RESEARCH ARTICLE

VOICE RECOGNITION CONTROLLED ROBOTIC VEHICLE

Nishant Srivastava, Anand Singh, *Vishwajeet Singh, Bhanu Vikram Singh and Ankit Pandey

Rajarshi Rananjay Sinh Institute of Management and Technology, Amethi (U.P), 227405, India

ARTICLE INFO	ABSTRACT
	This paper describes of about voice recognition controllable robotic vehicle which works by virtue of human voice commands for remote area operations. It is a mobile robot whose motions can be controlled by user using specific voice commands. The specific commands are firstly to be registered in the database & then after processing the speech commands, the necessary motion instructions are to be given to the mobile platform via Radio Frequency link. The following paper carries the detailed information of the prototype of the robotic vehicle. The commands are to be sent to the robotic vehicle using push buttons or voice commands present. The receiver would decode the data before feeding it to another microcontroller, to drive DC motors via motor driver Integrated Circuit for required work. This technology has an advantage over long communication range as compared to Radio Frequency technology. Further the project can also be developed using IoT (Internet of Things) technology where a user can control the robot from any corner of the world.

INTRODUCTION

Here in proposed design of robotic vehicle, the aim is to control the movements of the vehicle prototype using voice commands from the user. These commands will be directed at the specific Android Application on the user's phone which is connected to the robot using a Bluetooth Module. The commands issued will then be relayed over an Radio Frequency (RF) channel and will be received by the Module. The aim of Voice Controlled Robotic Vehicle (VCRV) is to listen and act accordingly on the commands received from the user. Here, the system would require the specific training from the user (for the accent of voice) after which the device will start understanding the commands issued. This is accomplished by adding required commands to the microcontroller through a well-defined code. The special feature of the software is to train itself for the voice commands. The graphical user interface running along with software provides a convenient method for the users to train. Furthermore, this bio-electronics project can also be enhanced& improved using DTMF technology.

All about Voice/Speech Recognition

It is defined as the process of enabling a computer to identify and respond to sound produced in human speech.

Corresponding author: **Vishwajeet Singh,**
Rajarshi Rananjay Sinh Institute of Management and Technology, Amethi (U.P), 227405, India.

Voice recognition is the process of taking spoken word as an input to the program, Speech signals are captured coming from the microphone or telephone attached & converting them into a digitally stored set of words. It is the process of capturing spoken words and commands, it is the ability of the machine to receive and interpret dictation, or to understand and carry out spoken commands. Both speech & voice recognition basis use recordings of human voice (user). Voice recognition stripes out personal differences in frequency of voice to detect the words particularly. Speech recognition criteria typically disregards the language & meaning to detect the person behind the speech (voice). For the project, to make it user friendly, Voice Recognition is the best methodology to control the robotic vehicle. Two crucial factors that decide the accuracy of the proposed voice recognition system are

- Accuracy in detecting the human words
- Processing those words at the desired speed so that the commands are executed with the least delay

Specifications of Voice Recognition Controlled Robotic Vehicle

- For training, Artificial Neural Networks is used
- Software is written using MATLAB
- For identifying the words, Linear Predictive Coding (LPC) is used, popular method for extracting speech characteristics from sample values
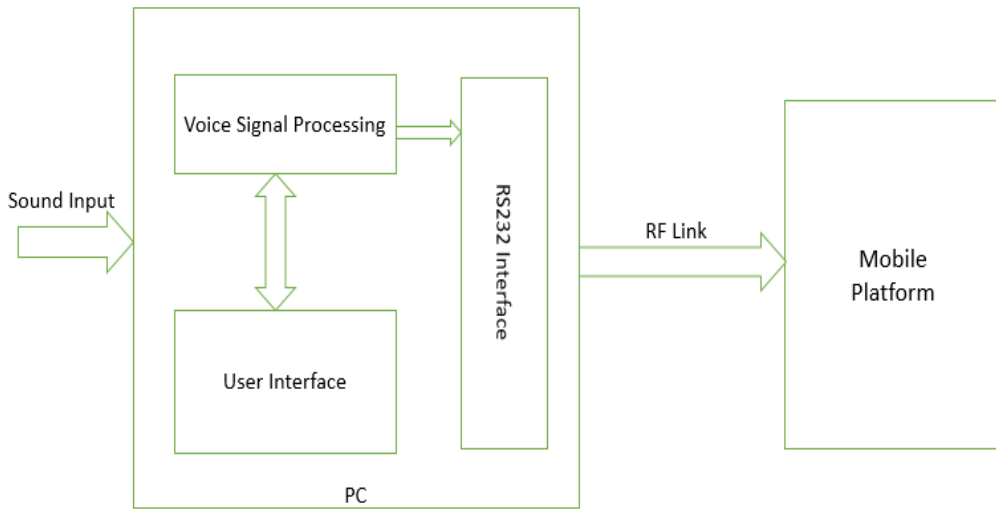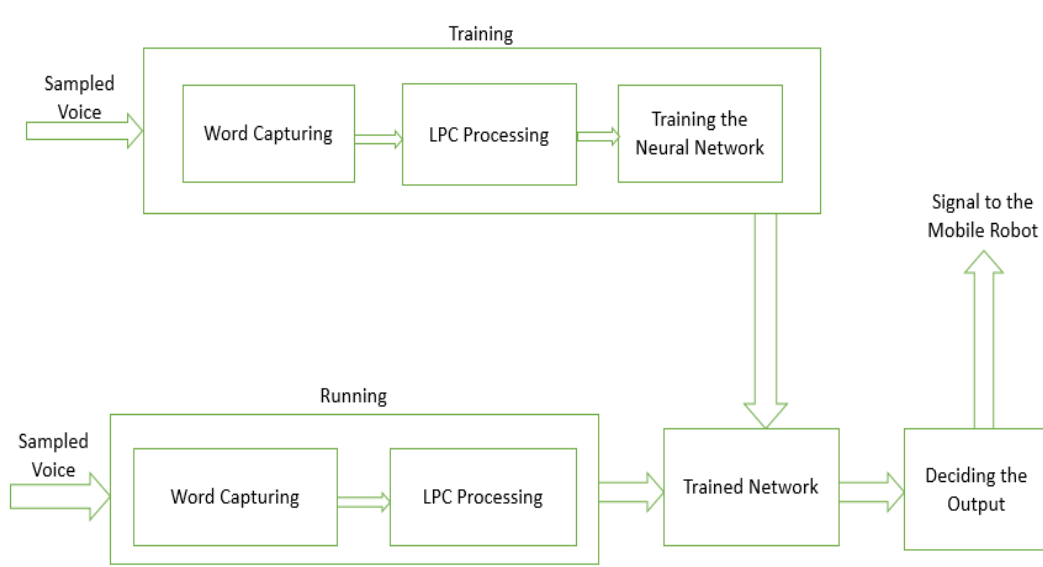
Fig. 1. Vehicle linkage Prototype

Fig. 2. Flowchart displaying working of Robotic Vehicle

Firstly, a spoken word is recognized & then a set of parameters called "Cepstral Co-efficient" is calculated from the voice data samples belonging to that based on LPC. Then those coefficients are processed by the trained neural network to decide whether that word is any of the specific commands in database. If the word is identified as any one of the commands, then a relevant signal is sent to the mobile platform via RS232. The overall operation is illustrated in the diagram above:

Description of robotic vehicle

The Voice Controlled Robot is controlled through voice commands given by the user who is operating the project. These voice command needs to be given through an android app which is installed on the user's android mobile. Please note that user should have good internet connection in order to have a smooth operation of the android application. Speech recognition is done within the android app and then a respective command is sent to the voice controlled robot vehicle. Microcontroller fitted on the Robot decodes these commands and gives an appropriate command to the motors connected to the robot. In voice operated robot the robotic vehicles movement is controlled via voice command. This is android base project.

In this project we use Bluetooth device to receive command from user. User use android application to give command to Bluetooth device. Bluetooth device receive command from application and transmit same to 8051 microcontrollers, then microcontroller controls the robotic vehicle as per command. Microcontroller is the main unit of the project. It's a central processing unit (CPU) of the Robot. It receives various commands from Bluetooth decoder and gives the respective output via motor driver ICs.
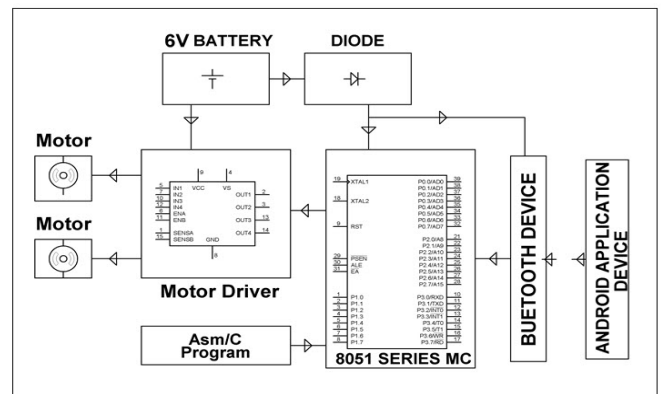

Fig. 3. Graphical Description of Robotic Vehicle

This robotic vehicle operated on commands like forward, reverse, left, right, stop, left U turn, right U turn. The communication between android application and Bluetooth is serial communication. Bluetooth decoder is always connected to Robot and another mobile phone is used to control the movements of the Robot. It gives ASCII code output. This receiver enables wireless transmission & reception of serial data. It has 10 meters range. DC motor is used for robotic vehicle movement

Transmitter

On the Transmitter section, commands are given to the Mobile Application through the micro-phone of the mobile handset. This mobile handset is connected to the moving vehicle via Bluetooth module. The mobile application used, is programmed in such a way that the voice commands given to the handset are received by the micro-phone and these analogy voice commands are converted to digital word sequences (A to D conversion). These stored sequences are than transmitted to the robot via Bluetooth transceiver module and are sent to the transceiver controller (MAX 232).
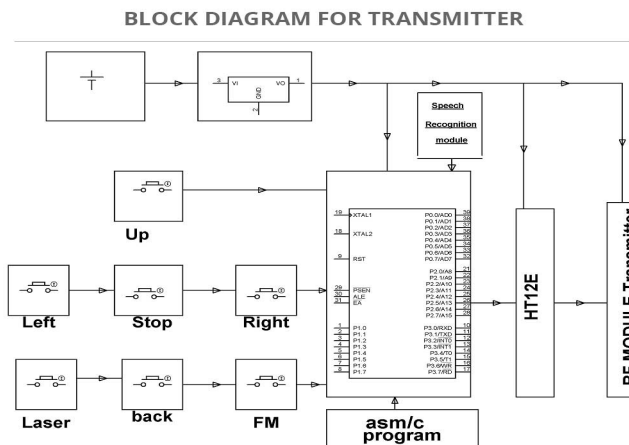

Fig. 4. Transmitter's Block Diagram

Receiver

MAX 232 transceiver is used to decode the received signal and for serial communication with the Bluetooth module. The controller compares these digital signals with the stored programme commands in it and convert them into voice strings. The voice strings are then used to run the servo motors for the desired interval of time.

Transducers

MQ6 Gas sensor are used to detect the gases in the vicinity. As soon as a gas is detected it turn on the buzzer and the level of gas is shown on the LCD. LM35 is the temperature sensor used to detect the temperature in the desired vicinity. The sensor detects the temperature and if it exceeds 50°C it turns on the buzzer. Thus, it acts as temperature detector.

Word Capturing

The signals coming from the microphone is processed only when something is spoken. The program waits until the sample value exceeds some threshold value. When the program is triggered by a significant example, a no. of samples is captured to process. After that, to determine the actual boundaries of the

word spoken, "edge detection" is performed. Here the centre of gravity of the energy distribution of the signal is calculated & then from the point intervals where the amplitude level lies below a threshold level are removed. Finally, we can have a set of voice samples corresponding to a particular word free of silent periods.

Linear predictive coding processing

The steps followed for the extraction of speech characteristics from captured samples using LPC is described as follows: --

Pre-Emphasis

This operation is necessary for removing DC & low frequency components of the incoming speech signal. It also makes the signal spectrum flatter. It is done using a first order FIR filter which can be described by the transfer function,

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

Here we used a = 0.9, FIR filtering was applied to the signal in the time domain using the MATLAB function "filter".

Frame Blocking

Each signal is now converted into a set of fixed length frames, with some no. of samples in each frame is overlapping. If the frame length is L & each frame is shifted by M samples away from the adjacent frame, then nth frame can be denoted by,

$$X_n[i] = s [(n-1) * M + i]$$

where n = 1,2,3 …………………… N
and i = 1,2,3 …………………… L.

Windowing

Each individual frame is windowed to minimize the signal discontinuities at the borders of each frame. We used the "Hamming Window" for this purpose. The set of samples for each frame is multiplied by the time domain version of the Hamming Window with size equal to the frame length.

LPC Calculation

First step of calculating LPC parameters is to get the auto-correlation vector for each frame. If the order of the auto-correlation is P, then the auto-correlation vector, 'r' can be given as

$$r(m) = \sum^{N-1-m} x(n) * x (n+m)$$

where, m = 0,1,2 ………………… P & x(i)s (i = 1,2,3 ………………………… L) are sample values in the windowed frame. Then Hermitian Toeplitz matrix of a is computed as shown below: --

$$A = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & \cdots & a_{-(n-1)} \\ a_1 & a_0 & a_{-1} & \ddots & & \vdots \\ a_2 & a_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & a_{-1} & a_{-2} \\ \vdots & & \ddots & a_1 & a_0 & a_{-1} \\ a_{n-1} & \cdots & \cdots & a_2 & a_1 & a_0 \end{bmatrix}$$
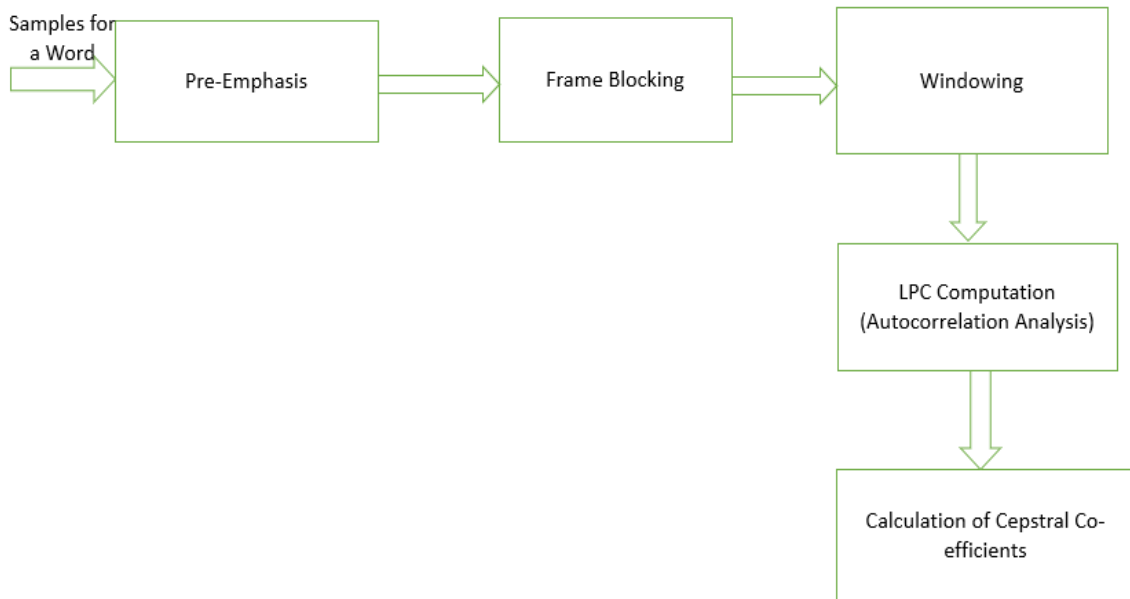
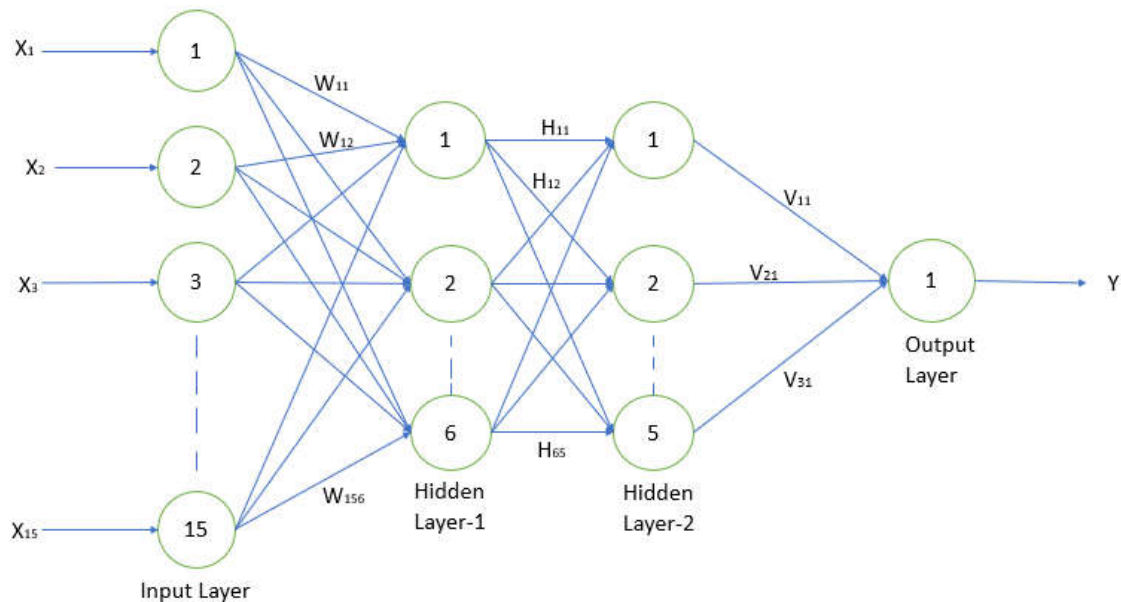**Fig. 5. Module for Decoding Speech/Voice using LPC**

Fig. 6. Neural Network Model

Finally, the LPC parameter matrix, 'x' is calculated by matrix multiplication of inverse of A and a.

$$x = A^{-1} * a$$

In our software, we used order (P), as 10.

Cepstral Co-Efficient Calculation

Cepstral co-efficient are Fourier transform representation of the log magnitude spectrum. Use of Cepstral co-efficient makes the application more robust & reliable. As low order, LPC parameters are too sensitive to the spectral slope & low order parameters are sensitive to noise, Cepstral co-efficient are weighted by a tapered window. So weighted set of Cepstral co-efficient, 'C_m' are found by,

$$C(m) = [1+(Q/2) \sin(\pi m/Q)] * C(m) \text{ where } Q = 1.5 * P$$

For each frame, belong to the captured word, this sequence of operations are performed & finally we end up with a no. of sets of Cepstral co-efficient. Next values for all the frames are averaged to get a single set of Cepstral co-efficient for that spoken word.

Artificial neural network

To recognize the speech mobile robot should have intelligence. Hence artificial neural network is used to make intelligence through the learning process. Here we used Multi-Layer Feed Forward Network & error back propagation learning algorithm to train the network. This network consists of two Hidden layers with the Input & Output layers. Linear transfer function used to the input layer & sigmoid transfer function applied to hidden layer as well as output layer. Following section describe the mathematical background of this network & learning algorithm.
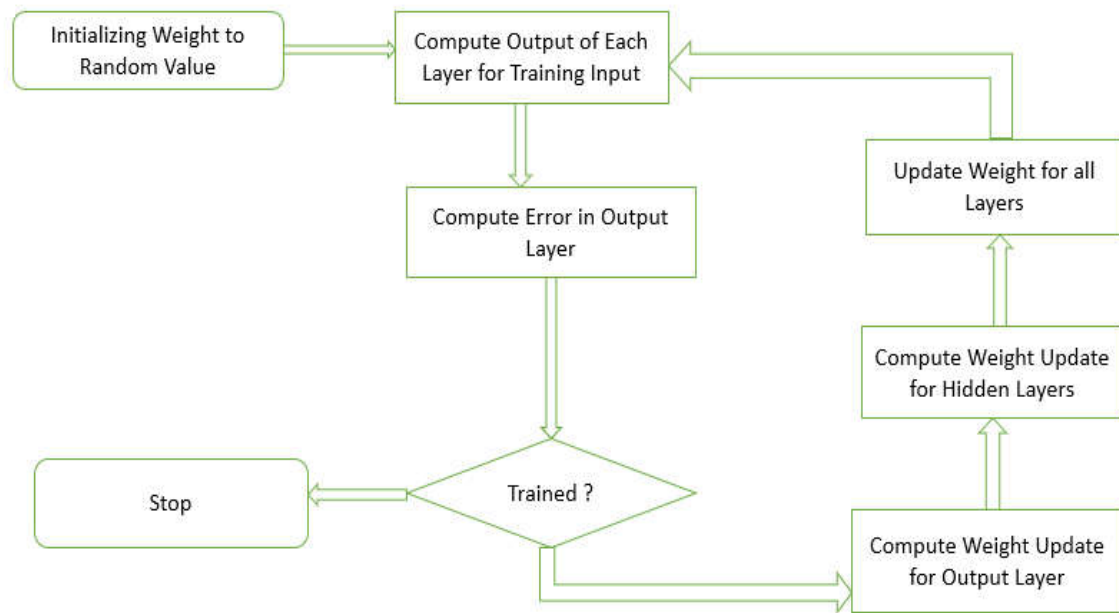
Fig. 7. Backpropagation Training Algorithm

Input Layer Function, y = mx

Hidden Layers & Output Layers Transfer Function, $y = 1 / (1 + \exp(-\lambda x))$

The neural network is built on MATLAB computing environment& it process data in real time & the program produces the output according to recognized voice command. First of all, calculated set of Cepstral co-efficient for that spoken word is fed to input layer of Artificial Neural Network & calculate the output. During the training process, set of voice command samples is input with the relevant output network to be produced. For each iteration network, actual network output is calculated & compared with the specified output & error is back propagated to minimize the error in next iteration.

Tuning process

The Neural Network we used is capable of identifying two voice commands. In order to identify more commands, we can use more separate networks. While using this kind of process we need to set output threshold levels according to the way that it had trained. For this, we would give known voice patterns to the already obtained neural network & set the threshold levels accordingly.

Working

This project Voice Controlled Robotic Vehicle helps to control robot through voice commands received via android application. The integration of control unit with Bluetooth device is done to capture and read the voice commands. The robotic vehicle then operates as per the command received via android application. For these 8051 microcontrollers is integrated in the system which makes it possible to operate the vehicle via android application. The controlling device may be any android based Smartphone/tab etc. having an android OS. The android controlling system provides a good interactive GUI that makes it easy for the user to control the vehicle. The transmitter uses an android application required for transmitting the data. The receiver end reads these commands and interprets them into controlling the robotic vehicle.

The android device sends commands to move the vehicle in forward, backward, right and left directions. After receiving the commands, the microcontroller then operates the motors. The communication between android device and receiver is sent as serial communication data. The microcontroller program is designed to move the motor through a motor driver IC as per the commands sent by android device.

Hardware specifications

8051 series Microcontroller, Motor Driver IC, DC motors, Robot Body, Crystal, Voltage Regulator, Bluetooth Device, Speech recognition module, Push Buttons, RF module, Encoder, Decoder, Resistors, Capacitors, Diodes, Batteries

Software specifications

Keil µVision IDE, MC Programming Language: Embedded C, Android App

Algorithm

- Start
- Establish Bluetooth connectivity between Android Application and the Bluetooth module on the robot
- Check whether the device is connected
- If connected, give the pre-defined instructions/ commands to the micro-phone of the mobile handset
- The voice commands should be trained to the EasyVR module
- Then the stored voice commands are represented in the form of binary numbers such as move forward – 001, move backward – 010 etc
- These binary values are transmitted via zigbee module which is a transceiver
- The transmitted binary values are then received by another zigbee module which is present on the receiver side
- Microcontroller will take those binary values and performs action (servo motors) according to the binary values

- If failed to connect at step 3 than again go to step 2
- Stop

Applications

- This line following robot can also be modified to a walking robot in any surface by using only one motor with homemade fixtures.
- Similarly, the same walking robot can be attached with special magnetic material as per instructions to climb a wall of any vertical metallic surface say a refrigerator body.
- This voice controlled robot vehicle has important application in industries where user can control the robotic vehicle via voice commands.

Conclusion and scope for future work

This project completely reforms the robotic vehicle & gives it a new dimension. It can easily recognize the voice commands and runs smoothly. Further enhancement in project can be used for Home security and military purposes where the commands can be given to robot without risk by increasing the range and by installing cameras.

- This research work has been narrowed down to short range Bluetooth module. Using a long-range modules and other connectivity devices will result in connectivity with the robot for long distances.
- Power Optimization such sleep and wakeup schedules can be incorporated.
- Image processing can be implemented in the robot to detect the colour and the objects.
- A thermal camera can be installed to sense the heat emitted by bodies useful in military purposes to detect enemies on the lines.
- Automatic Targeting System can be implemented in the robot for tracking the target.

RESULTS

In this paper, a method of voice (word) speech recognition system is proposed to control the voice operated robotic vehicle & thus the proposed technique/criteria would be more efficient in real time operation used in controlling.

The hardware used for making of the vehicle is also specified and the android app working which will control the robotic vehicle. The communication channel which would carry the signal, will be of 2 types that is "cellular connection" for word recognition internet connection & for clear connection to the robot instead of using an analogy connection, we have used a high digital connection that is via Bluetooth.

REFERENCES

Deng, Li, Douglas O'Shaughnessy, 2003. Speech processing: a dynamic and optimization-oriented approach. Marcel Dekker. pp. 41–48. ISBN 0-8247-4040-8.

Genin, Y. 1987. A survey of the eigenstructure properties of finite Hermitian Toeplitz matrices, Integral Equations and Operator Theory, 10, pp. 621-639.

MCS 51, MCS 251 and MCS 96 Microcontroller Product Lines, the Intel 186, Intel386 and Intel486 Processors Product Lines, and the i960 32 Bit RISC Processor, PCN 106013-01, Product Discontinuance, Reason for Revision: Add Key Milestone information and revise description of change" (PDF). Intel. 2006-05-02

Mohri, M. 2002. "Edit-Distance of Weighted Automata: General Definitions and Algorithms" (PDF). *International Journal of Foundations of Computer Science,* 14 (6): 957–982. doi:10.1142/S0129054103002114.

Sutskever, L., Vinyals, O., Le, Q. 2014. "Sequence to Sequence Learning with Neural Networks" (PDF). Proc. NIPS.

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., Lang, K. J. 1989. "Phoneme recognition using time-delay neural networks". IEEE Transactions on Acoustics, Speech and Signal Processing, 37 (3): 328–339. doi:10.1109/29.21701

Zen, Heiga; Sak, Hasim, 2015. "Unidirectional Long Short-Term Memory Recurrent Neural Network with Recurrent Output Layer for Low-Latency Speech Synthesis" (PDF). Google.com. ICASSP. pp. 4470–4474.

https://dsp.stackexchange.com/questions/8245/how-to-decide-the-number-of-cepstral-coefficients

https://www.engineersgarage.com/contribution/android-based-voice-controlled-robot

https://www.lairdtech.com/product-categories/connectivity-solutions/bluetooth-modules

https://www.mathworks.com/products/matlab.html

****