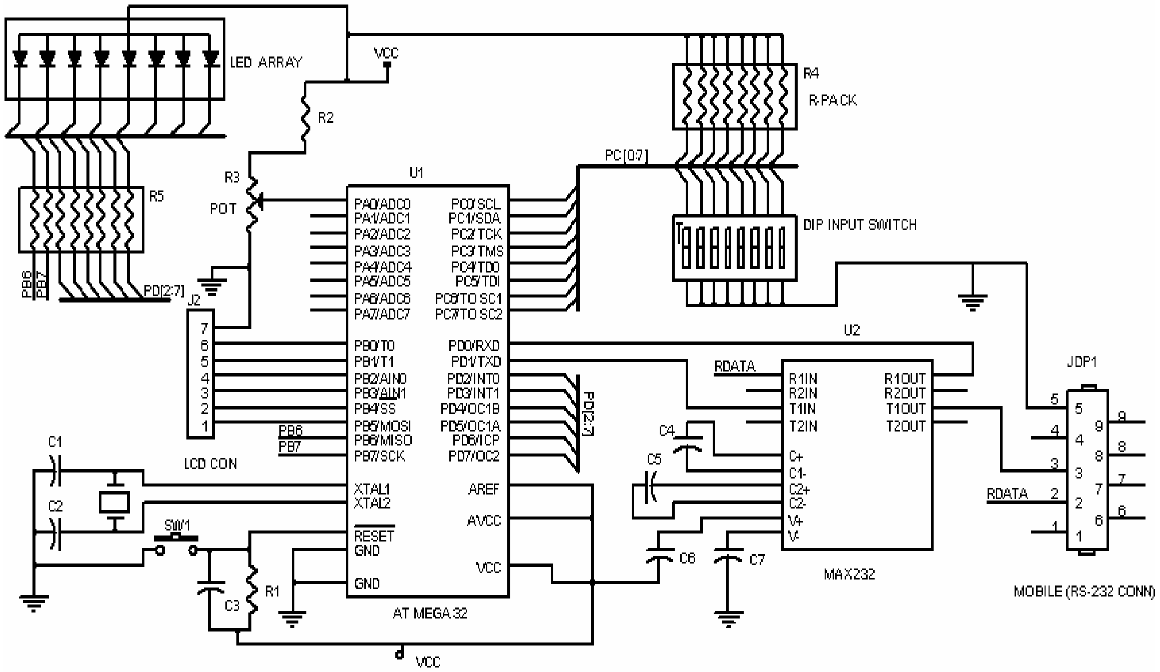

Fig. 2. Interfacing diagram of Micro-controller based card.

If the modem of control system mobile fails to respond to any of these commands, microcontroller sends “COMMUNICATION FAILURE” message to LCD display.

Micro-controller then checks digital input port bits and analog input at periodic intervals and sends SMS whenever there is change in status of input port bits or deviation of analog input from prescribed limits. Whenever control system mobile receives SMS, it sends unsolicited result code along with text message to the microcontroller through serial interface. Microcontroller interprets the text message and carries out the specified command. For example, if command to set device 6 ON is received, it sets fifth bit of Port D and sends SMS to user mobile through control system mobile using ‘AT + CGMS’ command to inform fulfillment of command.

IV. Spoken Commands Recognition

A voice signal contains psychological and physiological properties of the speaker as well as dialect differences, acoustical environment effects and phase differences [9]. There are many approaches for speech recognition [10-12]. A feature vector can represent each spoken word. It is decided to use two types of spoken command phrases as valid text message for this application as shown in Table I. Presently thirteen words namely alpha, device, one, two, three, four, five, six, seven, eight, on, off and status are used for recognition of commands. Initially all possible words in the selected command phrase are spoken and stored as speech templates and features are extracted. These features are used as inputs to Learning Vector Quantization (LVQ) based neural network for training. After successful completion of training, the system is tested for recognition of spoken words.

The block diagram of the feature extraction and processing for speech recognition is shown in Fig. 3. The basic steps in the feature extraction includes

A. Preprocessing:

The speech signal s(t) from microphone is passed through pre-amplifier. In-built 16-bit ADC on motherboard of PC then digitizes the filtered speech signal at sampling frequency of 22.05 KHz. Short-time energy in speech signal is used for detecting end points to isolate words in a phrase. Then speech signal is passed through a low pass first order pre-emphasis filter to spectrally flatten the original signal and to make it less susceptible to finite precision effects later in the signal processing [13]. The transfer function of the pre-emphasis filter is

$$H(z) = 1 - 0.95z^{-1}$$

The pre-emphasized speech signal is divided into frames of 512 samples with 50% overlapping. Then each frame is windowed to minimize the signal discontinuities at the beginning and end of each frame using Hamming window.

TABLE I

Choice Of Spoken Command Phrases

No.	Pass-word	Option	Param-eter1	Para-meter2	Remarks
1	Alpha	Device	1/2/3/4/5/ 6/7/8	ON / OFF	Select Device & switch it ON/OFF
2	Alpha	Status	1/2/3/4/5/ 6/7/8	----	Read Status of Input

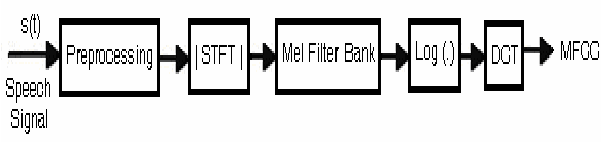

Fig. 3. Block Diagram of Speech Extraction & Processing

B. Mel Frequency Cepstral Coefficient (MFCC) Computation:

The Mel-cepstrum exploits auditory principles as well as decorrelating property of the cepstrum. In MFCC implementation, triangular filters are used. These filters follow Mel scale whereby band edges and corner frequencies are linear for low frequencies (<1000 Hz) and logarithmically increases with increase in frequency. FFT of each frame is taken to get the spectrum. A magnitude spectrum is computed and frequency warped in order to transform the spectrum into the Mel frequency in which filter bank is uniformly spaced. 24 Mel-scale triangular filters are multiplied with the magnitude spectra of the frame [2]. The energy in the Short Time Fourier Transform (STFT) weighted by each Mel-scale filter is given by

$$E_{mel}(l) = (1/A_l) \sum_{k=L_l}^{U_l} | V_l(\omega_k) \times S(\omega_k) |^2$$

where $V_l(\omega_k)$ is the frequency response of l^{th} filter, U_l and L_l denote the lower & upper energy indices over which filter is non-zero.

A_l normalizes the filter according to its varying bandwidth to give equal bandwidth of flat linear spectrum.

$$A_l = \sum_{k=L_l}^{U_l} | V_l(\omega_k) |^2$$

Finally discrete cosine transform (DCT) of filter bank coefficients is taken to get MFCC as under –

$$C_{mel}(k) = \sum_{l=1}^{N} \log\{E_{mel}(l)\} \cos[(k(l-0.5)\pi /N], \ k= 1, 2, ...12$$

where $\log\{E_{mel}(l)\}$ is log filter bank energies & $C_{mel}(k)$ is the k^{th} MFCC and N is number of filters.

It has been observed that performance is reasonably well for 24 filters in MFCC implementation [14].

If there are n frames in a word and 12 MFCCs are computed for each frame, we get feature vector of length $12 \times n$. However, the number of frames (n) varies from word to word, which in turn changes the length of feature vector. In order to obtain feature vector of constant length, n values of each Mel Frequency Cepstral Coefficient are converted into 10 values using resampling technique. Thus for each word, constant length feature vector of 120 (12×10) elements is obtained. Principal Component Analysis (PCA) is carried out on the MFCC data thus obtained. PCA transforms the input data so that the elements of the input vectors are uncorrelated.

C. LVQ Classifier :

The LVQ is an algorithm for learning classifiers from labeled data samples. It models the discrimination function defined by the set of labeled codebook vectors and the nearest neighborhood search between the codebook and data. In classification, a data point x_i is assigned to a class according to the class label of the closest codebook vector. The training algorithm involves an iterative gradient update of the winner unit [15, 16].

The winner unit $\mathbf{w}^c$ is defined by

$$c = \arg\min_k || \mathbf{x}_i - \mathbf{w}^k ||$$

The update equation for the winner unit $\mathbf{w}^c$ defined by the nearest neighbor and a data sample $\mathbf{x}(t)$ is

$$\mathbf{w}^c(t+1) = \mathbf{w}^c(t) \pm \alpha(t) [\mathbf{x}(t) - \mathbf{w}^c(t)]$$

where sign depends on whether the data sample is correctly classified (+) or misclassified (-) and $\alpha(t)$ is learning rule and must decrease monotonically in time.

V. RESULTS

Fig. 4 shows the speech signal waveform for a spoken command phrase “Alpha Device Six On” and its energy. Fig. 5 shows the plot of MFCC coefficients $\{C_{mel}(1), C_{mel}(2), C_{mel}(3), C_{mel}(4)\}$ for the spoken word “Alpha”.

Fifty samples of each spoken word are stored out of which, 25 are used for training and remaining for testing. Thus database of 650 samples of words is used for experimentation. Accuracy of correct recognition for various words in spoken commands with principal component analysis (PCA) is shown in Table II. The accuracy for the words ‘on’ and ‘one’ is relatively less because of phonetic similarity in these words. However, these words can easily be discriminated due to difference in their utterance positions in spoken command phrase.

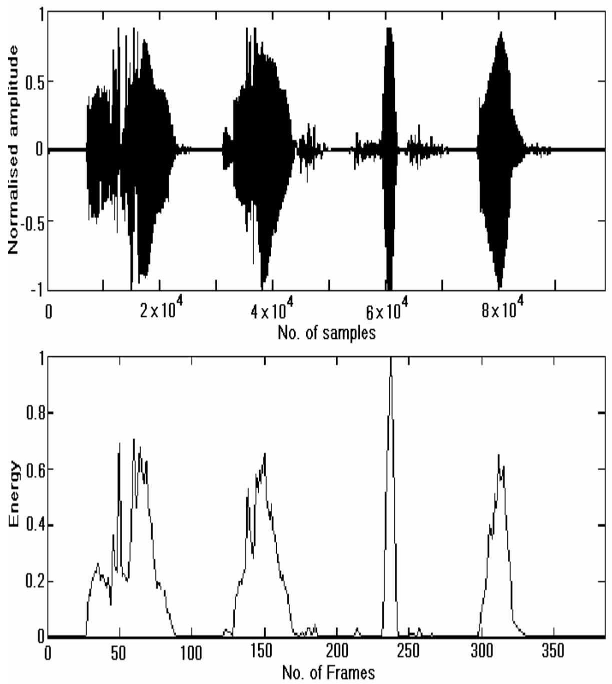

Fig. 4. Plot of spoken phrase ‘alpha device six on’ and its energy

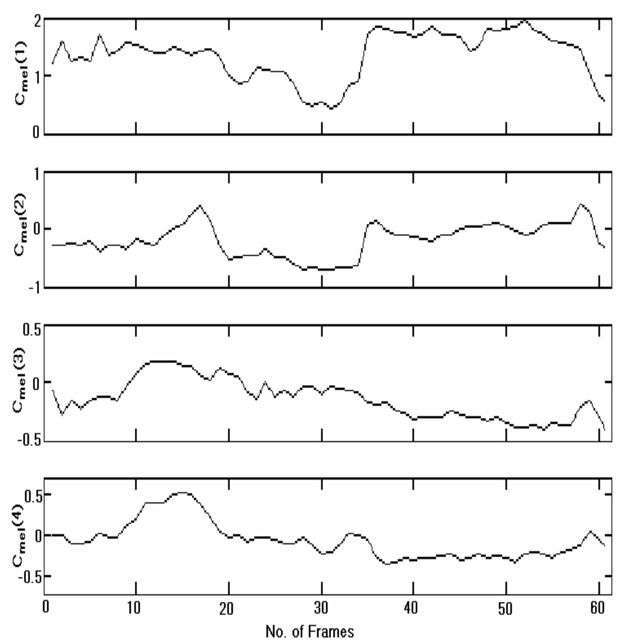

Fig. 5. Plot of MFFC Coefficients for spoken word 'alpha'

VI. Conclusions And Further Recommendations

Remote control of devices and retrieval of information relating present status of inputs using spoken commands have been successfully demonstrated through SMS based message transfer between user mobile and system mobile. There is scope for lot of improvement depending upon the user requirements like inclusion of greater number of desired commands, selection of suitable sensor for measurement of analog parameters, etc. This approach can be easily extended to develop many exciting products from remote process control to high-end security solutions. It can prove to be great boon to blind / physically handicapped persons due to its capability for remote control through speech commands. Many mobile companies are offering free SMS services or highly attractive schemes for SMS to divert customers from congested voice/data channels to signaling channels.

TABLE II
Accuracy Of Word Recognition

Spoken Word	Accuracy
Alpha	100%
Device	100%
Status	100%
ON	92%
OFF	96%
One	92%
Two	100%
Three	100%
Four	96%
Five	100%
Six	100%
Seven	96%
Eight	100%
Av. Accuracy	**97.85%**

However, this method is not suitable for time-critical applications as message transfer time to destination is variable. This problem can be alleviated to certain extent by adding time at which device should respond to message and sending the SMS well in advance before the scheduled event. The software can be modified in this case to check if timing information is present and accordingly schedule that event. Alternatively, it is recommended to use data calls (Fax/ GPRS) or DTMF based calls for immediate response from system with suitable modifications. The accuracy of spoken commands recognition system is about 98% (much better than our previous work [17]). Moreover, spoken phrase can be extended to carry out additional tasks like adding time duration for Device ON/OFF condition. Further, adding the speaker verification feature can enhance security level. Presently PC is used for generation of text message from voice command. For dedicated applications, PC can be replaced by DSP processor/ FPGA based system with higher initial development cost.

Acknowledgement

We thank the Principal, Prof. N M Gatphane of B N College of Engineering, Pusad for providing necessary facilities for carrying out this work. We acknowledge the diligent efforts of our students Sumit Mundhada, Manjit Sharma, & Rohan Patil in assisting us towards implementation of this idea.

References

[1] S. B. Davies and P. Mermelstein, "Comparison of Parametric Representation for Monosyllabic Word Recognition in Continously spoken Semantics", *IEEE Transanction on Acoustics, Speech & Signal Processing*, vol ASSP-28, Aug 1980, pp 357-366.

[2] Thomas F. Quatiero, "*Discrete time Speech Signal Processing – Principles and Practice*", Pearson Education (Singapore) Pvt. Ltd., Indian Branch, Delhi, India, 2004.

[3] N. P. Jawarkar, Vasif Ahmed & R. D. Thakare, "Remote world-wide control through SMS using Nokia Mobile", *IETE Journal of Education*, Vol 46, No. 4, Oct-Dec 2005, pp 165-170.

[4] http://www.etsi.org

[5] http://forum.nokia.com, AT Commands Set for Nokia GSM and WCDMA products, Version 1.2, July 2005.

[6] http://gatling.ikk.sztaki.hu/~kissg/gsm/at+c.html

[7] http://www.atmel.com/avr

[8] http://www.lancos.com/

[9] M. Bilginer Gulmezohu, Vakif Dzhafasov, Mustafa Keskin & Atalay Barkan, "A Novel Approach to Isolated Word Recognition", *IEEE Transactions on Speech & Audio Processing*, vol 7, No. 6, Nov 1999, pp 620-627.

[10] V.W. Zue, "The use of speech knowledge of ASR", *Readings in Speech Recognition*, A Waibel and K F Lee, Eds, San Mateo, CA: Morgan Kaufmann, 1990, pp 200-214.

[11] L. R. Rabiner, B. S. Atal and M. R. Sambur, "LPC Prediction on Error Analysis in variation with position of analysis frame", *IEEE Transactions on Acoustics, Speech & Signal Processing*, vol ASSP-25(5), Oct 1977, pp 434-442.

[12] G. M. White and R. B. Neely, "Speech Recognition Experiments with Linear Prediction, Bandpass filtering & Dynamic Programming", *IEEE Transactions on Acoustics, Speech & Signal Processing*, vol ASSP-24(2), 1976, pp 183-188.
[13] L. R. Rabiner and B. H. Juang, "*Fundamental of Speech Recognition*", Pearson Education (Singapore) Pvt. Ltd., 2005.
[14] S. Umesh, L. Cohen and D. Nelson, "Frequency warping and Mel scale". *IEEE Signal Processing Letter*, vol. 9, No. 3, March 2001, pp 104-107.
[15] Hollmen V. Tresp and O. Simula, "A Learning Vector Quantization Algorithm for Probabilistic Model", *Proceedings of EUSIPCO 2000 – X European Signal Processing Conference*, Volume II, pp 721-724.
[16] Kohonen T, "Improved version of Learning Vector Quantization", *International Joint Conference on Neural Networks*, San Diego, CA, 1990, pp z:545-550.
[17] N. P. Jawarkar, Vasif Ahmed and R. D. Thakare, "Remote Control using Mobile through Spoken Commands", *Proceedings of IEEE-ICSCN 2007*, MIT Campus, Anna University, Chennai, Feb 22-24, 2007, pp 622-625.

Naresh P Jawarkar had received M. Tech degree from Vishweswaraya Regional College of Engineering, Nagpur in 1986. His main areas of interest are speech processing, microprocessor applications, control systems and soft computing.

He has about 21 years of teaching experience. Presently he is working as Professor and Head, Electronics & Telecommunication Engineering Department at Babasaheb Naik College of Engineering, Pusad. He has organized many Training programmes/ conferences/ workshops/ seminars in the field of Electronics Engineering. He was member of Board of Studies of Nagpur and Amravati Universities. He has number of papers published in refereed journals and conference proceedings.

Prof. Jawarkar is Fellow of Institution of Electronics & Telecommunication Engineers (IETE), India; Fellow of the Institution of Engineers, India (IEI) and life member of Indian Society for Technical Education (ISTE).

Vasif Ahmed completed M.E. course from Maharaja Sayajirao University, Baroda in 1989. His major areas of interest are PC based control applications, mobile based applications, and Electronic Design Technology.

He has worked as Technical Partner in private firm 'AV System', Ahmedabad. In 1993, he joined Electronics & Telecommunication Engineering Department of Babasaheb Naik College of Engineering, Pusad and is presently working as an Assistant Professor. He has papers published in refereed journals and conference proceedings.

Mr Ahmed is Fellow of IEI society and member of ISTE.

Siddarth A Ladhake has completed Ph. D from Amravati University, Amravati. His main areas of interest are Digital System Design and VLSI. He is presently involved in research of multi-valued logic and code converters.

He has about 25 years of teaching experience and 1 year of industrial experience. Presently he is working as Principal, Sipna's College of Engineering & Technology, Amravati. He has number of papers published in refereed journals and conference proceedings. He has guided 07 students at M.E. level and 4 students are pursuing Ph. D. program under his supervision.

Dr. Ladhake is Fellow of IETE society and member of IEEE, IEI and ISTE professional bodies and is presently serving as Chairman of IETE, Amravati sub-centre.

Rajesh D Thakare had received M.E (Electronics) degree from Amravati University, Amravati in the year 2001. His main areas of interest are PC and microprocessor based control applications and Digital signal processing.

He has about 17 years of teaching experience. Presently he is working as Assistant Professor in Electronics Engineering Department at Priyadarshini College of Engineering, Nagpur. He has number of papers published in refereed journals and conference proceedings.

Mr Thakare is member of IETE and ISTE societies and is presently working as an executive member of IETE, Amravati sub-centre.