Asian Journal of Computer Science and Technology
ISSN: 2249-0701 (P) Vol.1 No.1, 2012, pp.115-117

DOI: https://doi.org/10.51983/ajcst-2012.1.1.1692

A Speech Recognizing Circuit for Home Service Robot

Merin Antony[1] and K. Mohanram[2]

[1]*PSNA College, Dindigul , Tamil Nadu, India*
[2]*MBC College, Peermade, Kerala, India*
E-mail: merink54@gmail.com

Abstract **- Due to the increased customer demand for various applications in various fields, service robots are becoming very popular. Human Robot Interaction (HRI) is very essential for the good performance of the home service robot. A robot can be controlled by many methods and here the robot is controlled by voice commands which are given orally. The robot must be able to perform various tasks such as speech source localization, source separation and classification etc. The robot must be capable to identify the source accurately. The speech recognizer and a speaker classifier are used to improve the performance of the robot. Reverberation deteriorates the quality and intelligibility of speech. This leads to the poor performance of classification systems. The noise effects can also degrade the performance of the classification systems. The channel may also have ISI. So channel must be equalized to nullify these effects. Here a new blind deconvolutive neural network is used for speech classification in the robot. Using this new method the robot can identify the commands given to it more accurately than by the existing methods.**

Keywords **- Reverberation, Blind equalization, Neural network, Robotics, Speech classification, HRI**

I. Introduction

Now a day's robots are gaining much importance in most of the fields. The demand for robot is increasing in many fields like health care, industry, house hold, security, home network etc. A robot can be controlled by various type of methods like by using computers, joysticks etc. But these all method had many complex method of controlling. Here a robot that can be controlled by voice command is considered. Such robot must be capable to recognize the commands given by its master accurately. The robot must be able to perform speech source localization, source separation and classification of the separated sources. However the robot must be capable to identify the source accurately. A speech recognizer and a speaker classifier are used for the improved performance of the robot. But the quality of the speech is degraded due to room reverberation. This leads to the poor performance of classification systems. Room reverberation parameters depend on the location of the speaker, the microphone the room geometry etc. The noise effects also degrade the performance of the speech classification and the speaker classification systems. The robot has to move according to the needs and thus the reverberation is constantly changing due to the relative movement of the speaker and the robot. This can affect the classification accuracy. A multi layer perceptron classifier can be used as classifier. Due to these movements the reverberation parameters are constantly changing with time. This results in non stationary signals and it adds further complications to the classification system. Moreover in most of the cases information-bearing signals transmitted between remote locations often travel through a signal-altering physical channel like coaxial cable, fibre optic, or twisted-pair cable in case of wired communications and the atmosphere or ocean in case of wireless communications .Each of these physical channels may cause distortion of the transmitted signal, including echoes and frequency-selective filtering of the transmitted signal. The signals are also affected by noise from various sources. The channels will also exhibit inter symbol interference. Linear channel equalization can be used to counter the effects of this linear channel distortion. The equalizer attempts to extract the transmitted symbol sequence by counteracting the effects of ISI. Thus it improves the probability of correct symbol detection. Histogram equalization (HE) is a very suitable tool for compensating the linear and non-linear distortions introduced into the speech signal by the environment .This can also be used in combination with noise reduction techniques.

It is common that the channel characteristics is unknown at the beginning or may vary with time. Hence the equalizer is a structure which must be adaptive in nature. There are two types of adaptive equalizers - trained equalizer and blind equalizer. Classical equalization techniques employ a time-slot during which a training signal, known in advance by the receiver, is transmitted. The receiver adapts the equalizer so that its output closely matches the known reference training signal. This may sacrifices the valuable channel capacity. Hence adaptation without the use of training, i.e., blind adaptation, is preferred. In most of the cases thus a teo stage structure consisting of a equalizer and a classifier is used so that the ill effects are nullified. A neural network constant modulus algorithm based equalizer is used prior to the multi layer perceptron classifier.

II. Literature Review

Wail Gueaieb has implemented an Intelligent Mobile Robot Navigation Technique Using RFID Technology which is based on processing some analog features of an RFID signal. Here the main consideration is the ability of speech classification in the service robot and many methods have been proposed for efficient speech classification. Maximum mutual information neural network for a hybrid connectionist-hidden Markov model speech recognition system was implemented by Gerhard. The neural network is used as a vector quantizer and is trained with a new learning algorithm. The error is reduced by this method. Mark D. Skowronski and John G. Harris have implemented a Noise-Robust Automatic Speech Recognition Using a Predictive Echo State Network. The predictive ESN classifier is formed by combining the ESN with a state machine framework. Edmondo Trentin and Marco Gori has implemented speech recognition system by combining neural networks and hidden markov models. They defined automatic speech recognisation as the classification of correct words from a given set or class. There are various other conventional speech classifiers like the linear prediction coefficients decision tree, the linear prediction cepstral coefficient decision tree, the Mel-cepstral coefficient decision tree etc. These are feature based classification systems. But the classification accuracy of these systems is less due to the effect of noise and other disturbances. In most of these classification systems, there will be a set of features or class of features as reference. The input is compared with these references and after comparison it is decided to which class or set the given input belongs.

III. Description

The existing methods suffer from lack of classification accuracy. This is because the input will be contaminated by noise in most of the cases and hence the noise and other nonlinearities must be equalised. Hence the new method in which equalisation is done prior to classification is used. Thus a two stage structure consisting of a equalizer and a classifier is used. The blind equalizer is based on the neural network constant modulus algorithm. This part consists of M layers. W1 is the weight and β2 is the constant modulus. The aim is to minimize the cost function. The learning algorithm for the NNCMA is similar to standard CMA. μ0 is the learning rate. The biases have to be updated properly.

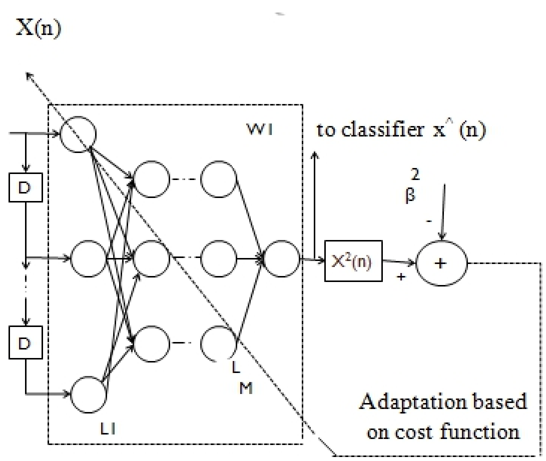

Fig.1 NNCMA based equalizer

After proper adaptation the minimum cost function is obtained. The output of the equalizer is given to the classifier. The equalizer is acting as the pre-processing stage. An MLP classifier is used. w2 is the weight. μ1 is the learning rate. The aim is to minimize the error function. The output of the layers is compared with the true class. Y (n) is the output and d (n) is the

desired output. The output compared with the desired output gives the error signal e (n).

$$e(n) = (1/2)(d(n) - y(n))^2$$

The weights are adjusted so that the error function is minimized.

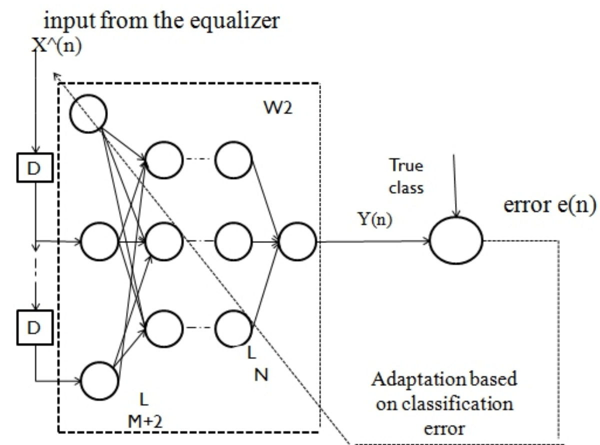

Fig.2 MLP based classifier

The BDNN network consists of N layers. The BDNN is divided into two parts consisting of M layered CMA based equalizer. The remaining layers are for the classifier. This speech recogniser consisting of two stages is used in the voice controlled robot. So the robot can efficiently classify the commands given by the master. For security purpose the system is trained in such a way that the robot will respond only to the master and not to any other person.

IV. Implementation Details

The speech recognising circuit is interfaced with the service robot. The speech recognising circuit is a completely assembled and easy to use programmable speech recognition circuit. We can train the words that the circuit want to recognise. The circuit has many applications such as Speech controlled appliances and toys, Speech assisted computer games, Voice recognition security etc. The main parts of the system are microcontroller, driver circuit, motor, microphone, HM2007, and the robot model. This robot have flash type micro controller it can be programmed with embedded c program. Then micro controller is interfaced to voice control device (HM 2007). Then microphone can be connected to HM 2007. The driver circuit is connected to controller and motor. It can be fixed at the robot model. The controller is also programmed with the voice command for controlling the robot then user can be give the command trough the microphone. The voice control device (HM2007) can process the that command and also given to the controller it can be match the command and give the control signal to the driver circuit it can be control the motor then robot model can be activated. Then user can be given various type of command to the controller then robot can process that command function.

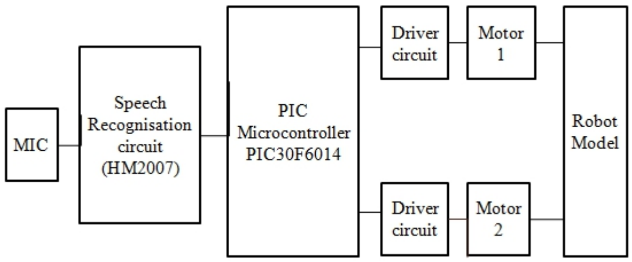

Fig.3 Block Diagram of the implemented circuit

Speech recognition is divided into two broad categories: speaker dependent and speaker independent. Speaker dependent systems are trained by the individual who will be using the system. These systems are capable of achieving a high command count and better than 95% accuracy for word recognition. The drawback to this approach is that the system only responds accurately only to the individual who trained the system. This is the most common approach employed in software for personal computers. This method is more secure. Speaker independent is a system trained to respond to a word regardless of who speaks. Therefore the system must respond to a large variety of speech patterns. The command word count is usually lower than the speaker dependent however high accuracy can still be maintain within processing limits. Industrial applications more often require speaker independent voice recognition systems. About 80 million people in India is suffering with various disabilities. Among them about18.4 million have locomotive disability, while some others suffers with various physical disabilities and some are visually impaired. These people are unable to operate their domestic appliances. Moreover, the existing voice activated home automation systems costly. Hence the idea of developing a low cost voice activated home automation system emerged, which would help people with mobility impairment. This system can control not only a robot but also domestic appliances with voice commands given in any appropriate language. The implementation of this system consists of two parts Electronic part and the Programming part. The Electronic part consists of the voice recognition IC HM2007, keypad, seven segment displays etc and other required circuits. This can be used as both voice dependent and voice independent. But for security purposes this system responds to the commands of the user only. HM2007 is trained for a specific number of commands. These commands are stored in SRAM. The Programming part consists of PIC16F877A microcontroller. 8 bit data from seven segment display is fed as input to the microcontroller. Programming is done to activate specific pins for various commands. The chip can take a maximum of 40 commands and thus can control 20 appliances using a single HM2007 IC. Output pins of microcontroller are connected to driver circuits which are in turn connected to relays. These relays control the operation of electrical appliances. This system is applied to a home service robot. The performance of the robot with and without the channel equalisation is compared. It can be seen that there is significant improvement in the classification accuracy.

V. Conclusion

Not even the robots but also other appliances like toys can be contolled by the voice command. For this a speech recognising circuit is essential and there are various methods of recognisation. But most of these methods are inefficient due to the presence of noise and other non linearities. Hence to reduce those effects an equalizer and a classifier is used for the recognisation and the classification accuracy in this case is higher than the other methods.

References

[1] D. Roviras, F. Castanie, F. Langlet,and H. Abdulkader,(2001) "UMTS channel equalization using neural networks: natural gradient algorithm," in *Proc. IJCNN,* vol. 1, pp. 699-702.

[2] D. Zhi and Y. G. Li, Blind Equalization and Identification. NewYork: Marcel Dekker, 2001.

[3] E. Trentin and M. Gori,(Nov. 2003)"Robust combination of neural networks and hidden Markov models for speech recognition," *IEEE Trans. Neural Netw.,* vol. 14, no. 6, pp. 1519-1531.

[4] F. Adnani, R. S. Hobson, P. D. Polur, Y. Jun, and Z. Ruobing, (Oct. 2001) "Isolated speech recognition using artificial neural networks," in Proc. 23rd Annu. *Int. Conf. IEEE Eng. Med. Biol. Soc. ,* vol. 2, pp. 1731-1734.

[5] F. Michaud, H. G. Okuno, J.-M. Valin, J. Rouat, K. Nakadai, and S. Yamamoto,(Aug. 2007.) "Robust recognition of simultaneous speech by a mobile robot," *IEEE Trans. Robot.,* vol. 23, no. 4, pp. 742-752.

[6] G. Rigoll,(Jan. 1994) "Maximum mutual information neural networks for hybrid connectionist-HMM speech recognition systems," *IEEE Trans. Speech Audio Process.,* vol. 2, pt. 2, no. 1, pp. 175-184.

[7] G. Yang, J. E. Meng, and Y. Song,(Jul. 2001)"A high performance neural-networksbased speech recognition system," in *Proc. Int. Joint Conf. Neural Netw., ,* vol. 2, pp. 1527-1531.

[8] H. G. Okuno ,H. Tsujino, J.-M. Valin, K. Komatani, K. Nakadai, M. Nakano, S. Yamamoto, and T. Ogata, (Oct. 9-15, 2006) , "Real-time robot audition system that recognizes simultaneous speech in the real world," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., ,* pp. 5333-5338.

[9] H. Shahhosseini, M.Reza Bonyadi, and S. Mostafa Rahimi Azghadi,(2007) "Gender classification based on feed forward back propagation neural network," in Artificial Intelligence and Innovations 2007: From Theory to Applications. New York: Springer-Verlag, pp. 299-304.

[10] I.Valganon, J. A. Salse, J. Palacin and X. Clua,(Oct. 2004.) "Building a mobile robot for a floor-cleaning operation in domestic environments," *IEEE Trans. Instrum. Meas.,* vol. 53, no. 5, pp. 1418-1424

[11] J. G. Harris,and M. D. Skowronski (Jul. 2007) "Noise-robust a speech recognition using a predictive echo state network," IEEE Trans. Audio, Speech, Lang. Process., vol. 15, no. 5, pp. 1724-1730.

[12] M. S. Miah and W. Gueaieb,(Sep. 2008) "An intelligent mobile robot navigation technique using RFID technology," *IEEE Trans. Instrum. Meas.,* vol. 57, no. 9, pp. 1908-1917.

[13] Jianyong Sun and Ata Kab´an (Aug 2010) "A Fast Algorithm for Robust Mixtures in the Presence of Measurement Errors", *IEEE Trans. Neural Networks.,* vol. 21, no. 8, pp. 1206-1220.