

# Gender Clustering and Classification Algorithms in Speech Processing: A Comprehensive Performance Analysis

M. Gomathy

Research Scholar, Department  
of CSE  
Shrimathi Indira Gandhi  
College, Trichy-2.

K. Meena

Phd, Vice Chancellor,  
Bharathidasan University,  
Principal and Director,  
Shrimathi Indira Gandhi  
College, Trichy.

K. R. Subramaniam

Phd, Professor, Department of  
Computer Application  
Shrimathi Indira Gandhi  
College, Trichy-2

## ABSTRACT

In speech processing gender clustering and classification is the most outstanding and challenging task. In both gender clustering and classification, one the most vital processes carried out is the selection of features. In speech processing, pitch is the most often used feature for gender clustering and classification. It is essential to note that compared to a female speech the pitch value of a male speech is much different. Also, in terms of frequency there is a considerable dissimilarity between the male and female speech. In some situations, either the frequency of male is almost same as female or the frequency of female is same as male. It is difficult to find out the exact gender in such conditions. This paper focus on rectifying these practical obstacles by extracting three significant features, namely, energy entropy, zero crossing rate, and short time energy. Gender clustering is performed based on these features. However, by means of Euclidean distance, Mahalanobis distance, Manhattan distance & Bhattacharyya distance methods the clustering performance is analyzed. Using fuzzy logic, neural network, hybrid neuro-fuzzy, and support vector machine the gender classification is done. A benchmark dataset and real-time dataset is used for testing to make sure the reliability of the performance. The test results show the performance of various techniques and distance algorithms for different datasets

## Keywords

Mahalanobis distance, Manhattan distance, Bhattacharyya distance, Neuro fuzzy, Support vector machine.

## 1. INTRODUCTION

In general during speech communication, initially a design about the speech is formed in the mind of speaker and by applying proper grammatical rules then it is converted in the form of words, phrases, and sentences [17]. By means of a speech production model the speech signals can be represented that views speech as the outcome of passing a glottal excitation waveform through a time-varying linear filter, which models the resonant characteristics of the vocal tract [18]. An elementary acoustic segmentation of speech given by the voiced-unvoiced-silenced (V/U/S) classification of speech signal is very important for speech analysis [19]. A sequence of individual sounds called phonemes that is approximately identical to the sounds of each letter of the alphabet makes the composition of human speech [20].

To convey their in-tensions and feelings, human beings use emotions widely [1]. The Gender Detection (GD) distinguishes between male and female gender speaker's audio segmentation,

audio event detection, and speech recognition [4]. In classifying multimedia data, audio plays a significant role, as it is easier to process when compared to video data, and also the audio data contains perceptually significant information [5]. Using different techniques the problem of automatic gender recognition in speech has been studied [2]. In speaker and speech recognition systems, gender identification is an imperative step. In both these systems, the gender identification step transforms the gender independent problem into a gender dependent one. As a result it can decrease the size and complexity of the problem [3]. To discover gender classification using learning algorithms is the main objective of this study [7]. The segmentation of the signal into acoustically homogeneous regions, ii) for classification of those segments according to background conditions, speaker gender and iii) for identifying all segments uttered by the same speaker [6].

Speech processing is the study of speech signals, where to process the speech several techniques are used. Speech processing is used in many applications like speech coding, speech synthesis, speech recognition, and speaker recognition technology [13]. Among the above, speech recognition is one of the most important system. By converting the audio signal received from a microphone or a telephone the speech recognition process produces a set of words [14]. Speech as input is more suitable for individuals challenged with a variety of physical disabilities, such as loss of sight or limitations in physical motion and motor skills [15]. The advancement of automated system the complexity for integration & recognition problem is increasing. When processing on randomly varying analog signals such as speech signals, the problem is found more complex [16].

To foster human-technology relationships gender classification is a difficult and challenging problem, on the other hand, is an interesting field for psychologists. In myriad of applications automatic gender classification take on an increasingly ubiquitous role, e.g. demographic data collection, etc. [8]. The gender based differences in human speech are partially due to physiological differences such as vocal fold thickness or vocal tract length and partially due to differences in speaking style [9]. The female speakers normally have higher formant frequencies with higher fundamental frequency (F0) and the (F0) differences are larger than the formant frequency differences between the male and female groups [11]. But for male speakers, because of the qualities like aggressiveness, body size, self-assurance, and assertiveness the F0 is lower, [10]. Energy entropy, pitch, zero crossing rate, spectral roll off, mel frequency spectral coefficients, spectrogram feature etc. are some of the commonly used audio features [12]

## 2. RELATED WORKS

Hongling Xie *et al.* [21] have proposed for using description reports of peer conflicts among a sample of African American children and adolescents from inner-city schools. The development and social functions of four types of aggressive behaviors are investigated by this study. Social, direct relational, physical, and verbal aggression, these are the four types of aggressive behavior. A total of 489 participants in grades 1, 4, and 7 were interviewed (220 boys and 269 girls). In peer conflicts, results showed that low levels of social aggression and high levels of physical aggression were reported. In the comparisons of same-gender conflicts at grade 7 gender differences on social, direct relational and physical aggression were chiefly observed. Across different forms of aggression, distinct configurations were recognized. Boys with configurations of physical and/or verbal aggression had higher levels of school social network centrality than non-aggressive boys. Girls with configurations of social and/or direct relational aggression showed relatively higher levels of network centrality than non-aggressive girls.

Wei Chu *et al.* [22]. have proposed an F0 Frame Error (FFE) metric. To objectively assess the performance of fundamental frequency (F0) tracking methods, this metric is developed as a combination of Gross Pitch Error (GPE) and Voicing Decision Error (VDE). To demonstrate the trade-off between GPE and VDE a GPE-VDE curve is then developed. A model-based Unvoiced/Voiced (U/V) classification front end which could be used by any F0 tracking algorithm is introduced additionally. We train speaker independent U/V models, and then adapt them to speaker dependent models in an unsupervised fashion in the U/V classification. The U/V classification result was taken as a mask for F0 tracking. For the pitch tracker TEMPO, experiments using the KEELE corpus with additive noise show that our statistically-based U/V classifier could decrease VDE and FFE in both white and babble noise conditions, and that minimizing FFE as an alternative of VDE results in a reduction in error rates for a number of F0 tracks algorithms, especially in babble noise.

S.G. Patil *et al.* [23]. have proposed Planning and design of coastal protection works like floating pipe breakwater need information about the performance characteristics of the structure in reducing the wave energy. To forecast the wave transmission through floating breakwaters by taking into consideration all the boundary conditions, a number of researchers had carried out analytical and numerical studies on floating breakwaters in the past but failed to give a simple mathematical model. Artificial neural networks (ANN), fuzzy logic, genetic programming and support vector machine (SVM) are the Computational Intelligence techniques which were productively used to solve complex problems and to expect wave transmission of horizontally interlaced multilayer moored floating pipe breakwater (HIMMFPB), a hybrid genetic algorithm tuned support vector machine regression (GA-SVMR) model was developed. Also, by genetic algorithm optimal SVM and kernel parameters of GA-SVMR models are determined. Using regular wave flume, the GA-SVMR model is trained on the data set obtained from experimental wave transmission of HIMMFPB. The results were compared with ANN and Adaptive Neuro-Fuzzy Inference System (ANFIS) models in terms of correlation coefficient, root mean square error and scatter index. GA-SVMR performance is found to be reliably superior. For the given set of data B-spline kernel function performs better than other kernel functions.

Yune-Sang Lee *et al.* [24] have proposed although much effort had been directed on the way to understand the neural basis of

speech processing, the neural processes concerned in the definite perception of speech had been comparatively less studied, and numerous questions stay open. In this functional magnetic resonance imaging (fMRI) the cortical regions mediating categorical speech perception using an advanced brain-mapping technique, whole-brain multivariate pattern-based analysis (MVPA). Usual healthy human subjects (native English speakers) were scanned as they listened to 10 consonant-vowel syllables along the /ba-/da/ continuum. To partition the fMRI data into /ba/ and /da/ conditions per subject outside of the scanner, individuals' own category boundaries were calculated. The whole-brain MVPA revealed that area and the left pre-supplementary motor area evoked distinct neural activity patterns between the two perceptual categories (/ba/ vs /da/) area was also found when the same analysis was applied to another dataset, which previously yielded the supramarginal gyrus using a univariate adaptation-fMRI paradigm. The consistent MVPA findings from two independent datasets strongly indicate that area participates in categorical speech perception, with a possible role of translating speech signals into articulatory codes. The difference in results between univariate and multivariate pattern-based analyses of the same data suggest that processes in different cortical areas along the dorsal speech perception stream were distributed on different spatial scales.

Marco Jeub *et al.* [25] have proposed to describes a new database of binaural room impulse responses (BRIR), referred to as the Aachen Impulse Response (AIR) database. The evaluation of speech enhancement algorithms dealing with room reverberation is the main field of application of this database. The measurements with a dummy head took place in a low-reverberant studio booth, an office room, a meeting room and a lecture room. It covers a wide range of situations due to the different dimensions and acoustic properties where digital hearing aids or other hands-free devices can be used. Besides the description of the database, instead of monaural measurements a motivation for using binaural was given. Furthermore to show the advantage of this database for algorithm evaluation, an example using coherence based DE reverberation technique was provided. The AIR database is being made available online.

## 3. EVALUATING THE PERFORMANCE ANALYSIS BY SPEECH CLUSTERING AND CLASSIFICATION ALGORITHMS

There are two processes speech processing and clustering classification. The gender clustering is approved by using different distance algorithm as well as Euclidean distance algorithm. The gender classification supported by different classification algorithms such as hybrid neural network and fuzzy logic method calculated results are means different clustering classification algorithm. Finally results are evaluated shows that gender clustering and gender classification processes are explored by means of different clustering and classification algorithms based on the selection results will be obtained for both gender clustering and gender classification. In several research works considered using pitch as feature for computing gender clustering and classification processes. Usually, the pitch value of male speech is higher than the female speech. However, in some situations, the frequency of male is almost analogous to female or frequency of female is analogous to male. In such circumstances, it is tricky to identify the exact gender considering all these problems into account in our proposed method.

Three features are: EE, ZCR, and STE are making used for identifying the exact gender. By using the aforementioned

features, the gender clustering and classification are approved out successfully.

### 3.1. Evaluating Types for Gender Classification & Clustering Algorithm

In both gender clustering and gender classification, feature selection is the most prominent process carried out. In the proposed method the three features that are considered are:

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

ZCR is the most prominent feature among the above mentioned three features. All such features are explained briefly:

The fundamental operations of these three features are discussed below separately.

#### 1. Energy Entropy (EE)

An abrupt change in the energy level of a speech signal is referred to as EE. To partition the speech signal into frames is the first step in EE calculation and the normalized energy for each frame is computed subsequently. The EE is calculated as,

$$EE, E = -\sum_{i=0}^{k-1} \sigma^2 \cdot \log_2(\sigma^2) \quad (1)$$

Where,

$$\sigma^2 = \sum_{a=1}^N \frac{\left(N * \frac{L}{M}\right)^2}{F}$$

$\sigma^2$  Signifies the normalized energy,  $N$  is the total number of blocks,  $L$  is the window length,  $M$  is the number of short blocks, and  $F$  is the frequency.

To examine the performance of the feature, above mathematical model is applied to the sample signals and from the result, it is found that EE for male is low and distributed, but for females it is high and stays only for a short period of time.

#### 2. (Short Time Energy) STE

The STE of speech signal is referred as an sudden rise in the energy level. Partitioned the signal into windows is the preliminary process for STE computation and later, the windowing function is carried out for each window. The short time energy is calculated as,

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

Where,  $h(s)$  is the impulse response, and  $y(r)$  is the input signal. To analyze the performance of the feature, above mathematical model is applied to the sample signals and from the result; it is found that STE for male is low, whereas for female it is high and continuous.

#### 3. (Zero Crossings Ratio) ZCR

The most imperative feature considered in our proposed method is ZCR. The ratio of number of time-domain zero crossings occurred to the frame length is known as ZCR. The ZCR is calculated as:

$$ZCR, Z = \frac{1}{N} \sum_{i=1}^{N-1} \frac{\text{sgn}\{x(i)\} - \text{sgn}\{x(i-1)\}}{2} \quad (3)$$

Where,  $N$  is the length of the sample, and  $\text{sgn}\{x(i)\}$  signify the sign function, i.e.

$$\text{sgn}\{x(i)\} = \begin{cases} 1; & x(i) > 0 \\ 0; & x(i) = 0 \\ -1; & x(i) < 0 \end{cases} \quad (4)$$

To examine the performance of the feature, above mathematical model is applied to the sample signals and from the result; it is found that the ZCR for female speech is higher compared to the male speech. Computing gender clustering and classification processes by using the above-mentioned three features is the next process. At first, we discuss about the performance of speech clustering process.

### 3.2. Gender Clustering: Performance Exploration by Different Clustering Algorithms

The gender clustering in speech process for using Euclidean distance algorithm. Computing beginning value for all the above three features and the male & female speech signals are separated based on this beginning value is the initial process in this method. For a certain speech signal Euclidean distance method (EDM) based on the number of features present in male and female gender. By applying diverse algorithms such as, Mahalanobis distance, Manhattan distance, and Bhattacharyya distance the presentation of gender grouping is analyzed.

#### 3.2.1. Performance Analysis Using Mahalanobis Distance Method

Here, instead of using EDM, to analyze the performance of different distance algorithms, Mahalanobis distance algorithm (MDA) is applied. The mathematical model used to calculate the distance using MDA is shown in equation 5,

$$D = \sqrt{(x_s - y_s)^* C^{-1} (x_s - y_s)^T} \quad (5)$$

Where,  $C$  is the covariance matrix,  $x = \{x_1, x_2, \dots, x_s\}^T$  and  $y = \{y_1, y_2, \dots, y_s\}^T$

By using the above mathematical model, based on the result obtained from MDA the distance is computed for the given set of speech signal and the given set of speech signal is clustered.

#### 3.2.2. Performance Analysis Using Manhattan Distance Method

Here, Manhattan distance algorithm is applied instead of using EDM to evaluate the performance of different distance algorithms, and the mathematical model used to compute the distance using Manhattan distance method is shown in equation 6,

$$D = \sum_{i=1}^T |x_i - y_i| \quad (6)$$

Where,  $x = \{x_1, x_2, \dots, x_s\}^T$  and  $y = \{y_1, y_2, \dots, y_s\}^T$

By using the above mathematical model, the distance is computed for the given set of speech signal and the given set of speech signal is clustered based on the result obtained from Manhattan distance algorithm

### 3.2.3. Performance Analysis Using Bhattacharyya Distance Method

Here, Bhattacharyya distance method (BDM) is applied instead of using EDM to examine the performance of different distance algorithms, and the mathematical model used to compute the distance using BDM is shown in equation 7,

$$D = 0.125 * (m_1 - m_2)^T P^{-1} * (m_1 - m_2) + 0.5 * \ln \left( \frac{\det P}{\sqrt{\det P_1 * \det P_2}} \right) \quad (7)$$

Where,  $m_i$  and  $P_i$  are the mean and covariance of the distributions.

$$P = \frac{P_1 + P_2}{2} \quad (8)$$

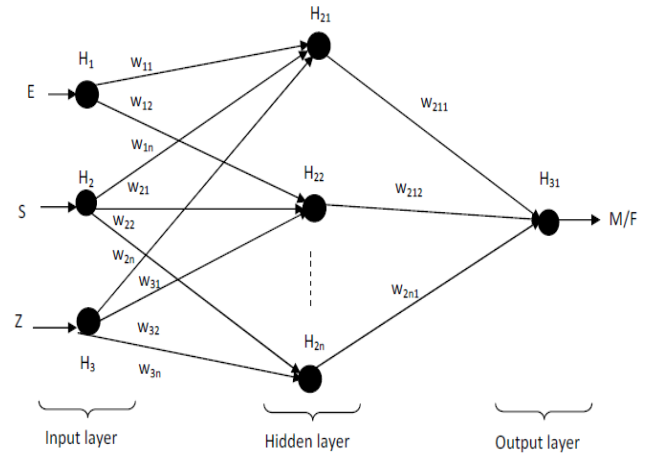
By using the above mathematical model, the distance is computed for the given set of speech signal and the given set of speech signal is clustered based on the result obtained from BDM.

### 3.3. Gender Classification: Performance Analysis Using Different Classification Algorithms

In gender classification, speech processing is obtained by using neural network (NN) and fuzzy logic (FL). For all the above three features, the initial process takes place in this method is computing threshold value and based on this threshold value the male & female speech signals are separated. Initially, by using different speech signal features values the FL and NN are trained. In the testing phase, initially it computes the three features and that feature values are given as input to the FL and NN, if a signal is given as input to FL and NN. NN and FL provide the percentage of male and female features present in the given speech signal as output and mean value is computed for FL and NN. The exact gender of the speaker is provided by the mean value of the given signal. Here by applying different other classification algorithms like Neuro fuzzy (NF) and support vector machine (SVM) instead of FL and NN. The performance of gender classification is analyzed.

#### 3.3.1. Performance Analysis Using Neuro Fuzzy

Here, NF is applied instead of NN and FL to analyze the performance of gender classification algorithms. It is well-known that Neuro fuzzy is a combination of NN and FL, where the fuzzy rules are created based on the number of input variables and the NN is trained using these rules. The three features namely, EE, ZCR, & STE are the input variables considered to NF and the output variable is male/female gender. The fuzzy rules are generated by considering the above mentioned input & output variables, and then NN is trained by considering the above generated fuzzy rules. In the testing phase, it provides the output as the speaker belongs to male or female gender if a certain speech signal is given as input to NF. The structure of the NN used in proposed method is shown in fig1.



**Fig. 1** Structure of neural network used in our proposed technique.

#### 3.3.2. Performance Analysis Using Support Vector Machine

SVM is applied instead of NN and FL to analyze the performance of gender classification algorithms. To know the speech signal belongs to which gender the SVM should be trained in order. The SVM is trained by an optimization function, which is given in equation [9].

$$f = \sum_{i=1}^n \alpha_i y_i K(s_i, x) + b \quad (9)$$

Where,  $\alpha$  is the coefficient,  $n$  is the number of samples,  $y$  is the label value of sample,  $K$  is the kernel function,  $b$  is the bias value,  $S$  is the supporting vector, and  $x$  is the feature vector to be classified. The SVM is trained by using the above equation, and after the completion of training process, if a speech signal as input, it provides the output as the given speech signal belongs to male or female gender.

## 4. EXPERIMENTAL RESULTS

The implementation of the proposed technique in MATLAB 7.11 is done and is tested with two different databases. The dataset used for evaluation, performance metrics and the results obtained from both the gender clustering and classification algorithms for both the datasets are described in this Section.

### 4.1. Dataset Description

For extensive analysis, we initially use a benchmark Harvard-Haskins database of 80 speech signals, which is partitioned into four datasets (will be further proceeded in the paper as Dataset I, II, III and IV). Each dataset holds equal number of male and female speech signals. Secondly, a database which is collected in a real environment is used. By selecting 25 samples with equal gender distribution i.e. 12 males and 13 females, the data collection is done. Each gender sample is asked to speak 10 different sentences. For evaluating the gender clustering and classification methodologies such real time dataset is also considered.

## 4.2. Performance Metrics

We make use of standard statistical performance measures, to study the performance of different gender clustering and classification methods as shown below

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (10)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (11)$$

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (12)$$

$$Specificity = \frac{TN}{(FP + TN)} \quad (13)$$

$$Likelihoodratio\ positive(LRP) = \frac{Sensitivity}{(1 - Specificity)} \quad (14)$$

$$Likelihoodratio\ negative(LRN) = \frac{(1 - Sensitivity)}{Specificity} \quad (15)$$

$$False\ positive\ rate, \alpha = \frac{FP}{(FP + TN)} \quad (16)$$

$$False\ negative\ rate, \beta = \frac{FP}{(FP + TN)} \quad (17)$$

In gender clustering, the performance of different distance algorithms is described in the following section.

## 4.3. Performance Evaluation on Harvard-Haskins Database

In Harvard-Haskins Database, the performance values obtained from gender clustering algorithm were presented followed by gender classification algorithms deals gender clustering only with Euclidean distance. However, Mahalanobis distance, Manhattan distance and Bhattacharyya distance are the three distance algorithm that was used in the proposed method instead of Euclidean distance and the results were presented.

Earlier, through fuzzy logic and neural network gender classification was performed. In this method, to examine the performance we use Neuro-fuzzy and SVM in the place of Neural Network and Fuzzy logic. Additionally, with respect to different threshold values we examine the performance variation, previously it was used as 0.5.

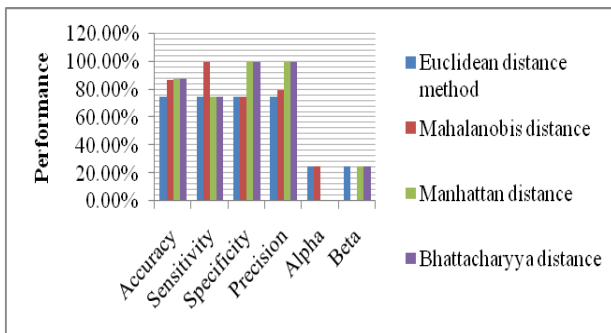


Fig. 2 Performance of different Distance Algorithms of Gender Clustering on Dataset I

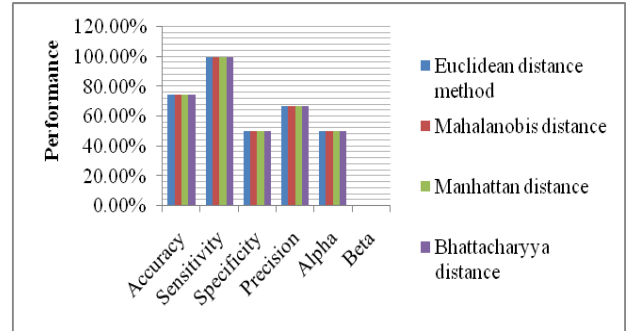


Fig. 3 Performance of Different Distance Algorithms of Gender Clustering On Dataset II

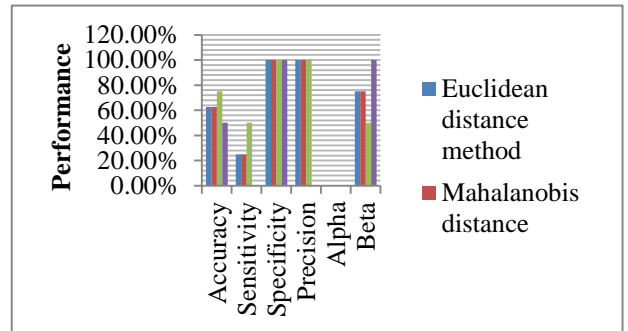


Fig. 4 Performance of Different Distance Algorithms of Gender Clustering On Dataset III

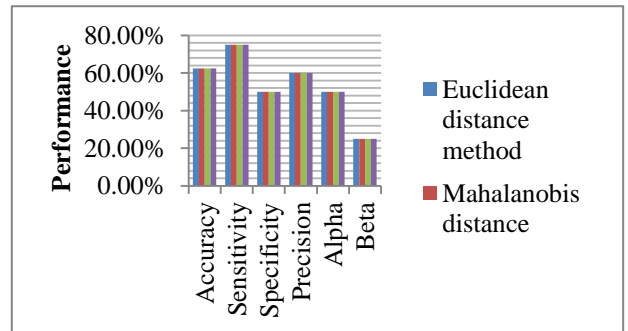
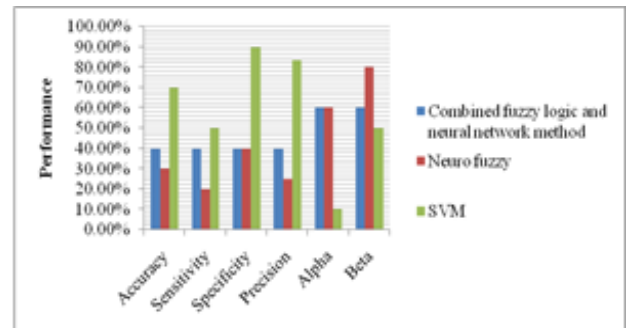
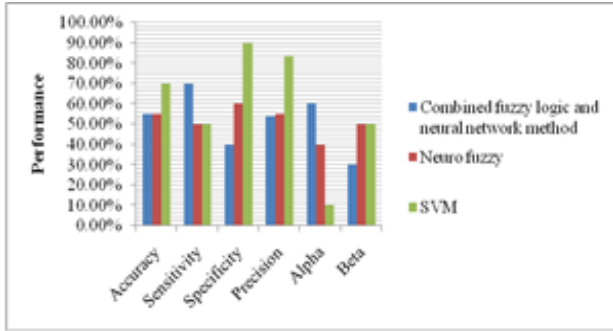


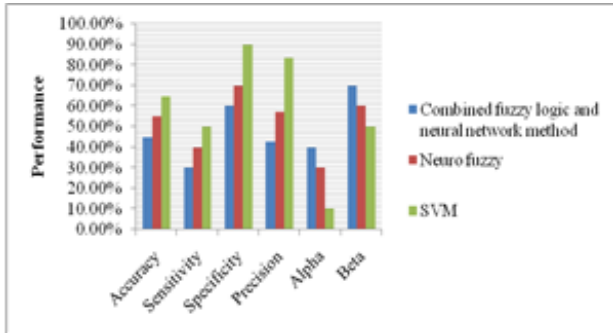
Fig. 5 Performance of Different Distance Algorithms of Gender Clustering On Dataset IV



(a)

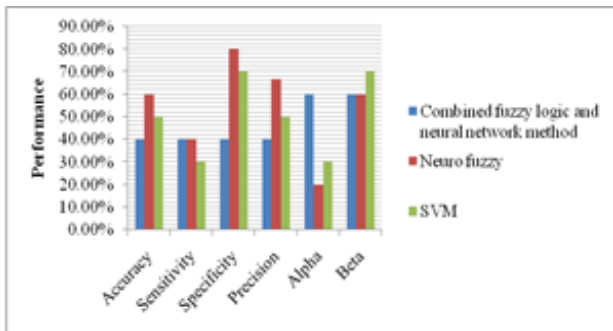


(b)

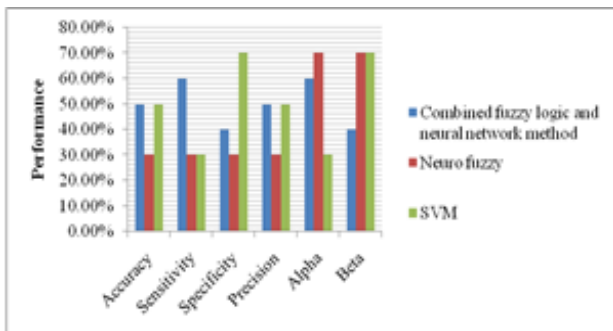


(c)

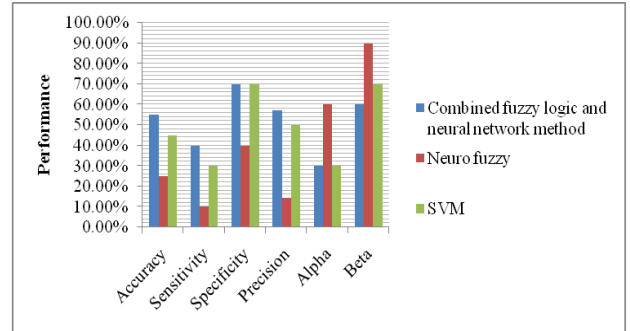
**Fig 6** Performance of Combined Fuzzy Logic and Neural Network, Neuro-Fuzzy and SVM Method on Dataset I at Threshold (A) 0.5, (B) 0.1 and (C) 0.8



(a)

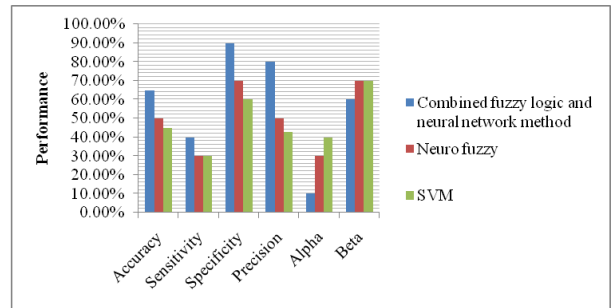


(b)

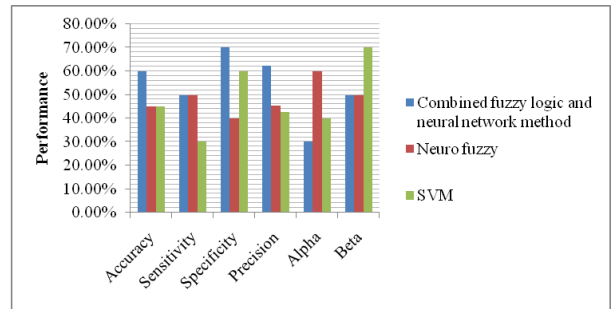


(c)

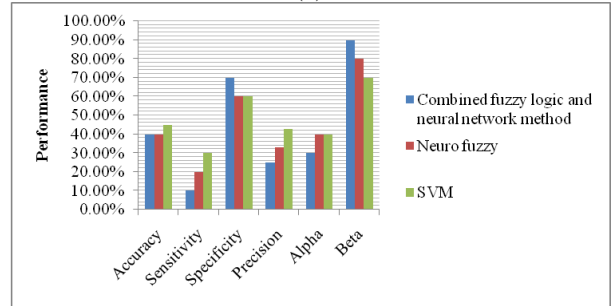
**Fig. 7** Performance of Combined Fuzzy Logic and Neural Network, Neuro-Fuzzy and SVM Method on Dataset II at Threshold (A) 0.5, (B) 0.1 and (C) 0.8



(a)



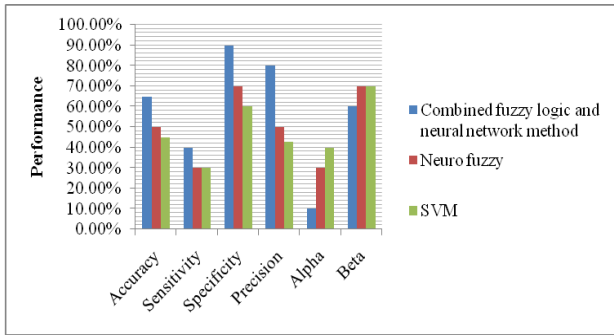
(b)



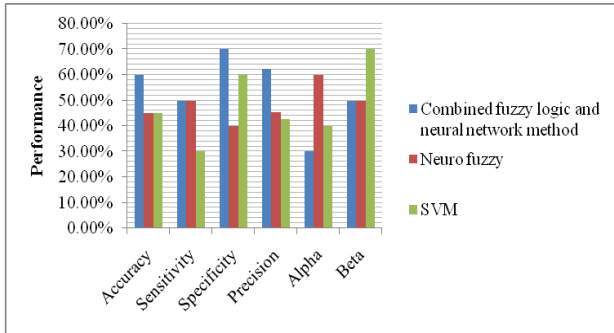
(c)

**Fig 8** Performance of Combined Fuzzy logic and Neural Network, Neuro-Fuzzy and SVM method on Dataset III at Threshold (a) 0.5, (b) 0.1 and (c) 0.8

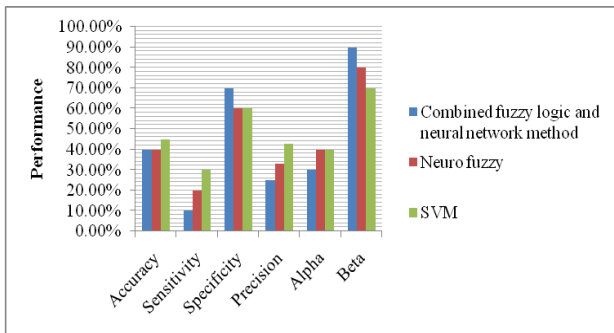




(a)



(b)



(c)

**Fig 9** Performance of Combined Fuzzy Logic and Neural Network, Neuro-Fuzzy and SVM Method on Dataset IV at Threshold (A) 0.5, (B) 0.1 and (C) 0.8

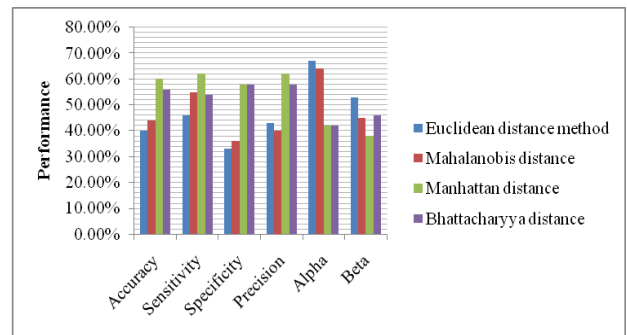
## Discussion

Fig 2, 3, 4 and 5 show the performance variations of gender clustering methods, when enabled with the four distance algorithms, on dataset i, ii, iii and iv, respectively. If it is seen, every distance algorithm relatively dominates and also relatively fails under certain circumstances. To validate the algorithms, average performance measures for all the datasets are taken here Euclidean distance achieved 68.75% of accuracy, mahalanobis distance achieved 71.75% of accuracy, Manhattan distance achieved 75% of accuracy and Bhattacharyya distance method achieved 68.75% of mean accuracy. Among the four, Manhattan distance is relatively higher accuracy, which is of 4.5% more accuracy than that of Manhattan distance. Likely, Manhattan distance achieved 75%, 75% and 82% of mean sensitivity, specificity and precision, respectively, which are again relatively higher than the other distance algorithms. Average false positive rate (or alpha) and average false negative rate (or beta) of Manhattan distance are 25% and 25% respectively, which are relatively lesser than (should be lesser) than the other distance measures.

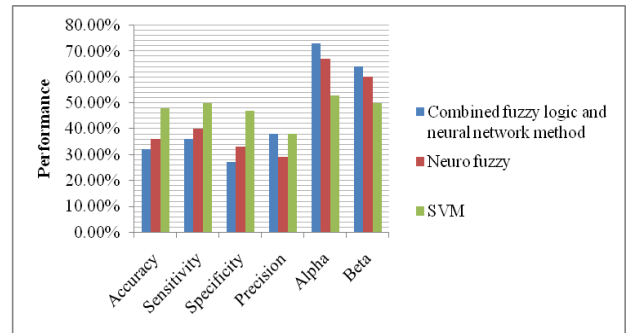
Fig 6, 7, 8 and 9 show the performance of different gender classification system on dataset i, ii, iii and iv respectively. when threshold is 0.5, the combined version of neural network and fuzzy logic achieved an average accuracy of about 52.5%, Neuro-fuzzy achieved 42.5% and SVM achieved 60%, whereas at threshold is 0.1, 50%, 38.75% and 58% and at threshold is 0.8, 46.25%, 38.75% and 56.25% were achieved by the combined neural network and fuzzy logic, Neuro-fuzzy and SVM, respectively. Similarly, in the case of other measures, SVM achieved 77.5%, 42.5% and 64.45% sensitivity, specificity and precision, whereas, Neuro-Fuzzy achieved 55%, 30% and 42.93% and combined neural network and fuzzy logic achieved 62.5%, 42.5% and 57.85%, sensitivity, specificity and precision respectively. The statistical values say that SVM enables accurate, precise and sensitive gender classification. Also, alpha and beta values are relatively lesser, when SVM is used.

## 4.4. Performance Evaluation on Real Time Database

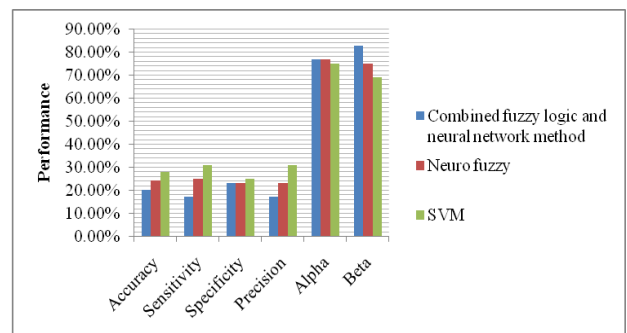
As similarly done on Harvard-Haskins Database, experiments are followed on the collected Real Time Database and the results are presented.



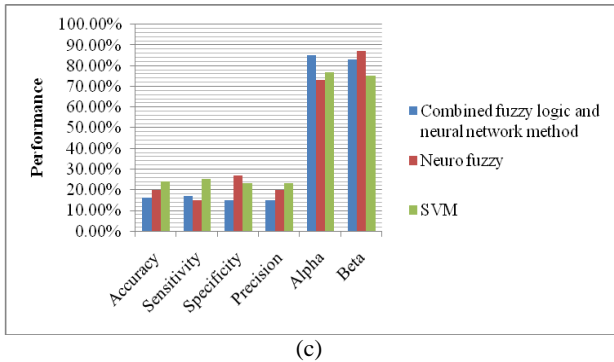
**Fig.10** Performance of different Distance Algorithms for Gender Clustering on Real-Time Dataset



(a)



(b)



**Fig. 11** Performance of Gender Classification at Threshold a) 0.5, (b) 0.1 and (c) 0.8, using Combined Fuzzy logic and Neural Network, Neuro-Fuzzy and SVM on Real-Time Dataset

**Discussion:** Fig 10 shows the performance values of gender clustering methodology with various distance algorithms such as Euclidean distance, Mahalanobis distance, Manhattan distance and Bhattacharyya distance. As the speech is acquired in a real environment and the signals are statements unlike the benchmark datasets, the gender clustering and classification performance is rather lesser than the performance on the benchmark dataset. In the real-time dataset, a maximum of 60%, 62%, 58% and 62% accuracy, sensitivity, specificity and precision is achieved in gender clustering, respectively, only when using Manhattan distance. Hence, again Manhattan distance is proved to be the better distance algorithm than the other three algorithms. Similarly, SVM shines in the gender classification methodology, when compared to the other classification methodologies (source: Fig 11).

## 5. CONCLUSION

By means of different algorithms the gender clustering and classification of speech signals were carried out and analyzed. We used a benchmark dataset and a real-time dataset for analysis purpose. By using Euclidean distance, Mahalanobis, Manhattan distance and Bhattacharyya distance gender clustering was done. Among the four, in both the benchmark dataset and real-time dataset the gender clustering using Manhattan distance performed well. Meanwhile, using Combined Fuzzy logic and neural network, Neuro-Fuzzy and Support Vector Machine gender classification is analyzed. Among the three variances, in both the datasets the gender classification using SVM dominated. Moreover, the optimal threshold could be set as 0.5 after detailed analysis under various datasets.

## REFERENCES

- [1] K.Sreenivasa Rao, Ramu Reddy, Sudhamay Maity, and Shashidhar G Koolagudi, "Characterization of Emotions Using the Dynamics of ProsodicFeatures", In Proceeding of Devices and Communications (ICDeCom) International Conference onThe Dynamics of ProsodicFeatures, The Dynamics Of ProsodicFeatures, pp.1-4, 2010.
- [2] Andrea DeMarco, and Stephen J. Cox, "An Accurate and Robust Gender Identification Algorithm", Journal of Neuroscience Methods, Vol. 172, No. 1, pp. 122-130, 2008.
- [3] Yakun Hu, Dapeng Wu, and Antonio Nucci, "Pitch-based Gender Identification with Two-stage Classification", Security and Communication Networks, Vol.5, No.2, pp.211-225, Feb 2012.
- [4] I. Trancoso, T. Pellegrini, J. Portelo, H. Meinedo, M. Bugalho, A. Abad, and J. Neto, "Audio Contributions to

Semantic Video Search", In proceeding of Multimedia and Expo, ICME. IEEE International Conference onSemantic Video Search, pp.630-633, July 2009.

- [5] AnvitaBajpa, and B.Yegnanarayana, "CombiningEvidencefrom Subsegmental and SegmentalFeaturesfor AudioClipClassification", In proceeding of IEEE Region Conference (TENCON), Hyderabad, India, No.10, pp. 1-6, Nov 2008.
- [6] Hugo Meinedo, and Joao Neto, "A Stream-based Audio Segmentation, Classification and Clustering Pre-processing System for Broadcast News using ANN Models", In proceeding of IEEE International Conference on Acoustics Speech and Signal Processing, ISBN: 0780376633, Vol.2, pp.237-240, 2003.
- [7] Arnulf B. A. Graf, and Felix A. Wichmann, "Gender Classification of Human Faces", Lecture Notes in Computer Science, Vol. 2525, pp.1-18, DOI: 10.1007/3-540-36181-2\_49, 2002.
- [8] Kotti. M, and Kotropoulos. C, "Gender Classification In Two Emotional Speech Databases", In Proceedings of 19th International Conference on Pattern Recognition, pp. 1-4, Tampa, Dec 2008.
- [9] Yu-Min Zengi, Zhen-Yang Wu, Tiago Falk, and Wai-Yip Chan, "Robust GMM based Gender Classification using Pitch and Rasta-PLP Parameters of Speech", In Proceedings of Fifth International Conference on Machine Learning and Cybernetics, pp.13-16, Dalian, Aug 2006.
- [10] Yoko Hasegawa, and Kazue Hata, "Non-Physiological Differences Between Male and Female Speech: Evidence from the Delayed F0 Fall Phenomenon in Japanese", In Proceedings of 1994 International Conference on Spoken Language Processing, pp. 1179-82, 1994.
- [11] Yoko Hasegawa, and Kazue Hata, "The Function of F0-Peak Delay in Japanese", In Proceedings of 21st Annual Meeting of the Berkeley Linguistics Society, pp. 141-151, 1995.
- [12] Theodoros Giannakopoulos, Aggelos Pikrakis, and Sergios Theodoridis, "A Speech/Music Discriminator for Radio Recordings using Bayesian Networks", In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 809-812, Toulouse, 2006.
- [13] M. Faúndez-Zanuy, S. McLaughlin, A. Esposito, A. Hussain, J. Schoentgen, G. Kubin, W. B. Kleijn and P. Maragos, "Non-linear Speech Processing: Overview and Applications, Control & Intelligent Systems", ACTA Press, Vol.30, No.1, pp. 1-10, 2002.
- [14] Gurpreet Singh, Akhil Junghare, and Priyam Chokhani, "Multi Utility E-Controlled cum Voice Operated Farm Vehicle", International Journal of Computer Applications, Vol. 1, No. 13, pp. 109-113, 2010.
- [15] Ramzi A. Haraty, and Omar El Ariss, "CASRA+: A Colloquial Arabic Speech Recognition Application", American Journal of Applied Sciences, Vol. 4, No.1, pp. 23-32, 2007.
- [16] Ibrahim Patel, and Y. Srinivas Rao, "SpeechRecognition using HMM with MFCC- an Analysis using Frequency Spectral De composition Technique", Signal & Image Processing : An International Journal (SIPIJ), Vol. 1, No. 2, pp.101-110, Dec 2010.



- [17] Anandthirtha. B. Gudi, and H. C. Nagaraj, "Optimal Curve Fitting of Speech Signal for Disabled Children", *International Journal of Computer science & Information Technology (IJCSIT)*, Vol. 1, No 2, pp. 99-107, Nov 2009.
- [18] R.J. McAulay, and T.F. Quatieri, "Speech Processing Based on a Sinusoidal Model", *The Lincoln Laboratory Journal*, Vol. 1, No. 2, pp. 153-168, 1988.
- [19] Yingyong Qi, and Bobby R. Hunt, "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, April 1993.
- [20] James A. Rodger, Parag C. Pendharkar, "A field Study of the Impact of Gender and User's Technical Experience on the Performance of Voice-Activated Medical Tracking Application", *Int. J. Human-Computer Studies*, Vol. 60, pp. 529–544, 2004.
- [21] Hongling Xie, Thomas W. Farmer, and Beverley D. Cairns, "Different forms of aggression among inner-city African-American children: Gender, configurations, and school social networks", *Journal of School Psychology*, Vol.41, pp.355 – 375, March 2003.
- [22] Wei Chu, and Abeer Alwan, "Reducing F0 Frame Error of F0 Tracking Algorithms Under Noisy Conditions With An Unvoiced/Voiced Classification Frontend", ' *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, Vol. 9, pp. 3969-3972, 2009.
- [23] S.G. Patil, S. Mandal, and A V Hegde, "Genetic algorithm based support vector machine regression in predicting wave transmission of horizontally Interlaced Multi-layer Moored floating pipe breakwater ", *Engineering & Instrumentation*, Elsevier, Vol.45, No. pp.203–212, 2012.
- [24] Yune-Sang Lee<sup>1</sup>, Peter Turkeltaub, Richard Granger, and Rajeev D. S. Raizada, "Categorical Speech Processing in Broca's Area: An fMRI Study Using Multivariate Pattern-Based Analysis", *The Journal of Neuroscience*, Vol.32, No.24, JAN 2012.
- [25] Marco Jeub, Magnus Schafer, and Peter Vary, "A Binaural Room Impulse Response Database for the Evaluation of Dereverberation Algorithms", *IEEE, IET, EURASIP*, Vol.16, pp.1- 4, July 2009.