Discrimination of Male and Female Voice Using Occurrence Pattern of Spectral Flux

Ghazaala Yasmin
Dept. of Comp. Sc. & Engg.
St. Thomas' College of Engineering & Technology
Kolkata, India
me.ghazaalayasmin@gmail.com

Suchibrota Dutta
Dept. of Computer Sc. & Application
Maharaja Manindra Chandra College
Kolkata, India
suchibrota@gmail.com

Arijit Ghosal
Dept. of Information Technology
St. Thomas' College of Engineering & Technology
Kolkata, India
ghosal.arijit@yahoo.com

***Abstract*— In this paper, a plain sailing scheme has been propsed for the purpose of discriminating male voice and female voice. Importance of gender identification system is increasing gradually due to its wide application area. Male-female voice discrimination portrays convincing role in the domain of security services, criminal investigation and speaker identification. Voice based identification of male and female is more robust than the other methods currently available. Constant research is going on to classify male and female voice in a better way. Researchers have worked with various types of features including some time domain features like Zero Crossing Rate (ZCR), Short Time Energy (STE) because of its capability to represent the physical characteristics of a certain audio signal. Inspiring from the certitude that male female voice discrimination is well reflected in frequency domain due to the contradiction in biological layout of their vocal cord. Being a frequency domain feature, spectral flux has been adopted for this work. This will expectantly discriminate male and female voice precisely. Spectral-flux based features have been used in the suggested effort which is frequency domain feature. The proposed work deals here to generate a co-occurrence matrix from the spectral flux and then to design feature set based on the extracted features from the matrix. For classification purpose some standard classifiers like RANSAC, k-NN and Neural-Net. The experimental result represents the strength of the proposed feature set.**

Keywords*— *Male-female voice discrimination; Spectral flux; Co-occurrence matrix; RANSAC

I. INTRODUCTION

Speech is generally supposed to convey some message. This message is passed using a sequence of legal sound units of a language. When a speaker speaks something, it delivers a lot of information. A spoken word predominantly carries the biological information about the speaker like gender, age, and emotion etc. along with the content of the message. Every speaker speaks in a unique way. Even if all speakers are uttering the same phrase of words their utterance become unique due to their biological differences. From the different utterance those speakers as well as their gender can be identified uniquely. Now-a-days detection of gender without any human intervention is gaining its popularity because of its wide application area like speech signal processing, applied physiology etc. Automatic voice recognition also helps us for gaining access control and it is also used in security systems. Male-female voice discrimination is the first step of voice or speaker recognition. When someone discriminates male and female by analyzing their personal appearance their voice plays a significant role along with other characteristics. As the biological structure of vocal cord is unique for every person so their voice will differ mainly from each other in frequency domain. This has motivated us to explore frequency domain features for the purpose of male-female voice discrimination. Moreover when huge set of voices will be required to discriminate into male and female voice it will not be possible to do that manually, computer based male-female voice discrimination is required.

In this anatomy, a simple way of discriminating male and female voice based on frequency domain feature set. Male and female voice mainly differs in frequency domain, the suggested effort have been concentrated into frequency domain features for the purpose of discrimination of them. A Feature set based on spectral flux has been generated which is a frequency domain feature. For classification purpose some well known standard classifiers like – RANSAC, k-NN and Neural-Network has been adopted. If compute mean and standard deviation from the spectral flux minute study of this important frequency domain feature will not be possible. Some researchers have already worked with spectral flux and their work reveals that it is an important frequency domain feature which has the capability to discriminate male and female quite efficiently. This has motivated us to minutely study this frequency domain feature. For this reason the concept of co-occurrence matrix has been introduced which helps to study the characteristics of any feature in detail. First co-occurrence matrix from spectral flux has been computed. Followed by the extraction of some well known features from the co-occurrence matrix .The purpose of the feature extraction is to minutely study the nature of spectral flux.

Researchers are spending time to generate feature set based on various types of features to capture acoustic heterogeneity between female and male voice. Gender classification system is drawing attention of all the speech researchers due to its wide application area. Starting from the domain of security, this male-female voice discrimination task renders important role in the field of speech processing, speaker identification,

978-1-5090-6106-8/17/$31.00 ©2017 IEEE

discriminating speech from background music etc. Moreover in the current scenario of different developing country, to make ITes relevant towards making a country digital, voice based computer applications is unavoidable. Male-female voice discrimination is the first step of these voice based applications. It has also been observed that a better result can be achieved, if separate aural model in case of female and male is applied.

Ali, Islam and Hossain [1] have worked with frequency domain features and they have used some frequency domain features like power spectrum density, frequency at maximum power in their system for gender recognition. Davis and Mermelstein [2] also used frequency domain features in their work. Deiv et. al [3] have detected gender in case of Hindi speech and they have used MFCC for detection of gender in Hindi speech. Feld, Burkhardt and Müller [4] have made a super vector system by mixing Gaussian Mixture Model and Support Vector Machine as classifier for discrimination of male and female gender and for recognition of speaker age. Ghai and Singh [5] have done a literature review on automatic speech recognition. Ghosal and Dutta [6] have used spectral flux as well as zero crossing rate and short time energy based features for male-female voice discrimination. Harb and Chen [7] have used statistics of first order spectrum and neural network for the purpose of gender detection. Hillenbrand et. al [8] also worked with frequency domain features. B. Jena and Beda Prakash Panigrahi [9] have classified gender based on pitch. Jung-Won Lee, Hong-Goo Kang, et al [10] has worked on the characteristics of vocal tract. Li et. al [11] has used utterance level features like prosodic, acoustic, and voice quality information along with MFCC in their work for the purpose of gender detection. Massida et al [12] have worked with cochlear implants (CIs) for discriminating voice gender for the purpose of recovering auditory abilities for deaf persons. Meena et. al [13] have used well-liked time domain features like short time energy, energy entropy and zero crossing rate as features and fuzzy logic and neural network for identification of gender of speaker. Pahwa and Aggarwal [14] have worked with Mel Frequency Cepstral Coefficient (MFCC) for the purpose of speech recognition in Hindi. Pronobis and Magimai.-Doss [15] have worked with the frequency domain features for male female voice classification. They have used Linear Prediction Cepstral Coefficients (LPCCs), Mel Frequency Cepstral Coefficients (MFCCs) and Perceptual Linear Prediction Coefficients (PLPs) as frequency domain features. They have also used Fundamental Frequency (F_0) in their frequency set as frequency domain feature. For classification purpose they have used Support Vector Machine (SVM). They have built a good robust dataset which comprised of different acoustic conditions for their work. R. Rajeshwara Rao and A. Prasad [16] have followed Linear Predictive Coding (LPC) approach in order to capture the information of vocal tract system like glottal excitement. They have applied MFCC on Linear Predictive Residual Signal. Sedaaghi [17] has worked with different classifiers like support vector machines (SVMs), probabilistic Neural Networks (PNNs), Gaussian mixture model (GMM) and the K nearest neighbor (K-NN). Shue and Iseli [18] have also picked up support vector machine for gender classification. Singh and Rajan [19] have calculated MFCC and inverted MFCC by Gaussian Filters to recognize speakers. Washani and Sharma [20] have reviewed speech recognition system. Yücesoy and Nabiyev [21] have used prosodic and spectral features and as classifier they have used Gaussian Mixture Model (GMM) and Support Vector Machine (SVM). Yu-Min Zeng and Zhen-Yang Wu et al [22] worked on classification of gender using Gaussian Mixture Model (GMM).

The paper is organized in this way: Our proposed solution is described in Section II. The experimental results are explained in Section III. Finally the conclusion is put in the Section IV.

II. PROPOSED METHODOLOGY

To discriminate male and female voice two steps has been instigated – i) pulling out the features from the input voice signal and then ii) determination of the category (male or female) of the input voice signal depending on the extracted features. But extraction of features is not an easy task as the features must represent the auditory characteristics of both male and female.

The proposed effort is in search of sound features intended for male-female voice discrimination. The aim is to generate a feature set which will be language independent so that this feature set can reflect a speaker's nonlinguistic characteristics. At the same time it has been intended to build the proposed model such a way that it works well in different acoustic conditions like speech of different languages, compressed speech, noisy speech, speech having telephonic quality, studio quality speech and so on.

A. Feature Extraction

Former dissertations reveal that male and female voice mainly differs in frequency domain due to biological differences in vocal region of every person. From the past study it is also get to know that spectral flux is a strong frequency domain feature which has the capability of differentiating male and female voice efficiently. Inspired from this learning, the prime attention has been paid on spectral flux to minutely study its characteristics. Researchers have worked with spectral flux but they have considered only the mean of the spectral flux. Perceiving from the fact that mean does not represent the overall characteristics of a feature. Consequently the spectral flux cannot be observed properly only from this conviction. In the field of image processing co-occurrence matrix is a very popular and widely used technique to measure or observe the detailed characteristics of a certain feature. Co-occurrence matrix is used to capture the distribution of co-occurring values within a feature. The concept to capture the distribution of co-occurring values of spectral flux has been employed here. In image processing visual texture contains variations of intensities which essentially form certain repeated patterns. These patterns can be caused by textural surface component, such as roughness, or they could result from reflectance differences, such as the color on a surface. But the differences observed by visual inspection are difficult to define in quantitative manner, which leads to demand of defining texture using some features. Among the current approaches used in image processing for describing texture, statistical approach is the widely used because it produces good results with low computational costs.

Once the co-occurrence matrix is formed, certain textural features are extracted from it – inertia, entropy, energy, inverse difference and correlation with distance 1 and angle 0 degree. The spatial relationship between two adjacent pixels can be expressed in several ways with different offsets and angles, the default one is between a pixel and its immediate neighbor to its right. Generally, possible relationships that are of directions 0^0 are specified and then they are implemented.

Mathematically, for a certain given image I of size L×L, the elements of a G×G gray-level co-occurrence matrix M_{CO} for a displacement vector d is defined as

$$M_{CO} = \sum_{q=1}^{L}\sum_{r=1}^{L}\begin{cases}1, \text{if } I(q,r) = i \text{ and } I(q+d_q, r+d_r) = j\\ 0, \text{otherwise}\end{cases} \quad (1)$$

Where I(q,r) indicates the pixel value at pixel (q,r) and the offsets d_q,d_r define the spatial link for which the corresponding matrix is computed.

Each element of the co-occurrence matrix is the number of times two adjacent values are neighborhood in distance d and angle θ. Thus, the matrix M_{CO} represents the distribution of pair wise occurrence of different pixel values.

The same concept of image processing is applied here to determine the textural features of spectral flux. Spectral flux is primarily the spectral variation. For an audio signal, spectral flux measures how fast the power spectrum of that audio signal revises. It is defined as the value of variation in spectrum between the two consecutive frames in a window which is short-time analyzed. Spectral flux is a good measure about the amount of spectral change of a signal and the fact that spectral change is quite different in male audio signal in comparison to female audio signal. For this reason spectral flux based features have been chosen though it is an important feature for discriminating music and speech. Spectral flux is determined as:

$$SF(X) = \sum_{j=0}^{t-1} s(X,j) - s(X-1,j) \quad (2)$$

where, t = total number of frames. Spectral flux for the X^{th} spectrum is represented by SF(X). Let the value of the j^{th} bin in X^{th} spectrum is s(X,j), then s(X-1,j) is similar for the spectrum before X. Now the values of each bin of the previous spectrum are subtracted from the values of its matching bin in the present spectrum and then summing up those differences to materialize a final value which is the desired final spectral flux for spectrum X.

Once the co-occurrence matrix of spectral flux is formed using equation (1), the following textural features are calculated from the co-occurrence matrix of spectral flux:

i) $Inertia = \sum_i \sum_j (i-j)^2 M_{CO}[i][j]$ (3)

ii) $Entropy = -\sum_i \sum_j M_{CO}[i][j] \log_2 M_{CO}[i][j]$ (4)

iii) $Energy = \sum_i \sum_j [M_{CO}[i][j]]^2$ (5)

iv) $Inverse\ Difference = \sum_i \sum_j M_{CO}[i][j]/|i-j|$ where $i \neq j$ (6)

v) $Correlation = (1/\sigma_x\sigma_y) \sum_i \sum_j (i-\mu_x)(j-\mu_y) M_{CO}[i][j]$ (7)

where,

$M_{CO}[i][j]$ = Value of co-occurrence matrix at position [i][j]

$\mu_x = \sum_i i \sum_j M_{CO}[i][j]$

$\mu_y = \sum_j j \sum_i M_{CO}[i][j]$

$\sigma^2_x = \sum_i (1-\mu_x)^2 \sum_j M_{CO}[i][j]$

$\sigma^2_y = \sum_j (1-\mu_y)^2 \sum_i M_{CO}[i][j]$

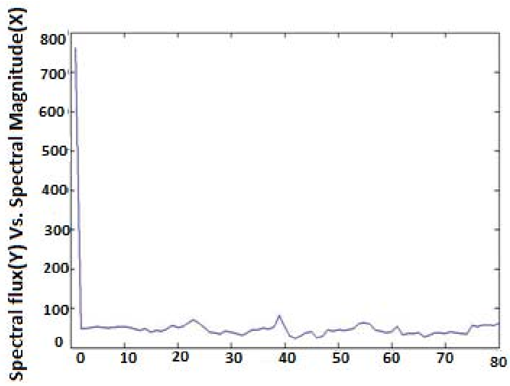

Fig. 1. Spectral flux plot of male voice

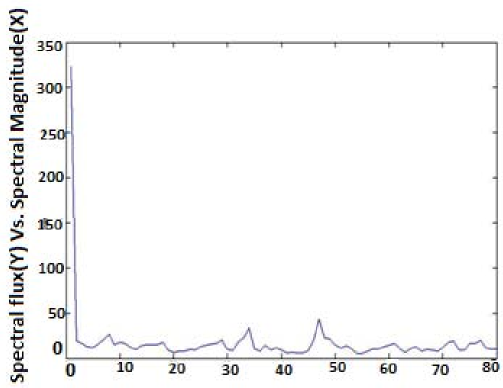

Fig. 2. Spectral flux plot of female voice

The mentioned work deals with the generation of co-occurrence matrix of spectral flux for 0 degree only because co-occurrence matrix cannot be formed as spectral flux is one dimensional.

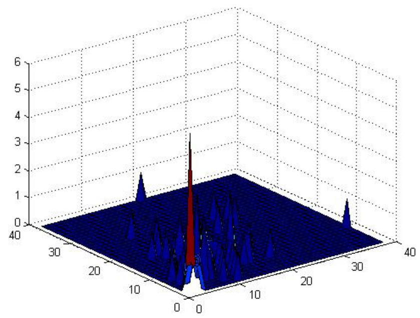

Fig. 3. Plot of 0^0 co-occurrence matrix of spectral flux for male voice

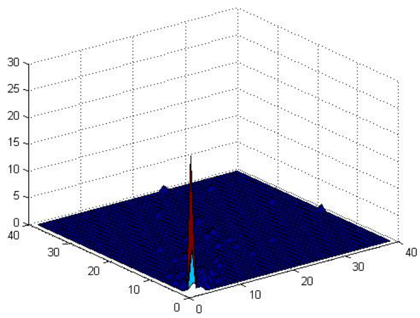

Fig. 4. Plot of 0^0 co-occurrence matrix of spectral flux for female voice

For this effort the above 5 textural features calculated from the co-occurrence matrix of spectral flux forms the 5-dimensional feature vector.

B. Classification

Each of the voice clips are represented by the above generated 5-dimmensional feature vector and then proceeding to perform the following task which is male-female voice discrimination. For that occasion, standard well used classifiers like RANSAC, k-NN and Neural-Net has been imported for the suggested approach.

RANdom Sample And Consensus (RANSAC) is pretty popular in the field of image processing. RANSAC is primarily a re-sampling strategy that yields solutions by applying the least number of examinations which is denominated as data points and is required to estimate the underlying model parameters. Dissimilar to conventional sampling approaches which generally apply as much of the data as possible for attaining an initial solution and then further proceed to prune outliers, RANSAC employs the smallest feasible set and then prosecutes to extend this embryonic set with the stable data points. The prime convenience of using RANSAC over other methods is that here the estimation is totally constructed based on inliers which indicates whose distribution can be described by a set of prototype of parameters.

At first, minimal numbers of data are essential to discover the dummy parameters and they are adopted randomly. Subsequently these parameters are resolved. Thereafter, the number of data from the set of all data which fit with a predefined leeway € are computed. Now, if the fragment of the amount of inliers compared to the total number of data in the set crosses the predefined entry €, dummy parameters are re-estimated using all the recognized inliers and then it is stopped. If it is not, then these steps are rehearsed for at most N times, where N is a high value chosen in such a way so that the probability p (usually 0.99) that an outlier does not get included in one of the sets of random specimens. RANSAC estimates the model depending on the inliers and compared to other methods it is less exaggerated by the noisy data.

The nearest neighbor (NN) method consists of assigning to the unlabelled feature vector the label of the training vector which is nearest to it in the feature space. *k*-Nearest Neighbor (k-NN) is the simplest supervised classifier which provides the simplest decision procedure for classification. Based on the value of the nearest neighbor, this classification discriminates the sample. It endeavors to observe the similarity between the test model and every model in the training set. The resemblance is decided by the nearest neighbor distance. The Euclidean distance and City Block distance has been considered to measure the distance while computing the nearest neighbor distance between test model and each pattern of the training model. The approach also regards to random rule and nearest rule for tie break.

In this algorithm k as number of nearest neighbors has been considered. The k-NN classifier takes the k nearest, i.e. the closest, neighbors around a sample and uses k nearest neighbors to assign a label. This is generally done by a majority-voting rule, which states that the label assigned should be the one, which occurs most among the neighbors. If $k = 1$, then the testing pattern is just assigned to a class of its nearest training model. The success rate or classification accuracy of *k*-nearest neighbor can be better than the nearest neighbor ($k = 1$) procedure. The classification accuracy also depends on the choice of training set. The best accuracy has been achieved for k=3, City Block distance and nearest rule for tie-break and its performance is shown in Table I.

Neural-Net (NN) is a statistical representation or computational reproduction which is indeed inspired by the construction and/or operative facets of the biological neural networks of human body. An NN is usually elucidated by the following three kinds of parameters:

1. The kindred nature existing among distinct layers of neurons
2. The training procedure to apprise the interconnections weights
3. The activation task which transforms a neuron's weighted input data to its corresponding output stimulation.

The N-dimensional feature vector is considered as input to this neural network, where N indicates the number of features. There is a layer in the network which contains hidden neurons. These hidden neurons are not the part of the input and output layer of neural network. These hidden neurons in practical help the network to comprehend troublesome jobs by extricating more salient characteristics from the input feature vectors. The total number of neurons present in the said hidden layer of the network is specified in the experiment conducted. The required number of output classes specifies the number of neurons to be put at the output layer of the network. In the task of male-female voice discrimination the classification of input audio data into male and female voice and here it is supposed to have two neurons in output layer corresponding to the male and female voice respectively. The model has been designed by considering five neurons in the input layer corresponding to 5-dimmensional feature vector and three neurons in the hidden layer.

III. EXPERIMENTAL RESULTS

The experiment has been conducted for a wide range of diversification in aural dataset. An audio database is prepared consisting of 200 male voice files and 200 female voice files. All of these voice files are around 90 seconds duration. These files are obtained from CD recording, some of them are recording of live programs and rest is downloaded from various sites in the Internet. Sampling frequency is maintained at 22050 Hz, 16-bit per sample and of all of these voice files are of type mono. Voices spoken in different languages are likewise considered here while building the audio database. Different age groups for both male and female are considered here. Moreover the examination has been prolonged on noisy auditory files too. This has procreated our approach more robust and reliable for real outline.

To assess the features, each of the audio input file is cut up into a collection of frames. Each of these frames comprises of 150 samples and these frames are 50% overlapped with the previous frame to avoid missing of any change of property at the marginal position of a frame.

TABLE I. OVERALL CLASSIFICATION ACCURACY FOR MALE-FEMALE VOICE DISCRIMINATION

Classifier	*Male Speech*	*Female Speech*
RANSAC	96.5%	95.5%
k-NN	92.5%	91.5%
Neural Network	91%	90%

During the experiment 50% of each type of data has been chosen for training and the rest of the data have been used for testing purpose. Then, the experiment is done by reversing the training set and test data set. Average of testing accuracy is reflected in Table I.

The performance of our proposed methodology is compared with that of the works of Ali, Islam and Hossain [1], Jena and Panigrahi [9] and Pahwa and Aggarwal [14]. All the systems are implemented and are applied on the same dataset used in this work. The comparative results in terms of accuracy are shown in tabular format.

In the work of Ali, Islam and Hussain [1], Fast Fourier Transform of the data, Power spectrum from transformer data and the sample point at maximum power is computed. They have tested their system only with 5 male and 5 female people, but methodologies of the mentioned scholars has been tested with the current supplied data dataset which comprises of 200 male and 200 female people.

In the effort of Jena and Panigrahi [9], Pitch value is computed for each of the male and female voices present in our dataset. It has also been computed the same for our dataset. They have not shown any performance accuracy in their work. The features used in the substantial work have been implemented on our dataset and the performance accuracy is tabulated in Table II. It has been observed that average pitch value for female voice sample are higher than that of male voice samples.

In the work of Pahwa and Aggarwal [14], MFCC, delta MFCC and delta-delta MFCC features has been computed. Like the mentioned methodology of the past scholar have also considered only the first 13 coefficients for each of the audio files of our dataset resulting 39 dimensional feature vectors. All of the 39 feature values have been considered in one time and in other time first Mel Coefficients has been ignored.

TABLE II. COMPARISON OF PERFORMANCE IN TERMS OF CLASSIFICATION ACCURACY

Method	*Male Speech*	*Female Speech*
Ali, Islam and Hossain	78%	77%
Jena and Panigrahi	85%	83%
Pahwa and Aggarwal (Considering MFCC1)	91%	96%
Pahwa and Aggarwal (2016)(Ignoring MFCC1)	88%	95%

Table II clearly shows that the proposed feature set performs better than the other method. For Pahwa and Aggarwal [14] accuracy in detecting female voice is little higher than our proposed feature set but at the same time accuracy in detecting male voice is less than our proposed feature set. As RANSAC is providing the best accuracy over the other classifiers in our work, the suggested work deals with the comparison of all others performance against RANSAC and the performance are tabulated in Table II.

IV. CONCLUSIONS

For discriminating male and female voice a simple but efficient has been proposed for good feature set based on spectral flux. The concept of co-occurrence matrix has been wielded in the proposed work for analyzing the feature vector more precisely. This notion is well preferred for dealing with images. Experimental result shows the capability of this feature set for male female voice discrimination. To highlight the strength of the proposed feature set simple classification scheme has been taken into consideration. RANSAC, k-NN and Neural-Net has been adopted as classifiers. In future, further sub-classification of both female and male voice for language identification may be done using the concept of occurrence pattern for extrication of perceptual facets.

REFERENCES

[1] M. S. Ali,. M. S.Islam, and M. A. Hossain, 'Gender Recognition System Using Speech Signal', *International Journal of Computer Science, Engineering and Information Technology (IJCSEIT)*, Vol. 2, No. 1, 2012, pp. 1-9.

[2] S. Davis, and P. Mermelstein, 'Comparison of Parametric Representations Monosyllabic Word Recognition in Continuously Spoken Sentences', *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 28, No. 4, 1980, pp. 357–366.

[3] D. S. Deiv, Gaurav and M. Bhattacharya, 'Automatic Gender Identification for Hindi Speech Recognition', *International Journal of Computer Applications*, Vol. 31, No.5, 2011, pp. 1-8.

[4] M. Feld, F. Burkhardt, and C.A. Müller, 'Automatic Speaker Age and Gender Recognition in The Car For Tailoring Dialog and Mobile Services', *INTERSPEECH*, 2010, pp. 2834-2837.

[5] W. Ghai, and N. Singh, N., 'Literature Review on Automatic Speech Recognition', *International Journal of Computer Applications*, Vol. 41, No. 8, 2012, pp. 42-50.

[6] A. Ghosal, and S. Dutta, 'Automatic Male-Female Voice Discrimination'. *International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT),* 2014, pp. 731-735. IEEE.

[7] H. Harb, and L. Chen, 'Gender Identification Using a General Audio Classifier', *International Conference on Multimedia and Expo, ICME'03 Proceedings*, Vol. 2, 2003, pp. II-733, IEEE.

[8] J. Hillenbrand, L.A. Getty, M.J. Clark, K. Wheeler, 'Acoustic Characteristics of American English Vowels', *The Journal of the Acoustical society of America,* Vol.97, No. 5, 1995, pp. 3099–3111.

[9] B. Jena and B.P. Panigrahi, 'Gender Classification by Pitch Analysis', *International Journal on Advanced Computer Theory and Engineering*, Vol. 1, Issue 1, 2012, pp. 106-108.

[10] J.W. Lee, H.G. Kang, J.Y. Choi, and Y.I. Son, 'An Investigation of Vocal Tract Characteristics for Acoustic Discrimination of Pathological Voices', *BioMed Research International*, Vol. 2013, 2013, pp. 1-11.

[11] M. Li, K.J. Han, and S. Narayanan, 'Automatic Speaker Age and Gender Recognition Using Acoustic and Prosodic Level Information Fusion', *Computer Speech & Language*, Vol. 27, No. 1, 2013, pp. 151-167.

[12] Z. Massida, M. Marx, P. Belin, C. James, B. Fraysse, P. Barone, and O. Deguine, 'Gender Categorization in Cochlear Implant Users', *Journal of Speech, Language, and Hearing Research*, Vol. 56, No. 5, 2013, pp. 1389-1401.

[13] K. Meena, K.R. Subramaniam, and M. Gomathy, 'Gender Classification in Speech Recognition Using Fuzzy Logic and Neural Network'. *Int. Arab J. Inf. Technol.*, Vol. 10, No. 5, 2013, pp. 477-485.

[14] A. Pahwa, and G. Aggarwal, 'Speech Feature Extraction for Gender Recognition', *International Journal of Image, Graphics & Signal Processing,* Vol. 8, No. 9, 2016, pp. 17-25.

[15] M. Pronobis, M. Magimai.-Doss, 'Analysis of F0 and Cepstral Features for Robust Automatic Gender Recognition', No. LIDIAP-REPORT-2009-020. 2009, Idiap.

[16] R.R. Rao, and A. Prasad, 'Glottal Excitation Feature Based Gender Identification System Using Ergodic HMM', *International Journal of Computer Applications*, Vol. 17, No. 3, 2011, pp. 31-36.

[17] M.H. Sedaaghi, 'A Comparative Study of Gender and Age Classification in Speech Signals'. *Iranian Journal of Electrical and Electronic Engineering*, Vol. 5, No. 1, 2009, pp. 1-12.

[18] Y.L. Shue, and M. Iseli, 'The Role of Voice Source Measures on Automatic Gender Classification'. *IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP*, 2008, pp. 4493-4496. IEEE.

[19] S. Singh, and E.G. Rajan, 'Vector Quantization Approach for Speaker Recognition Using MFCC and Inverted MFCC', *International Journal of Computer Applications*, Vol. 17, No. 1, 2011, pp.0975-8887.

[20] N. Washani, and S. Sharma, 'Speech Recognition System: A Review', *International Journal of Computer Applications*, Vol. 115, No. 18, 2015, pp. 7-10.

[21] E. Yücesoy, and V.V. Nabiyev, 'A New Approach With Score-level Fusion For The Classification of A Speaker Age and Gender. *Computers & Electrical Engineering*, Vol. 53, 2016, pp. 29-39.

[22] Y.M. Zeng, Z.Y. Wu, T. Falk, and W.Y. Chan, 'Robust GMM Based Gender Classification Using Pitch and RASTA-PLP Parameters of Speech', *International Conference on Machine Learning and Cybernetics*, 2006, pp. 3376-3379, IEEE.