

# Integrated Speaker and Speech Recognition for Wheel Chair Movement using Artificial Intelligence

Gurpreet Kaur

Research Scholar, I.K Gujral Punjab Technical University, Kapurthala-144603, India

Assistant Professor, University Institute of Engineering & Technology, Panjab University, Chandigarh-160025, India

E-mail: regs4gurpreet@yahoo.co.in

Mohit Srivastava

Professor, Chandigarh Engineering College, Landran, Mohali-140307, India

E-mail: mohitsrivastava.78@gmail.com

Amod Kumar

Scientist, Central Scientific Instruments Organisation, Chandigarh-160030, India

E-mail: csioamod@yahoo.com

**Keywords:** speaker recognition, speech recognition, mel frequency cepstral coefficients, artificial bee colony algorithm, feed forward back propagation neural network

**Received:** November 10, 2017

*Abstract: A speech signal is a result of the constrictions of vocal tract and different sounds can be generated by different vocal tract constrictions. A speech signal carries two things i.e. speaker's identity and meaning. For a specific applications of speaker and speech recognition like voice operated wheel chair, both the speaker and speech is to be recognized for movement of wheel chair. Automation in wheelchair is, today's requirement as the numbers of people are increasing with disabilities like injuries in spine, amputation, impairments in hands etc. They need assistance for moving their wheel chair. Voice operated wheel chair is one of the solution. The intention of this study is to use a speaker and speech dependent system to control the wheelchair and minimize the risk of unwanted accident. We have proposed a system in which speaker (patient) as well as speech (commands) is recognized based upon the acoustic features like Mel Frequency Cepstral Coefficients (MFCC). Optimization of the features using Artificial Bee Algorithm (ABC) is done to gain good accuracy with artificial intelligence technique as a classifier. We have tested our system on standard dataset (TIDIGITS) and our own prepared dataset. Also, validation of proposed work is done by generating control signal to actuate the wheel chair in real time scenario.*

*Povzetek: Predstavljen je nadzor invalidskega vozička s pomočjo govora in metod umetne inteligence.*

## 1 Introduction

In the recent years, speaker and speech recognition has become a major domain of research because of various applications in real world from home to health care services. Speaker and speech recognition is jointly used in voice operated wheel chair. Wheel chair should move only when specific person (speaker recognition) should give commands (speech recognition). To guarantee good accuracy and less learning time, feature extraction process is very important for any recognition system. We have done detailed analysis on various feature extraction methods [1-2]. MFCC are the most used features for the speaker and speech recognition systems. Various optimization algorithms like Genetic algorithms with convolutional neural networks have been used for various applications like human action recognition [3-4]. Weights can be optimized using Genetic Algorithm (GA) framework. Dervis Karaboga proposed ABC algorithm for optimizing numerical problems[5]. This algorithm can be used in various fields like data mining, designing

of filters, speech/speaker recognition etc. [6]. Initially DNN based system are used on phone recognition task. After that DNNs are used at large scale on big vocabulary continuous speech [7-10]. When DNN based systems are compared with other systems like dynamic time warping (DTW), Hidden Markov Models (HMM), Gaussian mixture models (GMM) based systems, then recognition accuracy is more for DNN based systems [11]. There is fast learning in deep neural network by parameterization of weight matrix by using periodic functions. In this way training time is reduced and classification accuracy improves. Different configurations of layers in neural network results in different results in terms of accuracy by considering complexity and memory requirements of the system [12-13]. Feed forward back propagation network (FFBPN) and recurrent neural network (RNN) are most used networks type [14-15]. This paper presents the integrated speaker and speech recognition system with optimized

MFCC features using ABC algorithm and FFBN is used for classification. The main aim of proposed work is to combined the features of speech according to the speaker and create a speech and speaker dependent system to recognize the command which is given by the speaker to operate the wheelchair. If the wheelchair is operated only on commands without recognizing speaker then anybody can provide the wrong commands. So we need to combine the speech with speaker for the movement of wheel chair without any error. Operating the wheelchair according to the command of speaker is a big task and the feature extraction technique plays an important role. If the extracted features of speech signal are not unique then the accuracy of recognition system is not acceptable and the chances of error will increase. So the selection of a feature extraction technique is major issue in the speech recognition system. MFCC is a better option to find out the features from the speech signals. But the chances of unwanted features are still there, so optimization is needed and we have used the artificial bee colony algorithm with a novel objective function. So in the proposed work, an efficient speech and speaker recognition system is presented using the artificial intelligence technique along with the ABC optimization algorithm.

## 2 Related work

There are many existing speech and speaker recognition system using different types of techniques.

Many researchers have developed smart wheel chairs with voice commands using commercially available recognition systems. Simpson et al. [16] developed a wheelchair with joystick as well as voice controlled module. They have used commercially available Verbox speech commander recognition system with nine commands. With the development of many algorithms in artificial intelligence domain, researchers have moved to this field. Pacnik et al. [17] proposed a voice operated intelligent wheel chair (VOIC) using LPC cepstral analysis with neural network. They have analyzed that recognition of speech is possible with neural network, giving 4% error. Another author Jabardi [18] developed a wheelchair based upon artificial neural network with 86% of accuracy for known speakers and 76% of accuracy for unknown speakers. Speech recognition field has gain more advancement in terms of recognition rate, environmental conditions, speaker variability etc. with the development of deep neural networks[19-24].

From the above survey we have decided to present an integrated speaker and speech recognition system using ABC algorithm with artificial intelligence technique for the movement of wheel chair.

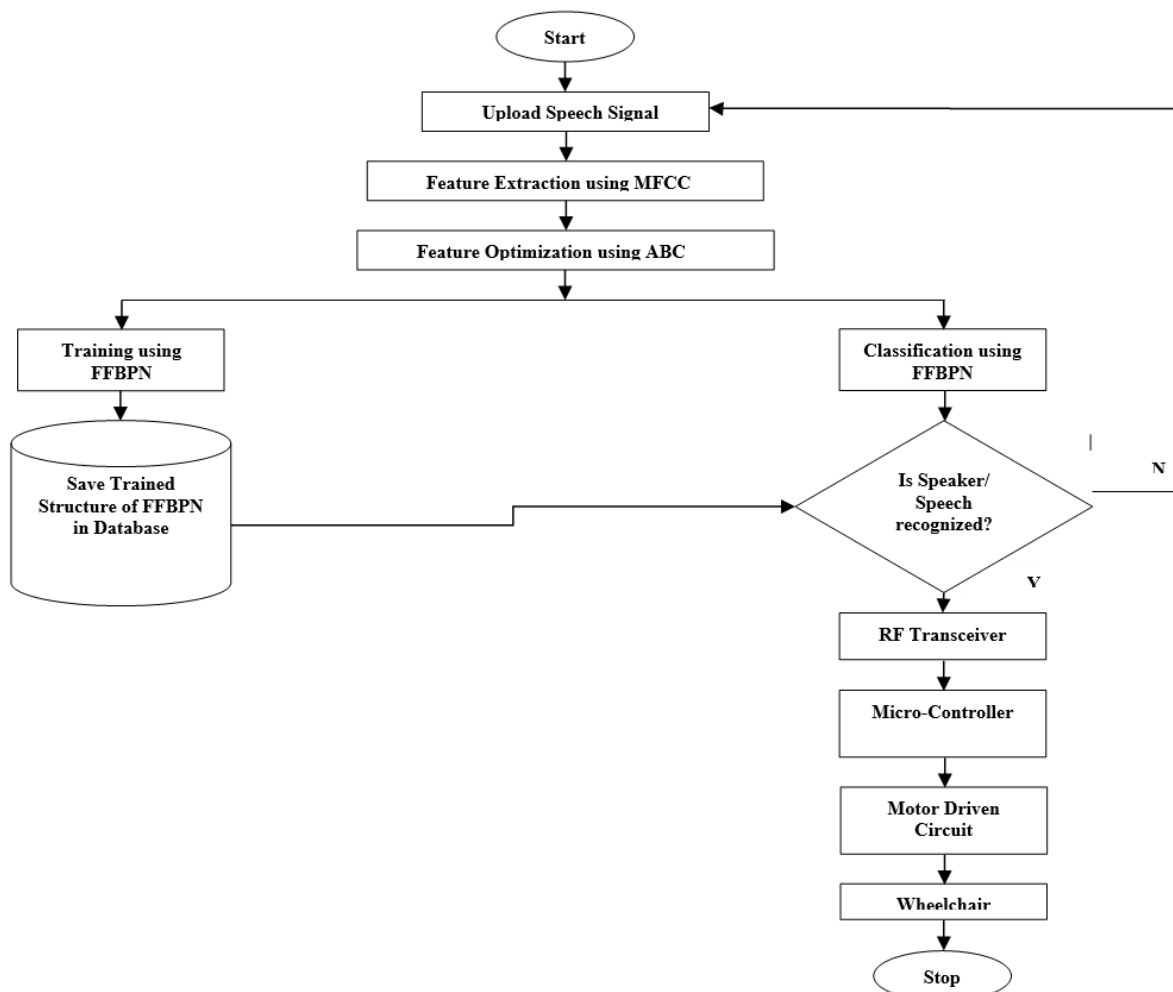


Figure 1: Integrated Speaker and Speech Recognition System for movement of wheel chair.

### 3 Proposed system

We have proposed an integrated speaker and speech recognition system as shown in figure 1.

Above figure represent the flow diagram of integrated speaker and speech recognition system based on the artificial intelligence concept. In the speaker and speech recognition system, firstly features are extracted from the speech signal. These features should be robust to noise and efficient enough so that classification can discriminate between the speakers and words. We have used MFCC features and to gain in accuracy, optimization of these features is done by ABC algorithm. Then classification is done with feed forward back propagation neural network. The FFBN output is what is recognized i.e. who is speaking and what he or she is speaking. The command word is given through the RF transceiver to the microcontroller. The MCU (ATMEGA 8) interprets the commands received and accordingly motor is controlled through driver circuit (L293D) to move the wheel chair.

#### 3.1 Feature extraction using MFCC algorithm

In this section, we have described the MFCC algorithm which is used to find out the feature set from the speech signal with respect to the speaker. The algorithm of MFCC is given below.

Firstly Initialized parameters

Tw = 25 (analysis frame duration (ms))

Ts = 10 (analysis frame shift (ms))

Alpha = 0.97 (pre emphasis coefficient)

R = [300 3700] (frequency range to consider)

M = 20 (number of filter bank channels)

C = 13 (number of cepstral coefficients)

L = 22 (cepstral sine lifter parameter)

Hamming = ((N)(0.54 - 0.46 \* cos(2 \* pi \* [0:N-1] ./ (N-1))))

MFCC<sub>features{i}</sub>

$$= \sum_{i=1}^n \text{MFCC}(\text{Signal}, fs, Tw, Ts, \text{Alpha}, \text{Hamming}, R, M, C, L)$$

Where Signal is the speech data which is uploaded by user and MFCC\_features are the extracted feature set from the uploaded speech data.

#### 3.2 Optimization with ABC algorithm

The probability of unwanted signals are more in extracted features and due to the unwanted signal the accuracy of work is degraded. So we need to enhance the features set by removing the unwanted signal using the artificial bee colony algorithm as an optimization technique. To optimize the features set, we have defined a novel objective function and fitness function of ABC algorithm as shown in equation 1.

$$ABC_{ff} = \begin{cases} Bee_{current} & \text{if } Bee_{current} > Bee_{onlooker} \\ Bee_{onlooker} & \text{else} \end{cases} \dots (1)$$

Where ABC<sub>ff</sub> is the output of fitness function and Bee<sub>current</sub> is the total bee which is called MFCC feature and Bee<sub>onlooker</sub> is the threshold value of feature set. The steps of ABC algorithm is given in below;

Upload dataset for Training

Select Case (B, F, L, R and S)

Choose Noise Type

A: Without Noise

B: White Gaussian Noise (WGN)

C: Adaptive WGN

If user=1 (Without Noise)

Speech\_signal=load (Speech Data)

$$\text{Speech\_MFCC}_{\text{features}\{i\}} = \sum_{i=1}^n \text{mfcc}(\text{Speech\_signal})$$

Initialize ABC Algorithm

Define - Employed bee

- Onlookers bee and

- Scouts bee

Set objective function:

$$ABC_{ff} = \begin{cases} Bee_{current} & \text{if } Bee_{current} > Bee_{onlooker} \\ Bee_{onlooker} & \text{else} \end{cases}$$

Optimized\_MFCC{i}

$$= \sum_{i=1}^r \sum_{j=1}^c \text{ABC}(\text{MFCC}_{\text{features}}, fs, ft)$$

where fs is selected value and ft is threshold value

else if user=2 (WGN Noise)

Speech\_signal\_WGN=load (Speech Data)

Speech\_WGN\_MFCC<sub>features{i}</sub>

$$= \sum_{i=1}^n \text{mfcc}(\text{Speech\_signal\_WGN})$$

Optimized\_WGN\_MFCC{i}

$$= \sum_{i=1}^r \sum_{j=1}^c \text{ABC}(\text{MFCC}_{\text{features}}, fs, ft)$$

where fs is selected value and ft is threshold value

else if user=3 (AWGN Noise)

Speech\_signal\_AWGN=load (Speech Data)

Speech\_AWGN\_MFCC<sub>features{i}</sub>

$$= \sum_{i=1}^n \text{mfcc}(\text{Speech\_signal\_AWGN})$$

Optimized\_AWGN\_MFCC{i}

$$= \sum_{i=1}^r \sum_{j=1}^c \text{ABC}(\text{MFCC}_{\text{features}}, fs, ft)$$

Where fs is selected value which is called current bee and ft is threshold value which is called onlooker bee. The best optimized is called scout bee which is the optimal feature from the MFCC feature sets.

end

Where speech\_signal is the speech data which is uploaded by users and Optimized\_AWGN\_MFCC is the

optimized features set which is used in the training of proposed system as an input of FFBPN algorithm.

### 3.3 FFBPN algorithm

A Feed forward back propagation neural network (FFBPN) is an authoritative machine learning technique from the field of deep learning. FFBPNs are trained using large collections of optimized features set. From these large collections, FFBPNs can learn prosperous feature representations for a wide range of features. The used algorithm of FFBPN is given as;

```

Load Optimized_MFCC_Data
Trainingdata = Optimized_MFCC_Data
Initialize FFBPN
Generate group of data = group
Set iteration = 1000
for i = 1: iteration
Weight = Optimized_MFCC_Data(i)
Hidden_Layer = [25, 25, 25] (tansig)
Net_algo = trainrp
Generate Net structure of FFBPN (net)
Net_Output = train (net, Trainingdata, group)
end

```

We have saved the Net\_Output as a training data and simulated with test data and appropriate results are calculated with feed forward back propagation neural network. The Net\_Output depends on the training data of network and it contains the categories of data which is used in the classification stage of proposed work. In the training phase we have considered the 25 neurons in each hidden layers with tan sigmoid transfer function. This is used as a carrier of signal from one layer to another layer of FFBPN. Each layer of FFBPN produces a response, or activation, to an input feature. However, there are only a few layers within a FFBPN that are suitable for feature training. Here we have set the 1000 iteration for the training of input data based on the performance criteria of FFBPN. In the each iteration FFBPN adjust the weight of input feature and create a structure of output according to the defined group at the time of initialization of network.

## 4 Proposed system

We have proposed an integrated speaker and speech recognition system as shown in figure 1.

Above figure represent the flow diagram of integrated speaker and speech recognition system based on the artificial intelligence concept. In the speaker and speech recognition system, firstly features are extracted from the speech signal. These features should be robust to noise and efficient enough so that classification can discriminate between the speakers and words. We have used MFCC features and to gain in accuracy, optimization of these features is done by ABC algorithm. Then classification is done with feed forward back propagation neural network. The FFBPN output is what is recognized i.e. who is speaking and what he or she is

speaking. The command word is given through the RF transceiver to the microcontroller. The MCU (ATMEGA 8) interprets the commands received and accordingly motor is controlled through driver circuit (L293D) to move the wheel chair.

### 4.1 Feature extraction using MFCC algorithm

In this section, we have described the MFCC algorithm which is used to find out the feature set from the speech signal with respect to the speaker. The algorithm of MFCC is given below.

Firstly Initialized parameters

Tw = 25 (analysis frame duration (ms))

Ts = 10 (analysis frame shift (ms))

Alpha = 0.97 (pre emphasis coefficient)

R = [300 3700] (frequency range to consider)

M = 20 (number of filter bank channels)

C = 13 (number of cepstral coefficients)

L = 22 (cepstral sine lifter parameter)

Hamming = ((N)(0.54 - 0.46 \* cos(2 \* pi \* [0:N-1] ./ (N-1))))

MFCC<sub>features{i}</sub>

$$= \sum_{i=1}^n \text{MFCC}(\text{Signal}, fs, Tw, Ts, \text{Alpha}, \text{Hamming}, R, M, C, L)$$

Where Signal is the speech data which is uploaded by user and MFCC\_features are the extracted feature set from the uploaded speech data.

### 4.2 Optimization with ABC algorithm

The probability of unwanted signals are more in extracted features and due to the unwanted signal the accuracy of work is degraded. So we need to enhance the features set by removing the unwanted signal using the artificial bee colony algorithm as an optimization technique. To optimize the features set, we have defined a novel objective function and fitness function of ABC algorithm as shown in equation 1.

$$ABC_{ff} = \begin{cases} Bee_{current} & \text{if } Bee_{current} > Bee_{onlooker} \\ Bee_{onlooker} & \text{else} \end{cases} \dots (1)$$

Where  $ABC_{ff}$  is the output of fitness function and  $Bee_{current}$  is the total bee which is called MFCC feature and  $Bee_{onlooker}$  is the threshold value of feature set. The steps of ABC algorithm is given in below;

Upload dataset for Training

Select Case (B, F, L, R and S)

Choose Noise Type

A: Without Noise

B: White Gaussian Noise (WGN)

C: Adaptive WGN

If user=1 (Without Noise)

Speech\_signal=load (Speech Data)

$$\text{Speech\_MFCC}_{\text{features}\{i\}} = \sum_{i=1}^n \text{mfcc}(\text{Speech\_signal})$$

Initialize ABC Algorithm

Define - Employed bee

- Onlookers bee and

- Scouts bee

Set objective function:

$$ABC_{ff} = \begin{cases} Bee_{current} & \text{if } Bee_{current} > Bee_{onlooker} \\ Bee_{onlooker} & \text{else} \end{cases}$$

Optimized\_MFCC{i}

$$= \sum_{i=1}^r \sum_{j=1}^c \text{ABC}(\text{MFCC}_{\text{features}}, fs, ft)$$

where fs is selected value and ft is threshold value

else if user=2 (WGN Noise)

Speech\_signal\_WGN=load (Speech Data)

Speech\_WGN\_MFCC<sub>features</sub>{i}

$$= \sum_{i=1}^n \text{mfcc}(\text{Speech\_signal\_WGN})$$

Optimized\_WGN\_MFCC{i}

$$= \sum_{i=1}^r \sum_{j=1}^c \text{ABC}(\text{MFCC}_{\text{features}}, fs, ft)$$

where fs is selected value and ft is threshold value

else if user=3 (AWGN Noise)

Speech\_signal\_AWGN=load (Speech Data)

Speech\_AWGN\_MFCC<sub>features</sub>{i}

$$= \sum_{i=1}^n \text{mfcc}(\text{Speech\_signal\_AWGN})$$

Optimized\_AWGN\_MFCC{i}

$$= \sum_{i=1}^r \sum_{j=1}^c \text{ABC}(\text{MFCC}_{\text{features}}, fs, ft)$$

Where fs is selected value which is called current bee and ft is threshold value which is called onlooker bee. The best optimized is called scout bee which is the optimal feature from the MFCC feature sets.

end

Where speech\_signal is the speech data which is uploaded by users and Optimized\_AWGN\_MFCC is the optimized features set which is used in the training of proposed system as an input of FFBN algorithm.

### 4.3 FFBN algorithm

A Feed forward back propagation neural network (FFBN) is an authoritative machine learning technique from the field of deep learning. FFBNs are trained using large collections of optimized features set. From these large collections, FFBNs can learn prosperous

feature representations for a wide range of features. The used algorithm of FFBN is given as;

Load Optimized\_MFCC\_Data

Trainingdata = Optimized\_MFCC\_Data

Initialize FFBN

Generate group of data = group

Set iteration = 1000

for i = 1: iteration

}Weight = Optimized\_MFCC\_Data(i)

}Hidden\_Layer = [25, 25, 25] (tansig)

Net\_algo = trainrp

Generate Net structure of FFBN (net)

Net\_Output = train (net, Trainingdata, group)

end

We have saved the Net\_Output as a training data and simulated with test data and appropriate results are calculated with feed forward back propagation neural network. The Net\_Output depends on the training data of network and it contains the categories of data which is used in the classification stage of proposed work. In the training phase we have considered the 25 neurons in each hidden layers with tan sigmoid transfer function. This is used as a carrier of signal from one layer to another layer of FFBN. Each layer of FFBN produces a response, or activation, to an input feature. However, there are only a few layers within a FFBN that are suitable for feature training. Here we have set the 1000 iteration for the training of input data based on the performance criteria of FFBN. In the each iteration FFBN adjust the weight of input feature and create a structure of output according to the defined group at the time of initialization of network.

## 5 Experiments and results

In this section, the simulation results and analysis of proposed work is described. The speech is acquired by sound recorder with the help of headphone at 16 KHz frequency at room environment in mono format. In our case, database is prepared for four speakers of age 27-34, two females (F1, F2) and two males (M1, M2). The words recorded are 'Forward, Backward, Left, Right, Stop'. Each word is recorded 80 times and hence 400 words are recorded for each speaker creating a database of 1600 words. It is much more difficult to recognize speech in presence of noise. Proposed work is tested on various types of noises like White Gaussian Noise (WGN), Adaptive White Gaussian Noise (AWGN) etc. We have tested our system on TIDIGITS database and our own created database. For the all types of signal, we have extracted the features and then optimized them to enhance the features set. After the optimization, we have trained the features with FFBN.

In the training phase, we have used the set of 25 neurons in each hidden layers with tan sigmoid transfer function to train the input feature data. After the training, we have tested the simulation with a test speech signal and process is repeated for testing phase.

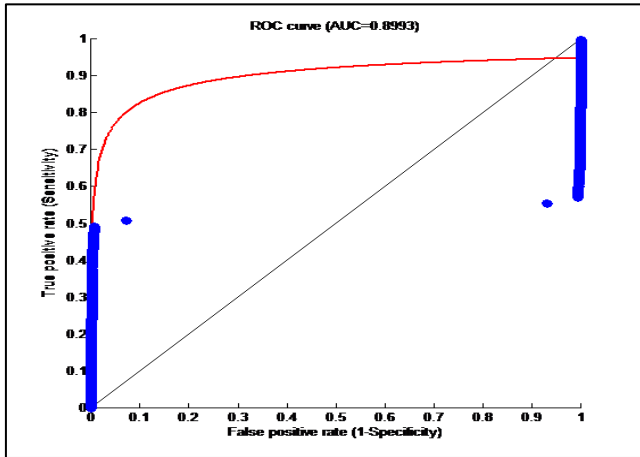


Figure 2: ROC Curve for Proposed Work.

Figure 2 shows the receiver operating characteristics (ROC) curve of proposed speech and speaker recognition system. It is a graphical method for comparing two empirical distributions where x-axis denotes the false positive rate and y-axis denotes the true positive rate. On the basis of ROC curve we have calculated the probability of recognition accuracy using the area under curve (AUC), which varies from 0 to 1. Form the figure 2, the AUC value is 0.8993 which indicates the training to system is good, therefore better classification rate. Table 1 shows the recognition accuracy for four persons including 2 men and 2 women. We have compiled the results for four persons using TIDIGITS dataset.

Speech Signal (Words)	Man 1	Man 2	Woman 1	Woman 2
One	96.88	97.11	97.35	96.20
Two	95.38	94.27	95.00	96.24
Three	99.42	98.30	97.91	99.35
Four	98.32	99.14	98.10	98.65
Five	99.16	97.88	99.02	99.35
Six	96.88	97.11	97.35	96.26
Seven	95.38	94.27	95.01	96.28
Eight	99.42	98.30	97.91	99.35
Nine	98.32	99.14	98.13	98.65
Zero	99.16	97.88	99.02	99.35

Table 1: Accuracy of integrated speaker and speech recognition for different isolated words (Clean Environment).

Figure 3 shows the accuracy of integrated speaker and speech recognition work for the digit database in clean environment. The average accuracy is more than 97% in clean environment.

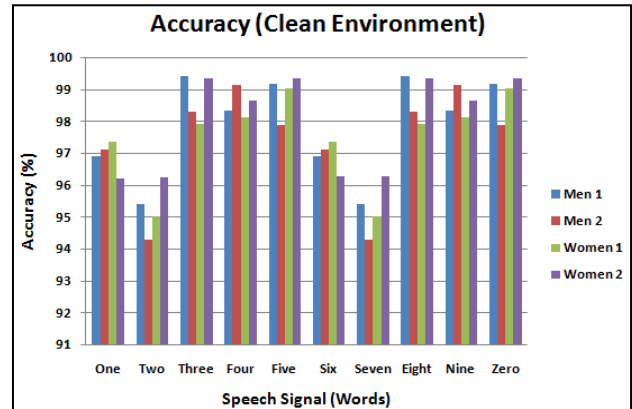


Figure 3: Accuracy of integrated speaker and speech recognition in clean environment.

Speech Signal (Words)	Man 1	Man 2	Woman 1	Woman 2
One	88.76	90.93	91.17	90.08
Two	89.26	88.09	88.82	90.03
Three	93.26	92.12	91.79	93.17
Four	92.19	92.96	91.92	92.47
Five	92.91	91.74	92.85	93.84
Six	90.75	90.93	91.19	90.56
Seven	89.21	88.09	88.87	90.07
Eight	93.24	92.15	91.72	93.17
Nine	92.14	92.99	91.92	92.45
Zero	92.98	91.72	92.84	93.17

Table 2: Accuracy of integrated speaker and speech recognition for different isolated words (Noisy Environment).

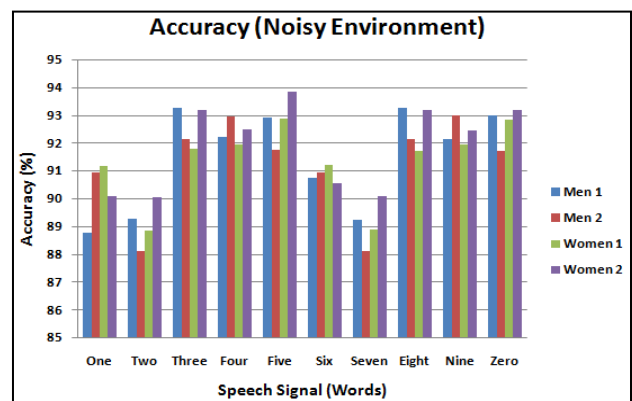


Figure 4: Accuracy of integrated speaker and speech recognition in noisy environment.

Figure 4 and table 2 shows the achieved accuracy in noisy environment. Speech signal is corrupted by adding White Gaussian Noise and then accuracy is measured. The average accuracy is more than 91% in noisy environment.

Table 3 shows the comparison of two methods: One with MFCC only and another is with MFCC and ABC

algorithm on our own created algorithm using FFBNP as a classifier.

No. of Iterations	Proposed work using MFCC	Proposed work using MFCC with ABC Algorithm
1	92.85	97.64
2	94.06	98.94
3	90.68	97.47
4	89.84	95.94
5	91.64	97.69
<b>Average (%)</b>	<b>91.82</b>	<b>97.53</b>

Table 3: Accuracy of integrated speaker and speech recognition for own created database.

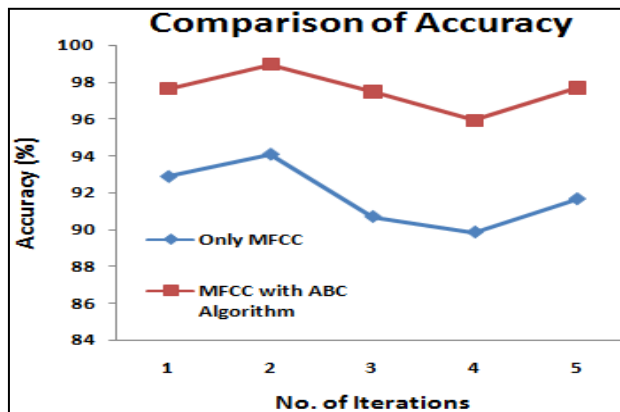


Figure 5: Comparison of accuracy.

In the figure 5 the comparison of accuracy for proposed work using optimization and without optimization is given. The accuracy is better for optimization case. So in integrated speaker and speech recognition, optimization is a better tool to create a unique feature set.

Further, we have tested our system in real time scenario for the movement of wheelchair. The command from MATLAB software is received using RF data modem. It works on 2.4 GHz frequency with adjustable baud rates of 9600 /115200 for direct interfacing with MCU. The MCU (ATMEGA 8) interprets the commands received and accordingly motor is controlled through driver circuit (L293D). Programming for MCU (ATMEGA 8) is done on ARDUINO Compiler. We have achieved average 87.4% accuracy for five isolated words in different environments like lab, canteen, office etc.

## 6 Conclusion

In proposed work, we have presented that speaker as well as speech recognition system with MFCC, ABC and FFBNP is helpful in achieving more accuracy. To be specific, we have found that optimization and feature extraction are very important as well as difficult steps in any pattern recognition system. In proposed work, we have extracted more useful feature set from speech signal using MFCC technique, feature optimization using ABC optimization algorithm and for the training and

classification of data, FFBNP is used. The experimental results analyzed that proposed method using MFCC with ABC algorithm provides good results with 97% of accuracy and it is 6% more than without using optimization technique. In real time scenario, average accuracy achieved is 87.4%.

## 7 References

- [1] Cutajar M., Micallef J., Casha O., Grech I., and Gatt E. Comparative study of automatic speech recognition techniques, *IET Signal Processing*, 7(1): 25–46, 2013. <http://dx.doi.org/10.1049/iet-spr.2012.0151>
- [2] Kaur, G., Srivastava, M., and Kumar, A. Analysis of feature extraction methods for speaker dependent speech recognition, *International journal of engineering and technology innovation*, 7(2):78–88, 2017.
- [3] Ijjina E. P. and Mohan C. K. Human action recognition using genetic algorithms and convolutional neural networks, *Pattern recognition*, 59: 199–212, 2016. <http://dx.doi.org/10.1016/j.patcog.2016.01.012>
- [4] Ijjina E. P. and Mohan C. K. Hybrid deep neural network model for human action recognition, *Applied soft computing*, 46: 936–952, 2015. <https://doi.org/10.1016/j.asoc.2015.08.025>
- [5] Karaboga D., and Akay B. A comparative study of Artificial Bee Colony algorithm, *Applied mathematics and computation*, 214(1): 108–132, 2009. <https://doi.org/10.1016/j.amc.2009.03.090>
- [6] Bolaji A. L., Khader A. T., Al-Betar M. A., and Awadallah M. A. Artificial bee colony algorithm, its variants and applications: A survey, *Journal of theoretical and applied information technology*, 47(2): 434–459, 2013. <https://doi.org/10.1504/IJAIP.2013.054681>
- [7] Chandra B., and Sharma R. K. Fast learning in deep neural networks, *Neuro-computing*, 171: 1205–1215, 2016. <https://doi.org/10.1016/j.neucom.2015.07.093>
- [8] Li K., Wu X., and Meng H. Intonation classification for L2 English speech using multi-distribution deep neural networks, *computer speech & language*, 43: 18–33, 2017. <https://doi.org/10.1016/j.csl.2016.11.006>
- [9] Richardson F., Member S., Reynolds D., and Dehak N. Deep neural network approaches to speaker and language recognition, *IEEE signal processing letters*, 22(10): 1671–1675, 2015. <https://doi.org/10.1109/LSP.2015.2420092>
- [10] Dahl G. E., Yu D., Deng L., and Acero A., Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition, 20(1): 30–42, 2012. <https://doi.org/10.1109/TASL.2011.2134090>
- [11] Solera-Urena R. and Garcia-Moral A. I. Real-time robust automatic speech recognition using compact

- support vector machines, *Audio speech and language processing*, 20(4): 1347–1361, 2012.  
<https://doi.org/10.1109/TASL.2011.2178597>
- [12] Mohamad D., and Salleh S. Malay isolated speech recognition using neural network : a work in finding number of hidden nodes and learning parameters, *International Arab journal of information technology*, 8(4): 364–371, 2011.
- [13] Desai Vijayendra A., and Thakar V. K. Neural network based Gujarati speech recognition for dataset collected by in-ear microphone, *Procedia computer science*, 93: 668–675, 2016.  
<https://doi.org/10.1016/j.procs.2016.07.259>
- [14] Abdalla O. A., Zakaria M. N., Sulaiman S., and Ahmad W. F. W. A comparison of feed-forward back-propagation and radial basis artificial neural networks: A Monte Carlo study,” *Proceedings 2010 International Symposium on Information Technology*, 2: 994–998, 2010.  
<https://doi.org/10.1109/ITSIM.2010.5561599>
- [15] Chen X., Liu X., Wang Y., Gales M. J. F., and Woodland P. C. Efficient training and evaluation of recurrent neural network language models for automatic speech recognition, *IEEE/ACM Transactions on audio speech and language processing*, 24(11): 2146–2157, 2016.  
<https://doi.org/10.1109/TASLP.2016.2598304>
- [16] Simpson, R.C. et al. NavChair: An assistive wheelchair navigation system with automatic adaptation, *Assistive technology and artificial intelligence*, 1458: 235–255, 1998
- [17] Pacnik, G., Benkic, K. and Brecko, B. Voice operated intelligent wheelchair - VOIC," *IEEE international symposium on industrial electronics*, 1221–1226, 2005.  
<https://doi.org/10.1109/ISIE.2005.1529099>
- [18] Jabardi, M.H., "Voice controlled smart electric-powered wheelchair based on artificial neural network," *International journal of advanced research in computer science*, 8(5): 31–37, 2017.  
<https://doi.org/10.26483/ijarcs.v8i5.3650>
- [19] Siniscalchi S. M., Svendsen T., and Lee C.-H. An artificial neural network approach to automatic speech processing, *Neurocomputing*, 140: 326–338, 2014.  
<https://doi.org/10.1016/j.neucom.2014.03.005>
- [20] Hossain A., Rahman M., Prodhan U. K., and Khan F. Implementation of back-propagation neural network for isolated Bangla speech recognition, *International journal of information sciences and techniques*, 3(4): 1–9, 2013.
- [21] Mansour A. H., Zen G., Salh A., Hayder H., and Alabdeen Z. Voice recognition using back propagation algorithm in neural networks,” *International journal of computer trends and technology*, 23(3): 132–139, 2015.
- [22] Qian Y., Tan T., and Yu D. Neural network based multi-factor aware joint training for robust speech recognition, *IEEE/ACM transactions on audio speech and language processing*, 24(12): 2231–2240, 2016.  
<https://doi.org/10.1109/TASLP.2016.2598308>
- [23] Dede G. and Sazlı M. H. Speech recognition with artificial neural networks, *Proceedings of the annual conference of the international speech communication association, INTERSPEECH*, 20(3), 763–768, 2015.  
<https://doi.org/10.1016/j.dsp.2009.10.004>
- [24] Shahamiri S. R. and Binti Salim S. S. Real-time frequency-based noise-robust automatic speech recognition using multi-nets artificial neural networks: A multi-views multi-learners approach, *Neurocomputing*, 129(5), 1053–1063, 2014.  
<https://doi.org/10.1016/j.neucom.2013.09.040>