

Gender Classification in Speech Recognition using Fuzzy Logic and Neural Network

Kunjithapatham Meena¹, Kulumani Subramaniam², and Muthusamy Gomathy³

¹Shrimathi Indira Gandhi College, Bharathidhasan University, India

²Department of Computer Application, Shrimathi Indira Gandhi College, India

³Department of Computer Science, Shrimathi Indira Gandhi College, India

Abstract: Nowadays classification of gender is one of the most important processes in speech processing. Usually gender classification is based on considering pitch as feature. The pitch value of female is higher than the male. In most of the recent research works gender classification process is performed using the abovementioned condition. In some cases the pitch value of male is higher and also pitch of some female is low, in that case this classification does not produce the exact required result. By considering the aforementioned problem we have here proposed a new method for gender classification method which considers three features. The new method uses fuzzy logic and neural network to identify the gender of the speaker. To train fuzzy logic and neural network, training dataset is generated by using the above three features. Then mean value is calculated for the obtained result from fuzzy logic and neural network. By using this threshold value, the proposed method identifies the speaker belongs to which gender. The implementation result shows the performance of the proposed technique in gender classification.

Keywords: Gender classification, fuzzy logic, neural network, energy entropy, short time energy, zero crossing rate.

Received July 16, 2011; accepted December 30, 2011; published online August 5, 2012

1. Introduction

In modern civilized societies for communication between human speeches is one of the common methods [3]. Different ideas formed in the mind of the speaker are communicated by speech in the form of words, phrases, and sentences by applying some proper grammatical rules [4]. By considering speech as one of the outcome of passing a glottal excitation wave form through a time varying linear filter which can be used to represent speech signal, so as a speech production model that models the resonant characteristics of the vocal tract [12]. By classifying the speech with voiced, unvoiced and silence (VAS/S) an elementary acoustic segmentation of speech which is essential for speech can be considered [15]. In succession to individual sounds called phonemes this technique can almost be identical to the sounds of each letter of the alphabet which makes the composition of human speech [18].

Speech processing is the study of speech signals, and the various methods which are used to process them. In this process various applications such as speech coding, speech synthesis, speech recognition and speaker recognition technologies; speech processing is employed [25]. Among the above, speech recognition is the most important one. The main purpose of speech recognition is to convert the acoustic signal obtained from a microphone or a telephone to generate a set of words [13, 23]. In order to extract and

determine the linguistic information conveyed by a speech wave we have to employ computers or electronic circuits [6]. This process is performed for several applications such as security device, household appliances, cellular phones ATM machines and computers [14].

Gender classification is applied in many fields. For example it is applied in various applications such as speech recognition, speaker diarization, speaker indexing, annotation and retrieval of multimedia database, synthesis, smart human computer interaction biometrics social robots etc. and it is a difficult and challenging problem [10]. We can identify physiological differences such as vocal fold thickness or vocal tract length and differences in speaking style of humans as partly the reason gender based differences in human speech [20, 26]. Normally the higher formant frequencies and fundamental Frequency (FO) are higher for female speakers and the FO differences are larger than the formant frequency differences between male and female groups [25]. For male speakers various speech qualities like aggressiveness, body size, self-confidence, and assertiveness are related to low FO [8].

In most of the previous research works classification is performed by considering pitch as feature. There are also certain limitations while considering this feature. To solve the abovementioned problem, here we have proposed a new method for gender classification using three features by using fuzzy logic and neural network.

The rest of the paper is structured as follows: The related works are briefly reviewed in section 2 and the proposed technique with adequate mathematical models and illustrations are detailed in section 3. The implementation results obtained are discussed in section 4 and section 5 concludes the paper.

2. Related Works

Some of the recent research works related to speech classification is discussed as follows.

Rakesh *et al.* [16] have proposed two different models by using several speech processing techniques and algorithms, and one of their models is used to produce formant values of the voice sample and the other model to produce pitch value of the voice sample. The gender biased features and pitch value of a speaker were extracted by employing these two models. The mean of formants and pitch of all the samples of a speaker were calculated by applying a model having loops and counters which generates a mean of Formant 1 and pitch value of the speaker. The speaker are classified between Male and Female by computing euclidean distance from the mean value of Males and Females of the generated mean values of formant 1 and Pitch by using a nearest neighbor technique. Using NI Lab VIEW, the algorithm is implemented in real time.

Rao and Prasad [17], have proposed that the different time-varying glottal excitation components of speech were used for text independent gender recognition studies. The excitation information in speech was represented by using a Linear Prediction (LP) residual. They have used a Hidden Markov Models (HMMs) to capture the gender-specific information in the excitation of different voiced speech. The decrease in the error during training and identifying genders during testing phase close to 100% precision have proved that the continuous Ergodic HMM can effectively capture the gender-specific information in the excitation component of the speech. In their gender identification study, they have also calculated the size of testing data on the gender recognition performance by using gender specific features in various HMM states, and mixture components. They have also performed the gender recognition studies on Texas Instruments and Massachusetts Institute of Technology (TIMIT) database.

Devi *et al.* [1], in their study have discussed that the background noise from noisy environment for example car, bus, babble, factory, helicopter, street noise and more have reduced the performance of speech-processing systems like speech coding, speech recognition etc. Thus the classification of noise is necessary to improve the performance of the speech recognition system. The selection of excellent set of features that can efficiently separate the signals in the

feature space was an important process in the design of a signal classification system. Noise classification is crucial process in order to reduce the consequence of environmental noises on speech processing tasks. They have proposed a fuzzy ARTMAP network and modified fuzzy ARTMAP network to classify the various background noise signals. Moreover in addition to it their experimental results were compared with both back propagation networks and Radial Basis Function Network (RBFN).

Sedaaghi [19], have discussed a comparative study of gender and age classification algorithms which is used in speech signal. Experimental results are used for the Danish Emotional Speech database (DES) and English Language Speech Database for Speaker Recognition (ELSDSR). Identification of the best classifier for gender and age classification when speech signals were processed, has been made by experimentally comparing the Bayes classifier using various techniques Sequential Floating Forward Selection (SFFS) for feature selection, probabilistic Neural Networks (PNNs), Support Vector Machines (SVMs), the K-Nearest Neighbor (K-NN) and Gaussian Mixture Model (GMM), as different classifiers. They have shown that gender classification can be carried out with a precision of 95% approximately by using speech signal either from both genders or from male and female individually.

Sigmund [21], have proposed an approach for automatic identification of gender in a short segment of normally spoken continuous speech. They have studied all the vowels separately to observe which phonemes are useful for gender recognition and have also evaluated the selected Mel-Frequency Cepstral Coefficients (MFCCs) based two different simple identifiers. More than 90% of accuracy is being achieved for gender identification in short-time analysis (20 msec) by using the vowel phonemes. Particularly there was no error for vowel "a". To recognize male/female speakers with the accuracy of more than 93%, the speech duration of 500 msec is enough for text-independent analysis. Automatic assessment of speaker's gender by her/his voice has been an important aspect for achieving high-quality dialogue systems.

Mahdi and Jafer [11], have suggested a wavelet-based algorithm for voice and unvoiced classification of speech segments. The classification process involves two steps:

1. Statistical analysis of the energy-frequency distribution of the different speech signals by means of wavelet transform.
2. Evaluation of the short-time zero-crossing rate of the signal.

For each time segment of the pre-emphasized speech, they have also calculated the ratio of the average energy in the low-frequency wavelet sub bands in

comparison to that of highest-frequency wavelet sub band by using a 4-level dyadic wavelet transform, and then have compared it to a pre-determined threshold. An experimentally confirmed criterion depends upon the results of comparison process was used to obtain the classification decision.

Silovsky and Nouza [22], in his research have presented a set of methods to categorize of various audio segments in a system for automatic transcription of broadcast programs. Their task is to decide:

1. Whether the segment should be labeled as speech or as non-speech and also in the previous case.
2. Whether the talking person is one of the speakers in the database.
3. Or else, the speaker belongs to which gender.

Extending the information obtained from transcription system and also by improving the performance of the speech recognition module was done by using the result of classification. Like all other modern speaker recognition systems, their proposed method is also based on GMM. Since the number of the database speakers can be large, they have developed a method that accelerates the recognition process in a significant way.

While reviewing these recent researches which has discussed the same problem, in most of the researches pitch is considered as feature and in some other researches other statistical features are considered. For testing their methods, in some researches emotional speech is utilized and in some other, continuous/real time speech data is considered and in other researches any words speech dataset was considered.

3. Gender Classification using Fuzzy Logic and Neural Network

Gender classification plays a major role in speech processing. This technique is used to identify the gender of the speaker. There are various methods used for gender classification. But the major problem is most of these works depends on pitch value. The pitch mainly depends on the frequency of sound. Normally the pitch of female is high and for male the pitch is low. In some cases the pitch of male is higher like the female and also the pitch of female is lower like male. In this situation speech classification using pitch will not produce appropriate results. By considering this drawback here we proposed a new method for speech classification using three features namely; energy entropy, short time energy, and zero crossing rates. Initially the three feature values are computed and given as an input to the fuzzy logic and neural network individually and it gives the percentage of male and female feature as output. Then mean value is taken and using this value gender classification is done. The process takes place in proposed method which is explained briefly in the below sections. Initially, we

have discussed about the features which are used in our method.

3.1. Feature Analysis for Speech Signals

Feature selection plays one of the important roles in gender classification. The gender classification fully depends on the feature which we have selected in proposed method. The three features used in our method are as follows:

- Short Time Energy (STE).
- Zero Crossing Rate (ZCR).
- Energy Entropy (EE).

Among these three features, the most important feature is ZCR. These features are explained briefly in [2]. Now we can see the basic operation of these three features one by one.

3.1.1. STE

The STE of speech signal is said to be the sudden increase in energy signal. To compute STE, initially the signal is split into s windows and then the window function is calculated for each window. The STE is calculated using the equation given below.

$$S = \sum_{r=-\infty}^{\infty} y(r)^2 \cdot h(s-r) \quad (1)$$

By using the above equation the STE is calculated. From the testing results we have observed that the energy entropy output for males is low whereas for females it is high and continuous.

3.1.2. ZCR

The ZCR is the most important feature considered in our method. The ZCR is defined as to be the ratio of number of time domain zero crossings occurred to the frame length. The equation 3 shows the formula to calculate zero crossing rate.

$$Z = \frac{1}{2N} \sum_{i=1}^{N-1} \{sgn\{x(i)\} - sgn\{x(i-1)\}\} \quad (2)$$

where, $sgn\{x(i)\}$ stands for the sign function, i.e.

$$sgn\{x(i)\} = \begin{cases} 1; & x(i) > 0 \\ 0; & x(i) = 0 \\ -1; & x(i) < 0 \end{cases} \quad (3)$$

By using the above equation the ZCR for each signal is calculated. From the testing results we observed that the ZCR for female speech is higher than that of the male speech.

3.1.3. EE

EE in speech signal is defined as the sudden different changes in the energy level of a speech signal. To calculate EE, initially the speech signal is split into k frames and then the normalized energy for each frame

is evaluated. The formula to calculate energy entropy is given below:

$$E = -\sum_{i=0}^{k-1} \sigma^2 \cdot \log_2(\sigma^2) \tag{4}$$

where, σ^2 is the normalized energy.

By using the above equation the EE is computed. From the testing results we have observed that the energy entropy for males is low and distributed while for females it is high and remains for a short period.

The features used in our method are explained in the above sections. Next process is to identify the percentage of male and female feature present in the given speech signal using fuzzy logic and neural network.

3.2. Identifying Male and Female Feature using Fuzzy Logic

Fuzzy Logic offers several unique parameters which alternatively produces better results in many control problems [5]. Fuzzy logic here is used to calculate the percentage of various male and female features presents in the given speech signal. Generally fuzzy logic consists of three important steps. This includes fuzzification, generating fuzzy rules and defuzzification. In the fuzzification process the system data is converted in to fuzzy data. For fuzzification process triangular membership function is used. Next process after this is generating fuzzy rules. Figure 1 shows the structure of fuzzy logic used in the proposed method with 3 input variables and one output variable.

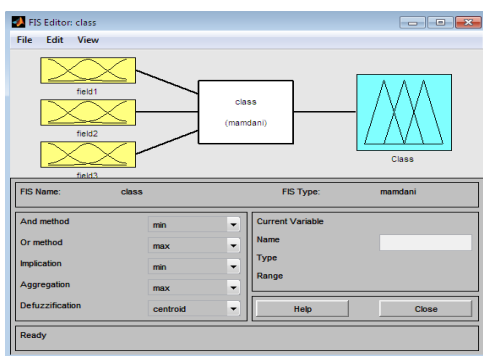


Figure 1. Structure of fuzzy used in our method.

3.2.1. Fuzzy Rules Generation

The input to our fuzzy logic is energy entropy (E), short time energy (S) and zero crossing rate (Z) and the output obtained from the fuzzy is the percentage of various male and female features which are present in the given speech signal. The input variables are fuzzified into three various sets namely; large, medium and small and the output variable is fuzzified into three sets namely; male, female/male and female. In female /male the speech signal belongs to either male or female. The fuzzy rules generated are shown in Table 1.

After generation of fuzzy rules the next step is to train fuzzy logic. The fuzzy logic is trained by using the rules shown in Table 1. To train fuzzy logic, training datasets are to be generated. The input training dataset is generated as $\{[E_{max}, E_{min}], [S_{max}, S_{min}], [Z_{max}, Z_{min}]\}$. After completion of training, the fuzzy logic obtained is ready for practical operation. In testing if we will give E, S and Z values as input to the fuzzy logic it will provide the output as the feature belong to male or female.

Table 1. Fuzzy rules.

| S. No | Fuzzy Rules for Gender Classification |
|-------|---|
| 1 | if E=high and S=low and Z=low, then Male |
| 2 | if E=high and S=low and Z=medium, then Female/Male |
| 3 | if E=high and S=low and Z=high, then Female |
| 4 | if E=high and S=medium and Z=low, then Female/Male |
| 5 | if E=high and S=medium and Z=medium, then Female |
| 6 | if E=high and S=medium and Z=high, then Female |
| 7 | if E=high and S=high and Z=low, then Female |
| 8 | if E=high and S=high and Z=medium, then Female |
| 9 | if E=high and S=high and Z=high, then Female |
| 10 | if E=medium and S=low and Z=low, then Male |
| 11 | if E=medium and S=low and Z=medium, then Male |
| 12 | if E=medium and S=low and Z=high, then Female |
| 13 | if E=medium and S=medium and Z=low, then Female/Male |
| 14 | if E=medium and S=medium and Z=medium, then Female/Male |
| 15 | if E=medium and S=medium and Z=high, then Female/Male |
| 16 | if E=medium and S=high and Z=low, then Female/Male |
| 17 | if E=medium and S=high and Z=medium, then Female/Male |
| 18 | if E=medium and S=high and Z=high, then Female |
| 19 | if E=low and S=low and Z=low, then Male |
| 20 | if E=low and S=low and Z=medium, then Male |
| 21 | if E=low and S=low and Z=high, then Male |
| 22 | if E=low and S=medium and Z=low, then Male |
| 23 | if E=low and S=medium and Z=medium, then Male |
| 24 | if E=low and S=medium and Z=high, then Female/Male |
| 25 | if E=low and S=high and Z=low, then Female/Male |
| 26 | if E=low and S=high and Z=medium, then Female/Male |
| 27 | if E=low and S=high and Z=high, then Female |

3.3. Identifying Male and Female Feature using Neural Network

The main aim of the classification ANNs is to produce an exact output based on the input parameters [7]. Neural networks are used here to calculate the percentage of female and male features present in a given speech signal. Basically neural network consists of three layers namely; input layer, hidden layer and output layer. In our method input layer has three variables, hidden layer has n variables and output layer has one variable. The input to the neural network is energy entropy, short time energy and zero crossing rate. The two stages of operation which takes place in neural network are training stage and testing stage. For training of neural network, training dataset is generated. The input training dataset is generated as $\{[E_{max}, E_{min}], [S_{max}, S_{min}], [Z_{max}, Z_{min}]\}$. Figure 2 shows the structure of neural network used in the proposed method with 3 input variables and one output variable.

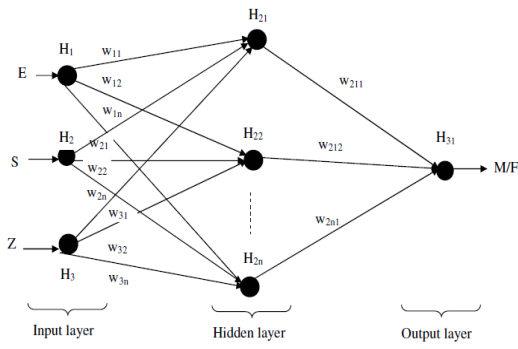


Figure 2. Structure of neural network used in proposed method.

3.3.1. Neural Network Training for Gender Classification

The steps for training the neural network are:

- *Step 1:* Initialize the input weight of each neuron.
- *Step 2:* Apply a training dataset to the network. Here E , S and Z are the input to the network and M/F is the output of the network.

$$M / F = \sum_{r=1}^n W_{2,r} y(r) \tag{5}$$

where,

$$y(r) = \frac{1}{1 + \exp(-w_{1,r} \cdot (E + S + Z))} \tag{6}$$

Equation 5, represents the activation function performed in the output and input layer respectively.

- *Step 3:* Adjust the weights of all neurons.
- *Step 4:* For each E , S and Z corresponding male and female feature is computed.
- *Step 5:* Repeat the iteration process till the output reaches its least value.

After completing the training neural network is ready for various practical applications. Next step after completion of training is testing neural network. During testing speech signal is given as input, it provides the percentage of male and female feature present in that signal.

3.4. Gender Classification for the Given Speech Signal

After completing the process of training fuzzy logic and neural network, the next process is to identify the gender of the speaker. The initial step is to compute the mean value of output obtained from fuzzy logic and neural network. The mean value is calculated using the equations given below:

$$S^{final} = \frac{S^{fuzzy} + S^{NN}}{2} \tag{7}$$

where, S^{fuzzy} is the output generated from fuzzy logic and S^{NN} is the output obtained from neural network.

After calculating these mean values, the speech signal is splitted in to male and female using a threshold value.

$$classification = \begin{cases} \text{if } S^{final} < S^{threshold} ; \text{ then female} \\ \text{if } S^{final} > S^{threshold} ; \text{ then male} \end{cases} \tag{8}$$

From the above equation we obtain the speaker belongs to which gender. The threshold used in our method is 0.5.

4. Result and Discussions

This proposed technique was implemented in MATLAB 7.10 and is tested for different speech signals from Harvard-Haskins database [24]. Here 80 speech signals are taken as an input and then splitted into four datasets. Initially the neural network and fuzzy logic is trained by using some speech signal, and testing is performed by using a set of speech signal as input to the proposed method so that it identifies the speaker gender. The result of proposed technique i.e., combination of fuzzy logic and neural network are compared with the Fuzzy Logic (FL) and Neural Network (NN), Naive Bayes (NB) and using pitch as feature. From the comparison results, it is clear that our method is better than the other methods.

The performance of the proposed method, fuzzy logic and neural network are explained separately in the below sections.

4.1. Performance Analysis of Gender Classification

The True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values are calculated from the results obtained to the proposed method, fuzzy logic and neural network. The above four values are used to compute performance parameters like false positive rate (α), false negative rate (β), sensitivity (SE), specificity (SP), Likelihood Ratio Positive (LRP), Likelihood Ratio Negative (LRN), Accuracy (Acc) and Precision (Pre) using the equations given below:

$$SP = \frac{TN}{(FP + TN)} \tag{9}$$

$$SE = \frac{TP}{(TP + FN)} \tag{10}$$

$$\alpha = \frac{FP}{(FP + TN)} \tag{11}$$

$$\beta = \frac{FN}{(TP + FN)} \tag{12}$$

$$LRP = \frac{Sensitivity}{(1 - specificity)} \tag{13}$$

$$LRN = \frac{(1 - Sensitivity)}{specificity} \tag{14}$$

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{15}$$

$$Pre = \frac{TP}{(TP+FP)} \tag{16}$$

Using these above equations the performance of proposed method, FL, NN, NB and using pitch is calculated and values obtained are displayed in the below table for all the four dataset. Here we have tested 80 speech signals which are divided into 4 datasets with 20 signals each.

Table 2. Performance analysis.

| | Data Set | Proposed Method | FL | NN | NB | Using Pitch |
|-----|----------|-----------------|-----|-------|------|-------------|
| SP | 1 | 0.8 | 1 | 0.8 | 0 | 0.1 |
| | 2 | 0.8 | 1 | 0.8 | 0 | 0.2 |
| | 3 | 0.8 | 1 | 0.8 | 0 | 0.1 |
| | 4 | 0.7 | 1 | 0.7 | 1 | 0.2 |
| SE | 1 | 0.5 | 0 | 0.5 | 0.5 | 1 |
| | 2 | 0.4 | 0 | 0.4 | 1 | 1 |
| | 3 | 0.3 | 0 | 0.3 | 1 | 1 |
| | 4 | 0.3 | 0 | 0.3 | 0 | 1 |
| TP | 1 | 5 | 0 | 5 | 5 | 10 |
| | 2 | 4 | 0 | 4 | 10 | 10 |
| | 3 | 3 | 0 | 3 | 10 | 10 |
| | 4 | 3 | 0 | 3 | 0 | 10 |
| TN | 1 | 8 | 10 | 8 | 0 | 1 |
| | 2 | 8 | 10 | 8 | 0 | 2 |
| | 3 | 8 | 10 | 8 | 0 | 1 |
| | 4 | 7 | 10 | 7 | 10 | 2 |
| FP | 1 | 2 | 0 | 2 | 10 | 9 |
| | 2 | 2 | 0 | 2 | 10 | 8 |
| | 3 | 2 | 0 | 2 | 10 | 9 |
| | 4 | 3 | 0 | 3 | 0 | 8 |
| FN | 1 | 5 | 10 | 5 | 5 | 0 |
| | 2 | 6 | 10 | 6 | 0 | 0 |
| | 3 | 7 | 10 | 7 | 0 | 0 |
| | 4 | 7 | 10 | 7 | 10 | 0 |
| α | 1 | 0.2 | 0 | 0.2 | 1 | 0.9 |
| | 2 | 0.2 | 0 | 0.2 | 1 | 0.8 |
| | 3 | 0.2 | 0 | 0.2 | 1 | 0.9 |
| | 4 | 0.3 | 0 | 0.3 | 0 | 0.8 |
| β | 1 | 0.5 | 1 | 0.5 | 0.5 | 0 |
| | 2 | 0.6 | 1 | 0.6 | 0 | 0 |
| | 3 | 0.7 | 1 | 0.7 | 0 | 0 |
| | 4 | 0.7 | 1 | 0.7 | 1 | 0 |
| LRP | 1 | 2.5 | 0 | 2.5 | 0.5 | 1.11 |
| | 2 | 2 | 0 | 2 | 1 | 1.25 |
| | 3 | 1.5 | 0 | 1.5 | 1 | 1.11 |
| | 4 | 1 | 0 | 1 | 0 | 1.25 |
| LRN | 1 | 0.625 | 1 | 0.625 | 0 | 0 |
| | 2 | 0.75 | 1 | 0.75 | 0 | 0 |
| | 3 | 0.875 | 1 | 0.875 | 0 | 0 |
| | 4 | 1 | 1 | 1 | 1 | 0 |
| Acc | 1 | 0.65 | 0.5 | 0.55 | 0.25 | 0.5 |
| | 2 | 0.6 | 0.5 | 0.6 | 0.5 | 0.55 |
| | 3 | 0.55 | 0.5 | 0.5 | 0.5 | 0.5 |
| | 4 | 0.5 | 0.5 | 0.5 | 0.5 | 0.6 |
| Pre | 1 | 0.714 | 0 | 0.714 | 0.33 | 0.526 |
| | 2 | 0.667 | 0 | 0.667 | 0.5 | 0.55 |
| | 3 | 0.6 | 0 | 0.6 | 0.5 | 0.526 |
| | 4 | 0.5 | 0 | 0.5 | 0 | 0.55 |

Table 2 shows the performance of proposed method and other methods like fuzzy logic, neural network, Naive Bayes and using pitch for various performance parameters. From the table obtained above, it is clear that the accuracy of the proposed method is very much better than the fuzzy logic and neural network, Naive Bayes and using pitch. The graph of accuracy, specificity, sensitivity and precision values obtained

from each dataset for proposed different methods, fuzzy logic, neural network, Naive Bayes and using pitch are shown below.

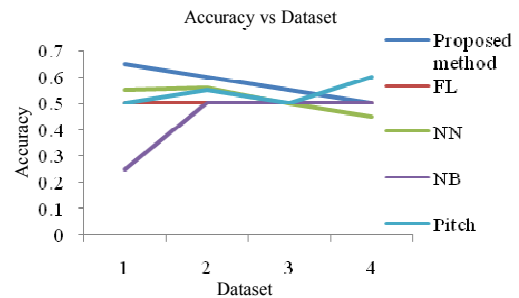


Figure 3. Comparison graph for accuracy vs dataset.

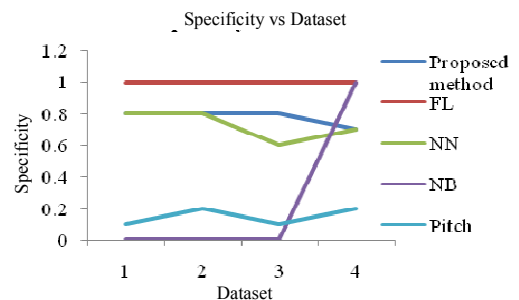


Figure 4. Comparison graph for specificity vs dataset.

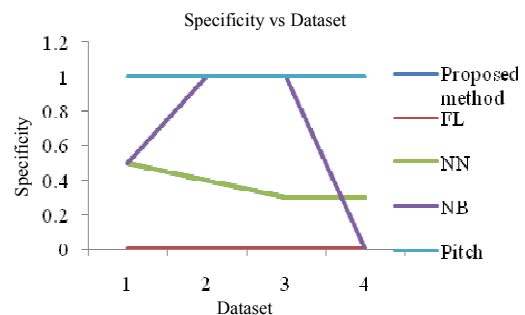


Figure 5. Comparison graph for sensitivity vs dataset.

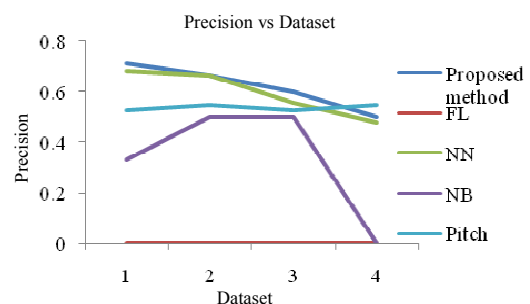


Figure 6. Comparison graph for precision vs dataset.

The above Figures 3, 4, 5, and 6 shows the accuracy, specificity, sensitivity and precision vs dataset graph respectively for proposed method, fuzzy logic, neural network, Naive Bayes and using pitch. From the above graphs it is clear that the proposed method is better than other methods.

Figure 7 shows the membership function used for training fuzzy logic and Figures 8, 9 and 10 shows the performance, regression and training graph obtained during the training of neural network respectively.

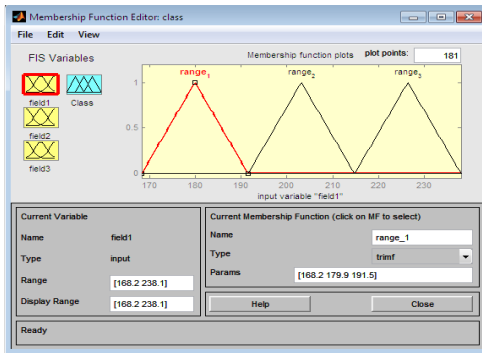


Figure 7. Fuzzy member ship function used in proposed method.

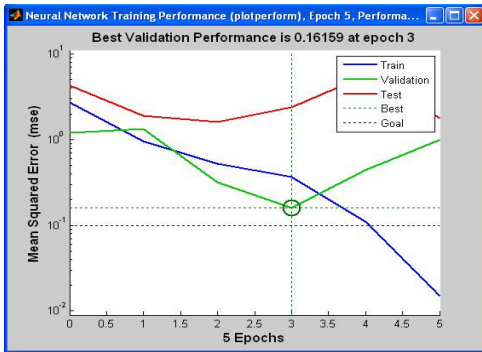


Figure 8. Performance graph obtained during neural network training.

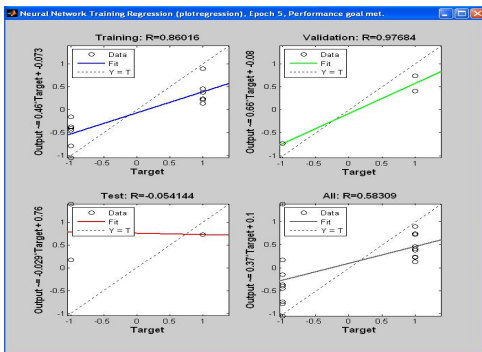


Figure 9. Regression graph obtained during neural network training.

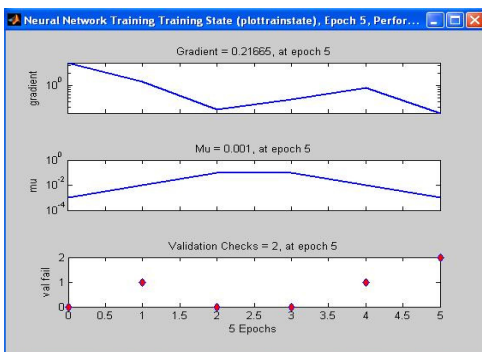


Figure 10. Training graph obtained during neural network training.

5. Conclusions

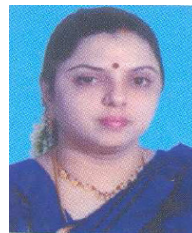
In this paper, a novel gender classification technique in speech processing using neural network and fuzzy logic was proposed. In this technique gender classification is performed by considering three different features such as energy entropy, short time

energy and zero crossing rates. Firstly mean values are calculated for three features by using training dataset and percentage of male and female features which are present in the speech signal are computed using fuzzy logic and neural network individually and then the mean value is taken to identify the gender of the speaker. This approach was implemented in the working platform of MATLAB for testing. The proposed method was tested using Harvard-Haskins database. During testing if a speech signal is given as input it will identify the gender of the speaker to which speaker belongs. The results obtained from proposed method are compared with the fuzzy and neural network, Naive Bayes and using pitch as feature. Comparison results have shown that our method is better than the other methods in gender classification.

Reference

- [1] Devi M., Kasthuri N., and Natarajan A., "Performance Comparison of Noise Classification using Intelligent Networks," *International Journal of Electronics Engineering*, vol. 2, no. 1, pp. 49-54, 2010.
- [2] Gomathy M., Meena K., and Subramaniam K., "Gender Grouping in Speech Recognition using Statistical Metrics of Pitch Strength," *European Journal of Scientific Research*, vol. 61, no. 4, pp. 524, 2011.
- [3] Gudi A., Shreedhar H., and Nagaraj H., "Signal Processing Techniques to Estimate the Speech Disability in Children," *IACSIT International Journal of Engineering and Technology*, vol. 2, no. 2, pp. 169-176, 2010.
- [4] Gudi A. and Nagaraj H., "Optimal Curve Fitting of Speech Signal for Disabled Children," *International Journal of Computer Science & Information Technology*, vol. 1, no. 2, pp. 99-107, 2009.
- [5] Haider T. and Yusuf M., "A Fuzzy Approach to Energy Optimized Routing for Wireless Sensor Network," *The International Arab Journal of Information Technology*, vol. 6, no. 2, pp. 179-188, 2009.
- [6] Haraty R. and Ariss O., "CASRA+: A Colloquial Arabic Speech Recognition Application," *American Journal of Applied Sciences*, vol. 4, no. 1, pp. 23-32, 2007.
- [7] Haraty H. and Ghaddar C., "Arabic Text Recognition," *The International Arab Journal of Information Technology*, vol. 1, no. 2, pp. 156-163, 2004.
- [8] Hasegawa Y. and Hata K., "Non-Physiological Differences between Male and Female Speech: Evidence from the Delayed F0 fall Phenomenon in Japanese," in *Proceedings of the International Conference on Spoken Language Processing*, Japan, pp. 1179-82, 1994.

- [9] Hasegawa Y. and Hata K., "The Function of F0-Peak Delay in Japanese," in *Proceedings of the 21st Annual Meeting of the Berkeley Linguistics Society*, pp. 141-151, 1995.
- [10] Kotti M. and Kotropoulos C., "Gender Classification in Two Emotional Speech Databases," in *Proceedings of the 19th International Conference on Pattern Recognition*, Tampa, pp. 1-4, 2008.
- [11] Mahdi A. and Jafer E., "Two-Feature Voiced/Unvoiced Classifier Using Wavelet Transform," *The Open Electrical and Electronic Engineering Journal*, vol. 2, no. 1874-3005, pp. 8-13, 2008.
- [12] McAulay R. and Quatieri T., "Speech Processing Based on a Sinusoidal Model," *The Lincoln Laboratory Journal*, vol. 1, no. 2, pp. 153-168, 1988.
- [13] Othman A. and Riadh M., "Speech Recognition using Scaly Neural Networks," in *Proceedings of World Academy of Science, Engineering and Technology*, vol. 38, pp. 253-258, 2008.
- [14] Patel I. and Rao S., "Speech Recognition using HMM with MFCC- an Analysis using Frequency Spectral Decomposition Technique," *Signal & Image Processing : An International Journal*, vol. 1, no. 2, pp. 101-110, 2010.
- [15] Qi Y. and Hunt B., "Voiced-Unvoiced-Silence Classifications of Speech using Hybrid Features and a Network Classifier," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 2, pp. 250-255, 1993.
- [16] Rakesh K., Dutta S., and Shama K., "Gender Recognition using Speech Processing Techniques in LABVIEW," *International Journal of Advances in Engineering & Technology*, vol. 1, no. 2, pp. 51-63, 2011.
- [17] Rao R. and Prasad A., "Glottal Excitation Feature Based Gender Identification System using Ergodic HMM," *International Journal of Computer Applications*, vol. 17, no. 3, pp. 31-36, 2011.
- [18] Rodger J. and Pendharkar P., "A Field Study of the Impact of Gender and User's Technical Experience on the Performance of Voice-Activated Medical Tracking Application," *International Journal of Human-Computer Studies*, vol. 60, no. 2, pp. 529-544, 2004.
- [19] Sedaaghi M., "A Comparative Study of Gender and Age Classification in Speech Signals," *Iranian Journal of Electrical & Electronic Engineering*, vol. 5, no. 1, pp. 1-12, 2009.
- [20] Shue Y. and Iseli M., "The Role of Voice Source Measures on Automatic Gender Classification," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, pp. 4493-4496, 2008.
- [21] Sigmund M., "Gender Distinction using Short Segments of Speech Signal," *International Journal of Computer Science and Network Security*, vol. 8, no. 10, pp. 159-162, 2008.
- [22] Silovsky J. and Nouza J., "Speech, Speaker and Speaker's Gender Identification in Automatically Processed Broadcast Stream," *Radio Engineering Journal*, vol. 15, no. 3, pp. 42-48, 2006.
- [23] Singh G., Junghare A., and Chokhani P., "Multi Utility E-Controlled Cum Voice Operated Farm Vehicle," *International Journal of Computer Applications*, vol. 1, no. 13, pp. 109-113, 2010.
- [24] Vesicle, available at: <http://vesicle.nsi.edu/users/patel/download.html>, last visited 2002.
- [25] Zanuy F., McLaughlin S., Esposito A., Hussain A., Schoentgen J., Kubin G., Kleijn W., and Maragos P., "Non-Linear Speech Processing: Overview and Applications, Control & Intelligent Systems," *ACTA Press*, vol. 30, no. 1, pp. 1-10, 2002.
- [26] Zengi Y., Wu Z., Falk T., and Chan W., "Robust GMM Based Gender Classification using Pitch and Rasta-PLP Parameters of Speech," in *Proceedings of the 5th International Conference on Machine Learning and Cybernetics*, Dalian, pp. 13-16, 2006.



Kunjithapatham Meena received her MSc, M.Phil, ME in computer science and engineering, MIE, PhD. She is the vice-chancellor of Bharathidhasan University. She is the principal and director MBA and MCA of Shrimathi Indira Gandhi

College, Trichirapalli. She has rich experience in the development of software tools for the assessment of specially abled children. Also, provides consultancy for organizing specific programmes for creating awareness/literacy about the computer and information technology among specific cross-sections of the society (Co-ordinator of the novel project IT ON WHEELS-from Lab to Land). Provides counseling for higher education, career placement and training.



Kulamani Subramaniam received his BSc, MSc degree in maths, MA (English), MEd, MSc (IT) and PhD degree in maths and computer applications from Madras, Annamalai University, Madurai and Bharathidhasan Universities, Tamil

nadu, India in the years 1966, 1969, 1982, 1977, 1983, 2009 and 2003 respectively. From 1969 to 2007 he has been an educationist for mathematics, english, educational technology and computer applications as lecturer and professor. He has headed the Department of Master of Computer Applications, Shrimathi Indira Gandhi College, Trichy-2, from 2007 to 2010.



Muthusamy Gomathy received BSc (Chemistry) degree in Holy Cross College in the year 1998. She has completed her MSIT degree in the year 2001 from Shrimathi Indira Gandhi College, Trichy, Bharathidhasan University, Then, She has completed her M.Phil degree from St.Josephs College, Trichy in the year 2002. From 2003 to 2010, She has been an Educationist for computer applications, Information Technology as Lecturer and Professor.