Applied Mathematics & Information Sciences
*An International Journal*

http://dx.doi.org/10.18576/amis/110331

# Automatic Gender Identification in Speech Recognition by Genetic Algorithm

*T. Jayasankar*[1,∗], *K. Vinothkumar*[2] *and Arputha Vijayaselvi*[3]

[1] Department of ECE, Anna University, BIT Campus, Tiruchirappllai, Tamilnadu, India
[2] Department of ECE, J.J College of Engineering Engineering and Technology, Tiruchirappllai,Tamilnadu, India
[3] Department of ECE,Kings College of Engineering Pudukkottai, Tamilnadu, India

Received: 2 Mar. 2017, Revised: 13 Apr. 2017, Accepted: 25 Apr. 2017
Published online: 1 May 2017

**Abstract:** Automatic gender classification is a system to classify the gender using speech signal developed for speech encoding, analysis, synthesis and gender identification. Generally generation of gender recognition system can be broadly in two levels namely front-end system and back-end system. The function of the front-end system is to represent by a set of vectors called feature such as pitch, short time energy (STE), energy entropy (EE), zero crossing rate (ZCR). The back-end system is also referred to classifier, and it includes to develop a gender model to recognize the gender from speech signal. In our existing system uses fuzzy logic and neural network approach does not produced the exact required result for gender classification due to complexity of training network. To overcome this problem various evolutionary algorithms like Genetic Algorithm (GA) is applied in gender classification. This work here applies GA to select the features that are responsible for gender identification. The implementation result shows the performance of the proposed method in gender identification.

**Keywords:** Speech processing, short time energy, energy entropy, zero crossing rate, classifier, genetic algorithm.

## 1 Introduction

Over the recent years, there is a steady and significant progress in natural language evolution; the spoken languages first evolved followed by its writings. The text was primary concern for communication between human and personal computer. Speech has often been adjudged as the key to the universal information access, since the speech mode is the natural way to interact, and it does not require literacy.

Speech technology can be broadly classified as (1) speech and audio coding; (2) text-to-speech synthesis (3) speech recognition (4) speaker recognition (5) speech enhancement and (6) spoken language understanding.

With the modern concern of security worldwide gender recognition has received great deal of responsiveness among of the speech researchers. Gender recognition is used to recognizing the gender from his or her voice. With the current gender recognition, can be classified as gender identification and gender verification.

In modern multimedia information retrieval systems, gender classification is applied several potential applications such as speech recognition, speaker diarization, smart human computer interaction, biometrics social robots, audio or video content indexing, etc [1, 2]. Automatic gender detection also useful in some cases of a mobile healthcare system i.e., there are some pathologies, such as vocal fold cyst. Also a quickly developing environment of computerization, one of the most vital problems in the developing world is gender identification [3].

Information about gender used for normalization of speech features, to decrease of word error rate in speech recognition. In common, gender identification of a speaker is significant for increased natural and personalized dialogue systems.

A comparative investigation on speech signals to devise a gender classifier based on pitch analysis of the speech signals through auto-correlation method [4]. A power spectrum estimation approach based gender classifications have developed [5] with the recognition accuracy of 80% on average.

A support vector machine (SVM) based gender identification is applied on discriminative weight training.

∗ Corresponding author e-mail: jayasankar27681@gmail.com

A novel gender classification system has proposed based on GMM, to model the characteristics of male and female speech [6–8]. The performances of gender classification system have been evaluated on the conditions of clean speech, with gender classification accuracy of 98% and remains 95% for most noisy speech.

Identification of the top classifier of gender and age classification when speech signals were processed has been made by experimentally comparing the various classifiers [9]. The approach for automatic identification of gender in a short segment of normally spoken continuous speech [10]. In [11], a new robust speech FHPD feature extraction has proposed for speaker identification under low SNR environments. A novel multimodal speech recognition system can be used independently or to be combined with any humanoid robot interaction applications proposed in [12].

In most of our existing system the feature values are calculated and given as an input to ANN and fuzzy logic [13–15] and MFCC approach [16, 17]. Gender classification by fuzzy logic and neural network provide better result but here we use two methods for classification so the complexity of the system is increased. Even though ANN has fast recognition rate and fault tolerance, it is not perfect because it's a local optimum problem. For these problems, meta-heuristic algorithms are considered as efficient tools to obtain optimal solutions such as Cuckoo search, Genetic Algorithm, Fuzzy logic etc. [18]. To overcome this issue, here we have proposed a new method for gender classification using‘ GA is adopted for obtaining optimal weight and improved gender identification performance.

This paper is organized as follows: Section 2 process of gender classification with genetic algorithm, Section 3 shows the experiments and results, and Section 4 concludes the paper.

## 2 Gender classification using genetic algorithm

The block diagram of gender identification using GA is shown in Fig. 1 In the first step, the features are extracted from the speakers' database and stored in a feature database. The database for features is now retrieved by the gender classification module to classify the gender for a new speech input, using GA for optimizing the search in the database. The result in the form of corresponding gender ID is shown as an output of the work from the database.

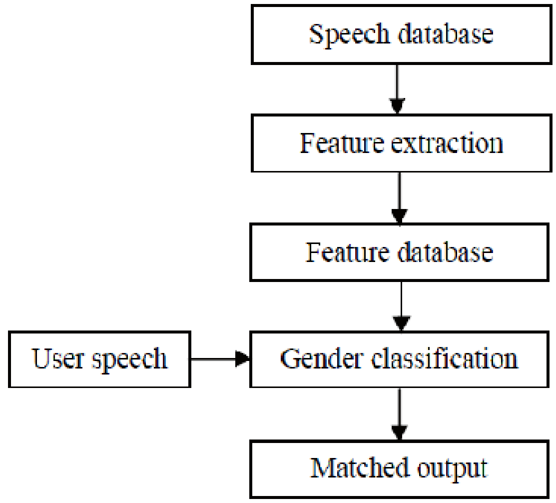


**Fig. 1:** Proposed methodology.

### 2.1 Feature Extraction for speech

Feature selection plays one of the most important parameters in gender classification. The features are used in our technique are as follows:

1.STE-Short Time Energy
2.ZCR-Zero Crossing Rate.
3.EE-Energy Entropy.

#### 2.1.1 STE

A speech signal's short time energy measurement determines voiced/ unvoiced speech. Short time energy detects changeover from voiced to unvoiced speech and vice versa. Normally, energy is high in voiced speech compared to unvoiced speech. Also, short time energy of speech signals ensures representation reflecting amplitude variations. This function is used to compute and estimate properties of the model's excitation function. The short time energy of speech signal is given by the equation1.

$$E_n = \sum_{m=-\infty}^{\infty} x^2(m)h(n-m) \tag{1}$$

where
$E(n)$—Short-time Energy,
$s(m)$—Speech signal,
$w(n)$—windowing function.

The STE is calculated from above equation, we have observed that the STE output for females is high and continuous compared to males.

#### 2.1.2 ZCR

A speech signal's short time zero crossing rate (ZCR) is used to categorize speech in voiced, unvoiced and silence. The ZCR meant as the ratio of number of time domain zero crossings occurred to the frame length. The short

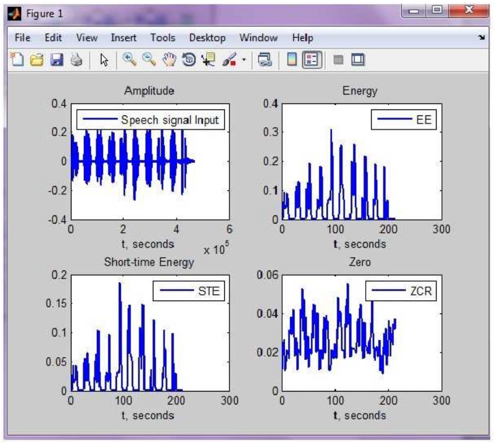


**Fig. 2:** Female speech signal and their EE, STE, ZCR.

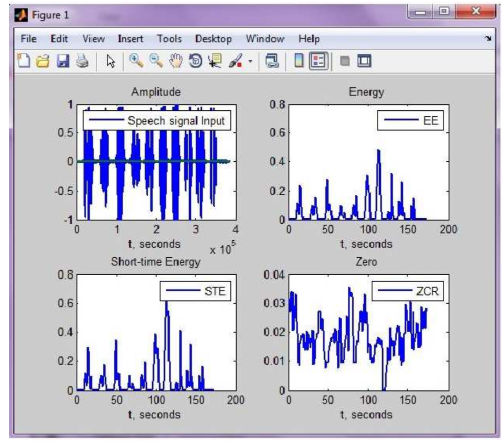


**Fig. 3:** Male speech signal and their EE, STE, ZCR.

time average energy (magnitude) crossing rate of digitally sampled speech signal is given by this (2).

$$Z_n = \sum_{m=-\infty}^{\infty} |\text{sgn}[x(m)] - \text{sgn}[(m-1)]|w(n-m)(2) \quad (2)$$

$$\text{where} \qquad \text{sgn}[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) \leq 0, \end{cases} \quad (3)$$

2.1.3 EE

A speech signal in EE represent as the sudden different changes in the energy level of a speech signal. To determine EE, the input signal is split into $k$ frames and then the normalized energy for each frame is evaluated. The EE of speech signal is given by the (4).

$$E = -\sum_{m=-\infty}^{\infty} \sigma(m) \cdot \log(\sigma) \quad (4)$$

where
$E$–Energy Entropy,
$\sigma$—Normalized Energy.

From the testing results we have perceived that the energy entropy. Fig. 2 and Fig. 3 shows the output for female and male waveform of EE, STE and ZCR. From the results EE for females is high compared to males and remains for a short period [15–17, 19]. The next procedure is to classify the percentage of gender feature present with genetic algorithm.

## *2.2 Genetic Algorithm*

In an evolutionary algorithm, are of recent interest, particularly for practical problems solving. This

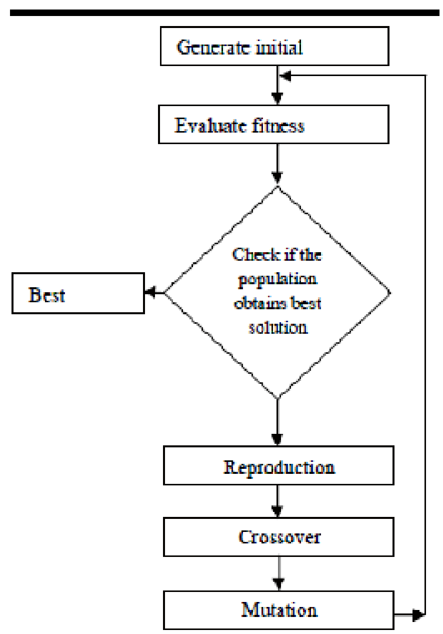


**Fig. 4:** The flow process of GA.

illustration scheme is elected by the researcher to state the set of solutions that form the search space for the algorithm. A number of individual solutions are produced to form an initial population. These steps are then repeated iteratively until a solution which satisfies a pre-defined termination criterion. Every individual is estimated using a fitness function, based upon their fitness values a number of individuals are elected to be parents [20].

The flow process of genetic algorithm is shown in Fig. 4. The generalized basic genetic algorithm is as follows: [20]:

–[start]: Genetic random population of $n$ chromosomes (suitable solutions for the problem)
–[Fitness]: Evaluate the fitness $f(x)$ of each chromosome $x$ in the population.
–[New population]: Create a new population by repeating following steps until the New population is complete.
  –[Selection]: select two parent chromosomes from a population according to their fitness (the better fitness, the bigger chance to get selected).
  –[crossover]: With a crossover probability, cross over the parents to form new offspring (children). If no crossover was performed, offspring is the exact copy of parents.
  –[Mutation]: With a mutation probability, mutate new offspring at each locus (position in chromosome)
  –[Accepting]: Place new offspring in the new population.
–[Replace]: Use new generated population for a further sum of the algorithm.
–[Test]: If the end condition is satisfied, stop, and return the best solution in current population.
–[Loop]: Go to step 2 for fitness evaluation.

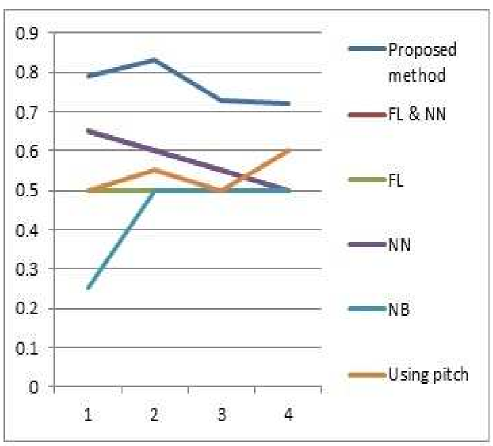


**Fig. 5:** Accuracy vs. Dataset.

**Table 1:** Performance analysis.

| | Data set | Proposed Method | FL & NN | FL | NN | NB | Using Pitch |
|---|---|---|---|---|---|---|---|
| SP | 1 | 0.87 | 0.8 | 1 | 0.8 | 0 | 0.1 |
| | 2 | 0.88 | 0.8 | 1 | 0.8 | 0 | 0.2 |
| | 3 | 0.73 | 0.8 | 1 | 0.8 | 0 | 0.1 |
| | 4 | 0.64 | 0.7 | 1 | 0.7 | 1 | 0.2 |
| *SE* | 1 | 0.77 | 0.5 | 0 | 0.5 | 0.5 | 1 |
| | 2 | 0.8 | 0.4 | 0 | 0.4 | 1 | 1 |
| | 3 | 0.75 | 0.3 | 0 | 0.3 | 1 | 1 |
| | 4 | 0.82 | 0.3 | 0 | 0.3 | 0 | 1 |
| *TP* | 1 | 10 | 5 | 0 | 5 | 5 | 10 |
| | 2 | 8 | 4 | 0 | 4 | 10 | 10 |
| | 3 | 6 | 3 | 0 | 3 | 10 | 10 |
| | 4 | 9 | 3 | 0 | 3 | 0 | 10 |
| *TN* | 1 | 9 | 8 | 10 | 8 | 0 | 1 |
| | 2 | 7 | 8 | 10 | 8 | 0 | 2 |
| | 3 | 8 | 8 | 10 | 8 | 0 | 1 |
| | 4 | 7 | 7 | 10 | 7 | 10 | 2 |
| *FP* | 1 | 2 | 2 | 0 | 2 | 10 | 9 |
| | 2 | 1 | 2 | 0 | 2 | 10 | 8 |
| | 3 | 3 | 2 | 0 | 2 | 10 | 9 |
| | 4 | 4 | 3 | 0 | 3 | 0 | 8 |
| *FN* | 1 | 3 | 5 | 10 | 5 | 5 | 0 |
| | 2 | 2 | 6 | 10 | 6 | 0 | 0 |
| | 3 | 2 | 7 | 10 | 7 | 0 | 0 |
| | 4 | 2 | 7 | 10 | 7 | 10 | 0 |
| $\alpha$ | 1 | 0.18 | 0.2 | 0 | 0.2 | 1 | 0.9 |
| | 2 | 0.13 | 0.2 | 0 | 0.2 | 1 | 0.8 |
| | 3 | 0.23 | 0.2 | 0 | 0.2 | 1 | 0.9 |
| | 4 | 0.37 | 0.3 | 0 | 0.3 | 0 | 0.8 |
| $\beta$ | 1 | 0.23 | 0.5 | 1 | 0.5 | 0.5 | 0 |
| | 2 | 0.2 | 0.6 | 1 | 0.6 | 0 | 0 |
| | 3 | 0.25 | 0.7 | 1 | 0.7 | 0 | 0 |
| | 4 | 0.18 | 0.7 | 1 | 0.7 | 1 | 0 |
| Acc | 1 | 0.79 | 0.65 | 0.5 | 0.65 | 0.25 | 0.5 |
| | 2 | 0.83 | 0.6 | 0.5 | 0.6 | 0.5 | 0.55 |
| | 3 | 0.73 | 0.55 | 0.5 | 0.55 | 0.5 | 0.5 |
| | 4 | 0.72 | 0.5 | 0.5 | 0.5 | 0.5 | 0.6 |
| Pre | 1 | 0.83 | 0.71 | 0 | 0.714 | 0.33 | 0.526 |
| | 2 | 0.89 | 0.66 | 0 | 0.667 | 0.5 | 0.55 |
| | 3 | 0.67 | 0.6 | 0 | 0.6 | 0.5 | 0.526 |
| | 4 | 0.69 | 0.5 | 0 | 0.5 | 0 | 0.55 |

## 3 Result and discussion

For developing the gender identification we have considered 80 speech signals as an input and then divided into 4 groups of dataset. The result of recommended technique i.e., Gender identification using Genetic algorithm by comparing different approach like the combination of fuzzy logic and neural network, Fuzzy Logic (FL), Neural Network (NN), Naive Bayes (NB). The proposed system is simulated in Matlab. It shows that our gender identification system is enhanced than the other techniques.

The performance of gender classification of proposed method and other techniques is given in Table 1. The performance can be evaluated by computing parameters like are $\alpha$, $\beta$, *SE*, *specificity*, *accuracy*, *LRP*, *LRN* and precision. The parameters are calculated by using the formula given below:

$$SP = \frac{TN}{FP + TN} \tag{5}$$

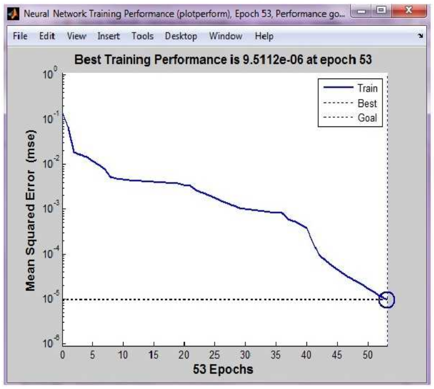


**Fig. 6:** Specificity vs. Dataset.

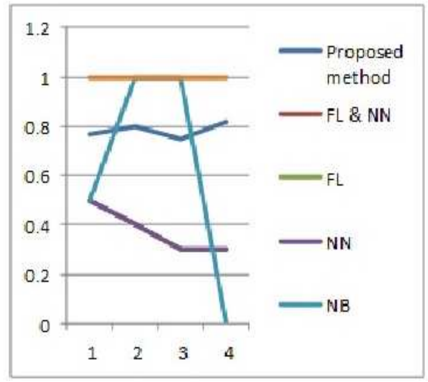


**Fig. 7:** *Sensitivity* vs. Dataset.

$$SE = \frac{TP}{TP+TN} \tag{6}$$

$$\alpha = \frac{FP}{FP+TN} \tag{7}$$

$$\beta = \frac{FN}{FN+TP} \tag{8}$$

$$LRP = \frac{Sensitivity}{1-specifity} \tag{9}$$

$$LRN = \frac{1-Sensitivity}{specifity} \tag{10}$$

$$Acc = \frac{TP+TN}{TP+FP+TN+FN} \tag{11}$$

$$Pre = \frac{TP}{FP+TP} \tag{12}$$

where $\alpha$—False positive rate,
$\beta$—False negative rate,
*TP*—True Positive,
*TN*—True Negative,
*LRP*—Likelihood Ratio Positive,
*FP*—False Positive,
*SE*—sensitivity,
*FN*—False Negative.

From the Table 1 it is observed that, the various other method of gender identification are compared with the proposed method to analyse the identification of gender and it is found the genetic approach is proved to be best with high accuracy, specificity, sensitivity and less complexity. As shown in Fig. 5 the Accuracy vs dataset graph respectively for proposed method, Fuzzy Logic (FL), Neural Network (NN), FLNN. Experiments show that proposed method (GA) method is better than other methods.

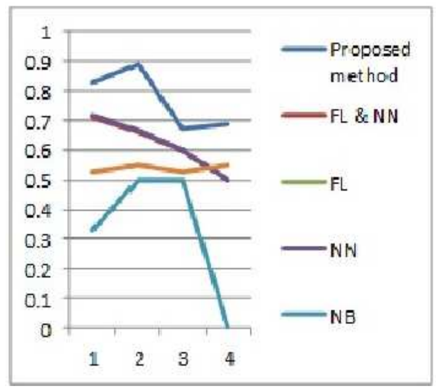


**Fig. 8:** Precision vs. Dataset.

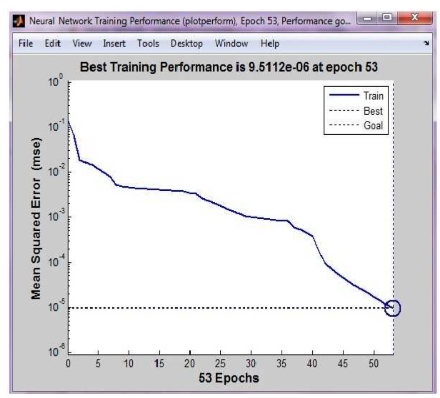


**Fig. 9:** Best training performance during neural network training.

The graph of specificity, sensitivity and precision values obtained from each dataset for proposed method as Figs. 6, 7 and 8. From the results above we can see that the GA approach is competitive with AN and the proposed systems further improve the performance of speaker identification system.

Fig. 9 shows the performance graph obtained during the training of neural network.

## 4 Conclusion

This research study proposes a novel Genetic algorithm based Gender identification designed and implemented for speech database. The given input speech is analysed and identified gender of speaker. It is found genetic algorithm shown results with its training and testing capabilities. In this study gender classification is attained by various features such as EE, STE and ZCR. The mean values are computed for three features by using a training dataset and percentage of identification of gender are computed using genetic approaches. This method is implemented in the MATLAB for testing and gender identification is used for obtaining efficient output.

## References

[1] H. Harb and L. Chen, Voice-based gender identification in multimedia applications, Journal of Intelligent Information Systems, 24(2), 179-198 (2005).
[2] D. Ververidis and C. Kotropoulos. Automatic speech classification to five emotional states based on gender information. In Proc. XII European Signal Processing Conf., volume 1, pages 341–344. Vienna, Austria,(2004).
[3] Musaed Alhussein, Zulfiqar Ali, Muhammad Imran,and Wadood Abdul, Automatic Gender Detection Based on Characteristics of Vocal Folds for Mobile Healthcare System, Mobile Information Systems, Article ID 7805217, 1–12 (2016).
[4] Bhagyalaxmi Jena and Beda Prakash Panigrahi, Gender Classification by Pitch Analysis,Electronics & Tele-communication Dept., Silicon Institute of Technology, 1(1), 2012.
[5] Md. Sadek Ali1, Md. Shariful Isla1 and Md. AlamgirHossain, Gender recognition of speech signal, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), 2(1), 1–9 (2012).
[6] S.-I. Kang and J.-H. Chang, Discriminative weight training-based optimally weighted MFCC for gender identification, IEICE Electronics Express 6(19), 1374–1379 (2009).
[7] Y. Hu, D. Wu, and A. Nucci, Pitch-based gender identification withtwo-stage classification, Security and Communication Networks, 5(2), 211–225 (2012).
[8] S. Gaikwad, B. Gawali, and S.C. Mehrotra, Gender identification using SVM with combination of MFCC, Advances in Computational Research, 4(1), 69-73, 2012.
[9] M. Sedaaghi, A Comparative Study of Gender and Age Classification in Speech Signals, Iranian Journal of Electrical & Electronic Engineering, 5(1), 1–12 (2009).
[10] M. Sigmund, Gender Distinction using Short Segments of Speech Signal, International Journal of Computer Science and Network Security, 8(10), 159–162 (2008).
[11] L. H. Zhang, Y. Bao and Z. Yang. Robust feature based on speech harmonic structure for speaker identification, Journal of Electronics & Information Technology, 28(10), 1786-1789 (2006).
[12] J. F. Lehman. Robo fashion world: a multimodal corpus of multi-child human-computer interaction, In Proceedings of the 2014 Workshop on Understanding and Modeling Multiparty, Multimodal Interactions, 1520. ACM, (2014).
[13] K. Rakesh, S. Dutta and K. Shama, "Gender Recognition using Speech Processing Techniques in LABVIEW, International Journal of Advances in Engineering & Technology, 1(2), 51–63 (2011).
[14] J. Park, F. Diehl, M.J.F. Gales, M. Tomalin and P.C. Woodland,The efficient incorporation of MLP features into automatic speech recognition systems, Journal, ELSEVIER, Computer Speech and Language, 25, 519–534 (2011).
[15] W.Y. Huang and R.P. Lippmann, "Neural Net andTraditional Classifiers," Proc. IEEE Corif'. on Neural Information Processing Systems-Natural and Synthetic,IEEE, New York, NY (1987).
[16] Jamil Ahmad, Mustansar Fiaz, Soon-il Kwon, Maleerat Sodanil, Bay Vo, and Sung Wook Baik, Gender Identification using MFCC for Telephone Applications–A Comparative Study International Journal of Computer Science and Electronics Engineering (IJCSEE), 3(5), 351–355 (2015).
[17] Rong Phoophuangpairoj and Sukanya Phongsuphap, Two-stage Gender Identification Using Pitch Frequencies, MFCCs and HMMs IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, 2879–2884,(2015).
[18] K. S. Lee and Z. W. Geem, A new meta-heuristic algorithm for continuous engineering optimization: harmony search theory and practice, Computer Methods in Applied Mechanics and Engineering, 194(36-38),3902-3933(2005)
[19] L.R. Rabiner, and R.W. Schafer, Digital Processing of Speech Signals, Englewood Cliffs, New Jersey, Prentice Hall, 512-ISBN-13:978013.
[20] D. Goldberg, editor. Genetic Algorithms in Search, Optimization, and Machine Learning, Addison Wesley, Reading, MA, 2136037, 1989.

**T. Jayasankar** received the B.E. degree in Electronics and Communication Engineering from Bharathiyar University, Coimbatore in 2001 and M.E. degree at Madurai Kamaraj University, Madurai in 2003 and Ph.D. in Speech Processing at Anna University Chennai 2017. At present, he is an Assistant Professor in the Electronics and Communication Engineering department, University College of Engineering, Anna University, Bharathidasan Institute of Technology Campus, Tiruchirappalli, Tamilnadu, India. He is a member of IEI, ISTE. He has been a lecturer at graduate and post-graduate level and has participated in a number of International and National level conferences and workshops. He has published around 25 papers in the reputed international journals and more than 15 papers in the international and national conferences and contributed one book chapter. His main interest is currently speech synthesis, speech and image processing and wireless networks.

**K. Vinoth Kumar** working as Assistant Professor (Senior Grade) in J.J. College of Engineering and Technology, Tiruchirappalli, Tamil Nadu, India. He received the Bachelor's degree in Electronics and Communication Engineering from the Kurinji College of engineering and technology, Manapparai, Tamil Nadu, India, in 2009. He received the Master's degree in Applied Electronics from the J.J. College of engineering and technology, Tiruchirappalli, Tamil Nadu, India, in 2011. He received the Ph.D. in Karpagam University Coimbatore, Tamil Nadu, India in 2017. He is a member in Universal Association of Computer and Electronics Engineer (UACEE) and member in International Association of Engineer (IAENG). He published more than ten international journals. His research interests include wireless communication, Mobile Ad hoc networks, and Sensor Networks and Communication networks. He has published in 2 science indexed journals and 5 Scopus Indexed journals.

**J. Arputha Vijaya Selvi** received her B.E. degree in Electronics and Communication Engineering from Government college of Engineering Tirunelveli, in 1991, and M.E. degree at National Institute of Technology Tiruchirappalli, in 1997 and Ph.D. degree from National Institute of Technology Tiruchirappalli, in 2006. From 2003 to 2004, she was the principal of Idhaya Engineering College for Women, TamilNadu. In 2007, she joined as Dean in Kings College of Engineering, Thanjavur. Her special interests are Adaptive Optics, Signal processing, Simulation and control.