ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

Emotion Recognition of Speech Signals Using Priori Information of Speaker's Gender

Nisha Chandran[1], Mahesh. B.S[2]

PG Student [Signal Processing], Dept. of ECE, College of Engineering, Kalloopara, Pathanamthitta, Kerala, India[1]

Assistant Professor, Dept. of ECE, College of Engineering, Kalloopara, Pathanamthitta, Kerala, India[2]

ABSTRACT: Speech emotion recognition is defined as the extraction of emotional state of the speaker from the audio signal registrations that makes human-machine interface more convenient. In this work a system is proposed that allows recognizing a person's emotional state starting from audio signal registrations.In this system a priori knowledge of speaker's gender is extracted. The system is able to recognize people emotions (anger, boredom, disgust, fear. happiness, sadness, neutral state). It is composed of 2 subsystems- Gender Recognition (GR) & Emotion Recognition(ER). Different Gender Recognition algorithms-Thresholding method using Pitch, gender recognition by binary SVM classifier and gender recognition by ANN are employed. It is aimed at providing 'a priori' information about the gender of the speaker. SVM based classifier (both male SVM & female SVM) which employs the gender information and reduced feature vectors as inputs are used for Emotion Classification.

KEYWORDS: Gender Recognition, Emotion Recognition, Pitch Estimation, MFCC, Support Vector Machine (SVM), Artificial Neural Networks (ANN).

I.INTRODUCTION

Speech emotion recognition is one of the latest challenges in speech processing and Human Computer Interaction (HCI) in order to address the operational needs in real world applications. Emotions play an extremely important role in human mental life. It is a medium of expression of one's perspective or his mental state to others. There are few universal emotions- including Anger, Surprise, Fear, Happiness, Sadness, Disgust and Neutral state which any intelligent system with finite computational resources can be trained to identify or synthesize as required. The importance of automatically recognizing emotions in human speech has grown with increasing role of spoken language interfaces in the field of human-machine interaction to make the human machine interface more efficient. Although emotion detection from speech is a relatively new field of research, it has many potential applications. In human-computer interaction systems, emotion recognition systems could provide users withimproved services by being adaptive to their emotions. Recent years have been marked by the increasing development of personal robots, either used as new educational technologies or for pure entertainment. Among the capabilities these personal robots need, the most basic is the ability to grasp human emotions and in particular, they should be able to recognize human emotions as well as to express their own emotions. Indeed, emotions are not only crucial to human reasoning, but they are central to social regulation, and in particular to control dialog flows. Despite the great progress made in artificial intelligence, we are still far from being able to naturally interact with machines, partly because machines do not understand our emotion states.

The main aim of this paper is to recognize the emotional state of a speaker with an accuracy level often higher than the evaluated methods taken from the literature and to check how a priori knowledge of the speaker's gender allows a performance increase.Investigation of algorithms that would perform the better classification for a given set of features and comparison of classifiers with and without a priori information of speaker's gender using different gender recognition methods is also employed.

II.LITERATURE SURVEY

Many researches provide an in-depth insight into the wide range of classification algorithms available, such as: Neural Networks (NN), Gaussian Mixture Model (GMM), Hidden Markov Model (HMM), Maximum Likelihood Bayesian Classifier (MLC), K-nearest Neighbors (KNN) and Support Vector Machine (SVM).

DOI:10.15662/IJAREEIE.2016.0506002

GENDER RECOGNITION FEATURES: Together with the Mel Frequency Cepstral Coefficients (MFCC) [10], pitch is the most frequently used feature[11],[14] since it is a physiologically distinctive trait of a speaker's gender. Other employed features are formant frequencies and bandwidths, open quotient and source spectral tilt correlates [12], energy between adjacent formants [15], fractal dimension and fractal dimension complexity [13], jitter and shimmer (pitch and amplitude micro-variations, respectively), harmonics-to-noise ratio, distance between signal spectrum ,formants [16], short time energy, zero crossing rate(ZCR)etc.ZCR for female speech is higher than that for the males.

EMOTION RECOGNITION FEATURES: Mean, variance, median, minimum, maximum and range of the amplitude of the speech of speech Energy, pitch, first 4 formants and first 12 MFCC[13][14].Spectrum shape features like Center of Gravity, Standard Deviation, Skewness and Kurtosis, mean and standard deviation of the glottal pulse period, jitter local absolute, relative average perturbation, difference of difference period.

CLASSIFIER SELECTION :In the speech emotion recognition system after calculation of the features, the best features are provided to the classifier. A classifier recognizes the emotion in the speaker’s speech utterance. Various types of classifier have been proposed for the task of speech emotion recognition. Gaussian Mixtures Model (GMM), K-nearest neighbours (KNN), Hidden Markov Model (HMM) and Support Vector Machine (SVM), Artificial Neural Network (ANN), etc. are the classifiers used in the speech emotion recognition system. Each classifier has some advantages and limitations over the others. An optimum classifier is needed to train and build a classification model by using machine learning algorithms to predict the emotional states on the basis of the speech instances. Each classifier requires an initial phase in which it is trained to perform a correct classification and a subsequent phase in which the classifier is tested.Only when the global features are extracted from the training utterances, Gaussian Mixture Model is more suitable for speech emotion recognition. All the training and testing equations are based on the supposition that all vectors are independent therefore GMM cannot form temporal structure of the training data. For the best features a maximum accuracy of 78.77% could be achieved using GMM. In speaker independent recognition typical performance obtained of 75%, and that of 89.12% for speaker dependent recognition using GMM. Other classifier that is used for the emotion classification is an artificial neural network (ANN), which is used due to its ability to find nonlinear boundaries separating the emotional states. Out of the many types, feed forward neural network is used most frequently in speech emotion recognition.

Transforming the original feature set to a high dimensional feature space by using the kernel function is the main thought behind the support vector machine (SVM) classifier, which leads to get optimum classification in this new feature space. The kernel functions like linear, polynomial, radial basis function (RBF) can be used in SVM model for large extent. In the main applications like pattern recognition and classification problems, SVM classifier are generally used, and because of that it is used in the speech emotion recognition system. SVM is having much better classification performance compared to other classifiers.

III.PROBLEM IDENTIFICATION AND OBJECTIVE

FOR GENDER RECOGNITION FEATURES:

Problem: In most cases gender classification was based on considering pitch as the main feature. But 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. There are also certain limitations while considering this feature. The accurate pitch extraction is not an easy task due to the non-stationarity and quasi-periodicity of speech signal, as well as the interaction between the glottal excitation and the vocal tract.

Objective: By considering the aforementioned problem here a new method is proposed for gender classification method which considers two features. Energy entropy and MFCC along with pitch .Also instead of using thresholding function a binary classifier is used to differentiate between the speaker genders.

FOR EMOTION RECOGNITION FEATURES:

The prosodic features are known as the primary indicator of the speakers emotional states. With the different emotional state, corresponding changes occurs in the speak rate, pitch, energy, and spectrum and hence each property of each

Speech data is obtained from statistical values like mean, median, standard deviation, minimum, maximum etc. Also thousands of paralinguistic features are extracted and used in experiments as a whole set or reduced to a subset using feature selection techniques and classified to one of three categories: Prosodic,Voice Quality, Spectral and Derived. The spectral features have been found, when used in combination to other categories of features (or even as a stand-alone feature vector), to improve (or to achieve good) performance.

Problem: The lack of a widely accepted taxonomy of emotions and emotional states; the strong emotion manifestation dependency of the speaker etc make emotion recognition an extremely difficult task. Selecting the most relevant subset from the original feature set increase the performance of the classifier and on the other hand decrease the computational complexity.

Objective: Mel Frequency Cepstral Coefficients (MFCCs) is an example of spectral feature that achieve good results not only on speech processing in general but also on emotion recognition. The MFCC parameterization techniques aim to simulate the way how a sound is perceived by a human. Also, the spectrum estimated using a windowed periodogram via the discrete Fourier transformation (DFT) algorithm. Despite having low bias, a consequence of the windowing is increased estimator variance. An elegant technique for reducing the spectral variance is to replace a windowed periodogram estimate with a multiple windowed spectrum estimate. Formant frequencies was found to be very much important in the analysis of the emotional state of a person based on Linear predictive coding technique (LPC) for estimation of the formant frequencies. The first three formant frequencies can be taken as relevant features in the complete feature vector. Hence for emotion recognition the features MFCC, Spectrum Estimate and Formant Analysis are taken into consideration that is with reduced number of features.

IV. PAPER CONTRIBUTIONS

The paper presents a gender-driven emotion recognition system whose aim, starting from speech recordings, is to individuate the gender of speakers and then, on the basis of this information, to classify the emotion characterizing the speech signals. Concerning the first step, the paper proposes a gender recognition method based on the pitch. This method employs a typical speech signal feature and a novel extraction method. Employing pitch and MFCC as features, gender recognition using SVM and gender recognition using ANN are also done.Concerning the emotion recognition approach, the paper proposes a solution based on traditional features sets and classifiers but, differently from the state of the art, it employstwo classifiers (i.e., two Support Vector Machines): the one trained on the basis of signals recorded by male speakers and the other one trained by female speech signals. The choice between the two classifiers is driven by the gender information individuated through the gender recognition method.

EMOTION RECOGNITION WITHOUT GENDER INFORMATION

ALGORITHM:

1. Read the signal.
2. Signal is divided into frames.
3. Power Spectral Density for each frame is estimated.
4. Formant analysis is performed based on Linear Predictive Coding (LPC).
5. Mean value of formants is calculated.
6. MFCC is calculated.
7. Multi Class SVM classifier is trained using the extracted features of all signals from database (BES).
8. Test Signal is obtained and features are extracted.
9. Test data is classified in accordance with SVM models of different emotion classes.
10. Six emotions and a neutral state is classified depending on the output of the classifier.

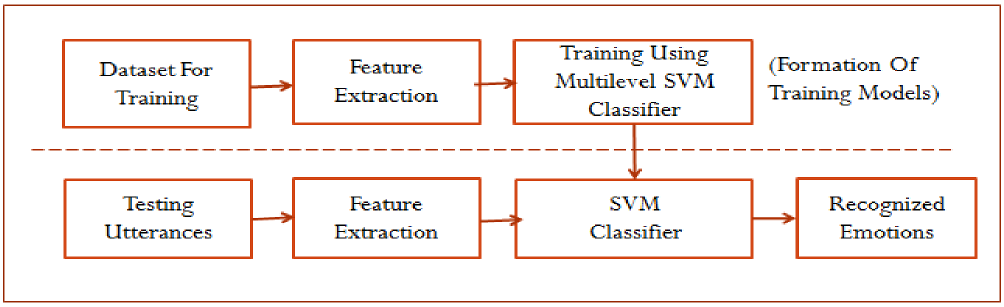


Fig.1 Emotion Recognition without Gender Information.

EMOTION RECOGNITION WITH GENDER INFORMATION:

ALGORITHM:
1. Read the signal.
2. Signal is divided into frames.
3. For each frame, MFCC, Energy and median of standard deviation of pitch frequency are calculated.
4. Extracted features are trained and tested using binary SVM classifier.
5. If the speaker is male, then multiclass SVM classifier is trained with extracted features- Formants, MFCC, Power Spectral Density of male speakers only.
6. If the speaker is female, then multiclass SVM classifier is trained with extracted features- Formants, MFCC, Power Spectral Density of female speakers only.
7. Test Signal is obtained and features are extracted.
8. Test data is classified in accordance with SVM models of different emotion classes which are gender dependent.
9. Six emotions and a neutral state is classified depending on the output of the classifier.

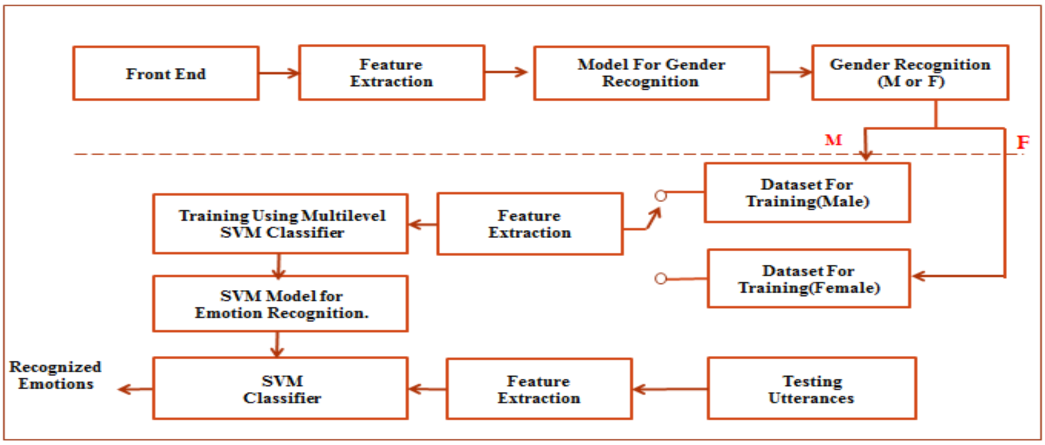

Front End
Feature Extraction
Model For Gender Recognition
Gender Recognition (M or F)
M
F
Dataset For Training(Male)
Dataset For Training(Female)
Feature Extraction
Training Using Multilevel SVM Classifier
SVM Model for Emotion Recognition.
Recognized Emotions
SVM Classifier
Feature Extraction
Testing Utterances

Fig.2 Emotion Recognition with Gender Information.

GENDER RECOGNITION (GR) SUBSYSTEM

GENDER RECOGNITION ALGORITHM USING PITCH:

The proposed GR method is designed to distinguish a male from a female speaker and has been thought to be realized over mobile devices, such as smartphones. It is designed to operate in an open-set scenario and is based on audio pitch estimation. In a nutshell: it is based on the fact that pitch values of male speakers are on average lower than pitch values of female speakers because male vocal folds are longer and thicker compared to female ones. In addition, being male and female pitch frequency separated, we realized that satisfying results in terms of accuracy of the GR can be obtained by using a single feature threshold classifier rather than more complex and time-consuming ones.

Speech signal exhibits a relative periodicity and its fundamental frequency, called pitch (frequency), is usually the lowest frequency component. A method based on the signal autocorrelation has been chosen because of its good applicability to voice and ease of implementation. In particular, given a real-value discrete-time signal, s(n); n ∈ [1….. N],$R(\boldsymbol{\tau})=s(n)s(n+\tau)$; $\tau€[0,1,\dots\dots N-1]$, R(τ) is the autocorrelation of lag τ.

Due to physiological reasons, is contained in a limited range of [P1,P2] (typically P1 = 50 [Hz] and P2 = 500 [Hz]) and limits the τ range between τ1and τ2 whereτ1=[Fs/P1] and τ1=[Fs/P2].Fs is the sampling frequency applied to the original analog signal to obtain the discrete-time signal s(n).

The autocorrelation function shows how much the signal correlates with itself, at different delays τ. Considering that, given a “sufficiently periodical" speech recording, its autocorrelation will present the highest value at delays corresponding to multiples of pitch periods defining the pitch period as $\boldsymbol{\tau}_{pitch}=\arg_{\boldsymbol{\tau}}\max\ R(\boldsymbol{\tau})$. The frequency of pitch is computed as$\rho_{pitch}= Fs/\boldsymbol{\tau}_{pitch.}$ From experimental tests, the employed threshold $\boldsymbol{\gamma}$ has been estimated to be 160 [Hz].

Short time energy of signals for each frame is calculated and samples with energy threshold above 0.6 is considered to be voiced. After the filtering samples with pitch value above the defined threshold is classified as females whereas those with pitch values below the threshold are classified as males. Pitch can be found by both time domain and frequency domain methods. Here auto correlation and cross correlation functions are used. For periodic signals, auto-correlation works as a mathematical tool for periodic sequence detector. Compared to auto-correlation, cross-correlation is much accurate for pitch detection as it is less affected by rapid variations in signal amplitude.

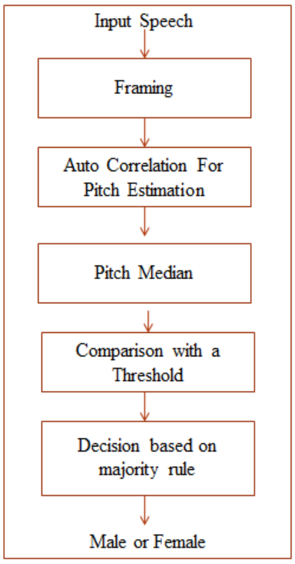

Input Speech
Framing
Auto Correlation For Pitch Estimation
Pitch Median
Comparison with a Threshold
Decision based on majority rule
Male or Female

Fig.3 Block diagram of auto correlation pitch detector.

SYSTEM MODEL FOR GENDER RECOGNITION USING SVM

ALGORITHM:

1. Read the signal.
2. Signal is divided into frames.
3. For each frame, MFCC, Energy and median of standard deviation of pitch frequency are calculated.
4. Extracted features are trained and tested using binary SVM classifier.
5. If the speaker is male, then multiclass SVM classifier is trained with extracted features- Formants, MFCC, Power Spectral Density of male speakers only.

6. If the speaker is female, then multiclass SVM classifier is trained with extracted features- Formants, MFCC, Power Spectral Density of female speakers only.
7. Test Signal is obtained and features are extracted.
8. Test data is classified in accordance with SVM models of different emotion classes which are gender dependent.
9. Six emotions and a neutral state is classified depending on the output of the classifier.

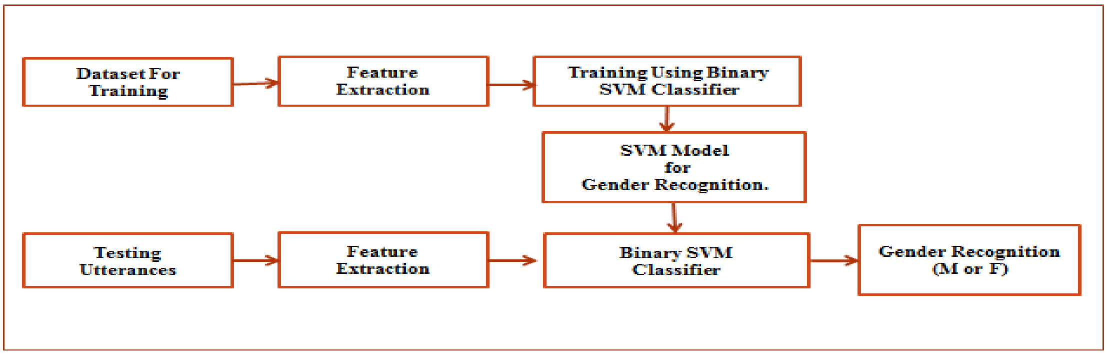

Dataset For Training
Feature Extraction
Training Using Binary SVM Classifier
SVM Model for Gender Recognition.
Testing Utterances
Feature Extraction
Binary SVM Classifier
Gender Recognition (M or F)

Fig.4Block Diagram Of Gender Recognition Using SVM

SVM is a simple and efficient computation of machine learning algorithms, and is widely used for pattern recognition and classification problems, and under the conditions of limited training data, it can have a very good classification performance compared to other classifiers. The idea behind the SVM is to transform the original input set to a high dimensional feature space by using kernel function. Therefore non-linear problems can be solved by doing this transformation. A support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification. A good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

SYSTEM MODEL FOR GENDER RECOGNITION USING ANN

ALGORITHM:
1. Read the signal.
2. Signal is divided into frames.
3. For each frame, MFCC, Energy and median of standard deviation of pitch frequency are calculated.
4. Extracted features are trained and simulated using Artificial Neural Network.
5. If the speaker is male, then multiclass SVM classifier is trained with extracted features- Formants, MFCC, Power Spectral Density of male speakers only.
6. If the speaker is female, then multiclass SVM classifier is trained with extracted features- Formants, MFCC, Power Spectral Density of female speakers only.
7. Test Signal is obtained and features are extracted.
8. Test data is classified in accordance with SVM models of different emotion classes which are gender dependent.
9. Six emotions and a neutral state is classified depending on the output of the classifier.

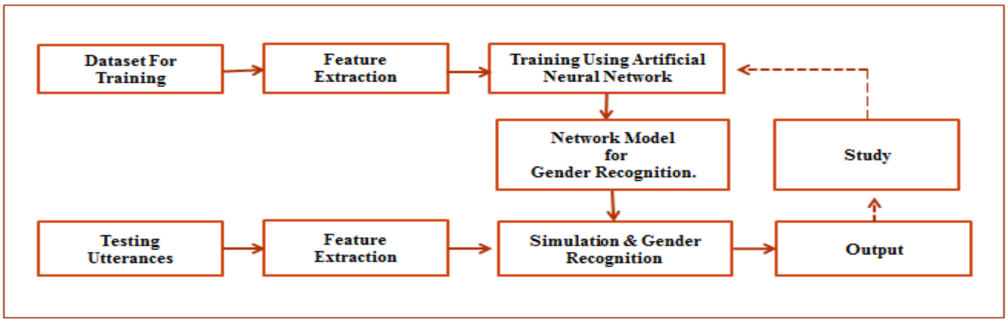
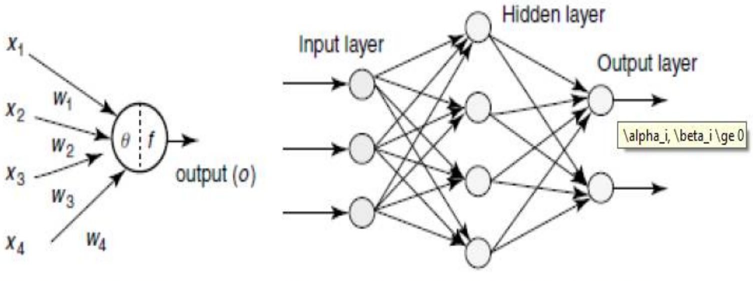

Dataset For Training
Feature Extraction
Training Using Artificial Neural Network
Network Model for Gender Recognition.
Study
Testing Utterances
Feature Extraction
Simulation & Gender Recognition
Output

Fig.5 Block Diagram Of Gender Recognition Using ANN.

Neural network is a computing system made up of a number of simple but highly interconnected processing elements, which process information by their dynamic state response to external inputs.Those highly interconnected processing elements are called neurons and belong to different layer. The main aim of the classification ANNs is to produce an exact output based on the input parameters.Basically neural network consists of three layers namely; input layer, hidden layer and output layer.



(a) Artificial neuron (b) Multilayered artificial neural network

The input to the neural network are given. 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 {[I1max,I1min], [I2max,I2min], [I3max,I3min]}. Figure shows the structure of neural network used in the proposed method with 3 input variables and two output variables. The neuron output signal O is given by the following relationship:.O=f(net)=f($\sum_{j=1}^{n}$ wjxj) where wj is the weight vector, and the function f(net) is referred to as an activation (transfer) function. The variable net is defined as a scalar product of the weight and input vectors and net=$w^Tx=w_1x_1+\ldots\ldots+w_nx_n$, where T is the transpose of a matrix, and, in the simplest case, the output value O is computed asO = f(net)= 1 if $w^Tx\geqslant\theta$ and 0 otherwise where θ is called the threshold level; and this type of node is called a linear threshold unit.

EMOTION RECOGNITION (ER) SUBSYSTEM

The implemented Emotion Recognition (ER) subsystem is based on two inputs: the features extracted by the Feature Extraction Block , in particular the sub-set ER of featuresneeded for the emotion recognition and the recognized speaker gender provided by the GR subsystem. Differently from the GR subsystem in which the employed feature has been individuated (the Pitch), concerning the ER subsystem the selection of feature(s) to be employed is still an open issue. The main features selected are pitch, MFCC and formants.

FORMANTS: In speech processing, formants are the resonancefrequencies of the vocal tract. The estimation of their frequency and their -3 [dB] bandwidth is fundamental for the analysis of the human speech as they are meaningful features able to distinguish the vowel sounds. In this paper, we employ a typical method to compute them, which is based on the Linear Predictive Coding (LPC) analysis. In more detail, the speech signal s(n) is re-sampled at twice the value Fmax = 5:5 [Hz], which is the maximum frequency applied within the algorithm to search formants. Then, a pre-emphasis filter is applied. The signal is divided in audio frames (0.05 [s] long in the case of formant extraction) and a Gaussian window is applied to eachframe. After that LPC Coefficients are computed by using the Burg method. Being Zi, the i-th complex root pair of the prediction (LPC) polynomial, the frequency, called $\boldsymbol{\gamma}_i$, and the -3 [dB] bandwidth, indicated with Δ_i, of the i-th formant relatedto the i-th complex root pair of the LPC polynomial, can be estimated by applying the following formulae$\boldsymbol{\gamma}_i$=(Fs/2Π)θ_i and Δ_i=-(Fs/Π)lnr_i. The algorithm finds all the formants in the range [0-max][Hz]. Some artifacts of the LPC algorithm can produce ``false" formants near 0 and Fmax [Hz] therefore the formants below 50 and over (Fmax -50[Hz]) are removed.

MEL-FREQUENCY CEPSTRUM COEFFICIENTS:The most prevalent and dominant method used to extract spectral features is calculating Mel- Frequency Cepstral Coefficients (MFCC). MFCCs are one of the most popular feature extraction techniques used in speech recognition based on frequency domain using the Mel scale which is based on the human ear scale. MFCCs being considered as frequency domain features are much more accurate than time domain features. Mel-Frequency Cepstral Coefficients (MFCC) is a representation of the real cepstral of a windowed short-time signal derived from the Fast Fourier Transform (FFT) of that signal. The difference from the real cepstral is that a nonlinear frequency scale is used, which approximates the behaviour of the auditory system. Additionally, these coefficients are robust and reliable to variations according to speakers and recording conditions. MFCC is an audio feature extraction technique which extracts parameters from the speech similar to ones that are used by humans for hearing speech, while at the same time, deemphasizes all other information and accurately represent the envelope.

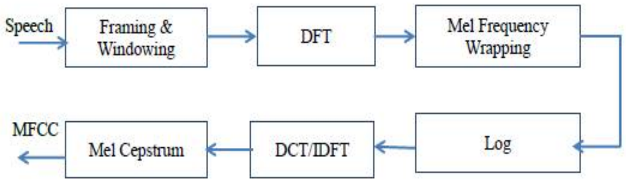
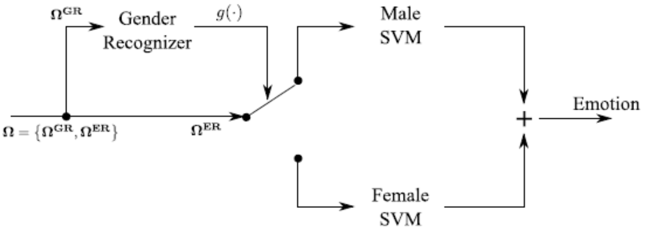


Fig.6 Block Diagram Of Gender Recognition Using ANN.

EMOTIONS CLASSIFIERS

Usually, in the literature of the field, a Support Vector Machine (SVM) is used to classify sentences. SVM is a relatively new machine learning algorithm introduced by Vapnik and derived from statistical learning theory in the 90s. The main idea is to transform the original input set into a high dimensional feature space by using a kernel function and, then, to achieve optimum classification in this new feature space, where a clear separation among features obtained by the optimal placement of a separation hyper plane under the precondition of linear separability. Differently from the previously proposed approaches, two different classifiers, both kernel-based Support Vector Machines (SVMs), have been employed in this paper.



The first one (called Male-SVM) is used if a male speaker is recognized by the Gender Recognition block. The other SVM (Female-SVM) is employed in case of female speaker. Male-SVM and Female-SVM classifiers have been trained by using speech signals of the employed reference Database (DB) generated, respectively, by male and female speakers. 70 % of the database is used for training and 15 % is used for testing and 15% for validation respectively. The two SVMs have been trained by the traditional Quadratic Programming (QP).

V. PERFORMANCE EVALUATION AND RESULTS

GENDER RECOGNITION

Gender recognition subsystem is implemented using three methods. Thresholding method using Pitch as a feature, Gender Recognition using SVM classifier and ANN.

RESULTS USING SVM :Gender classification by training and testing process using binary SVM classifier using Pitch obtained by Normalized Cross Correlation (NCCF) and MFCC on 10-fold Cross-Validation with an accuracy of 92.66%.

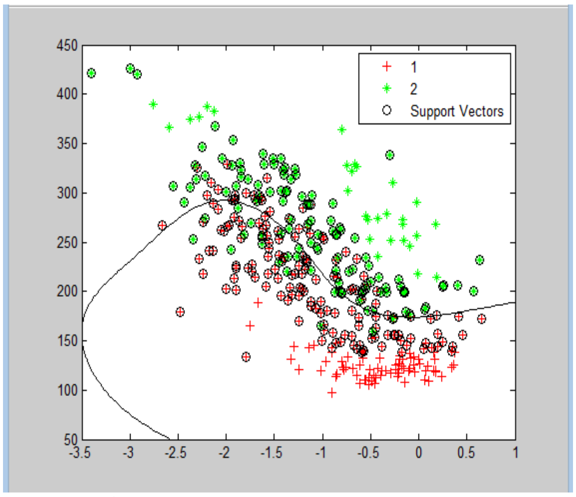


Fig.7 Binary Classification Using RBF Kernel For Gender Recognition.

RESULTS USING ANN :Performance shows mean square error dynamics for all the datasets in logarithmic scale. Training MSE is always decreasing, so its validation and test MSE you should be noted.

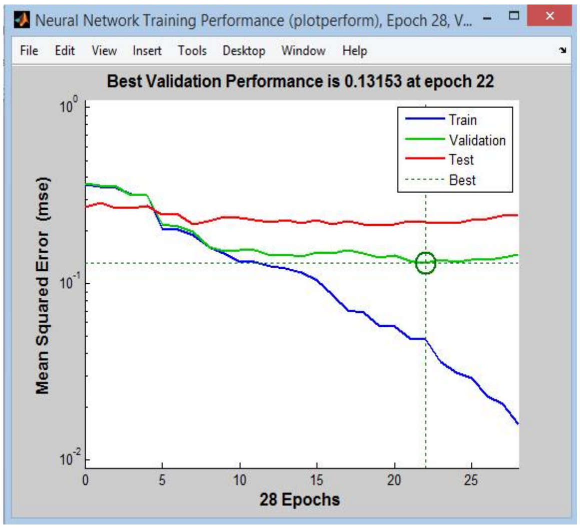


Fig.8 Performance in terms of Mean Square Error

From the figure, best validation performance is obtained at 22nd epoch with a gradient of 0.264. During training, the progress is constantly updated in the training window. Of most interest are the performance, the magnitude of the gradient of performance and the number of validation checks. The magnitude of the gradient and the number of validation checks are used to terminate the training. The gradient will become very small as the training reaches a minimum of the performance.

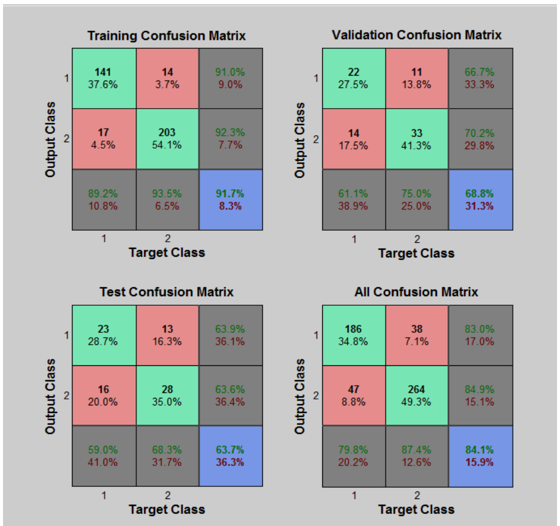
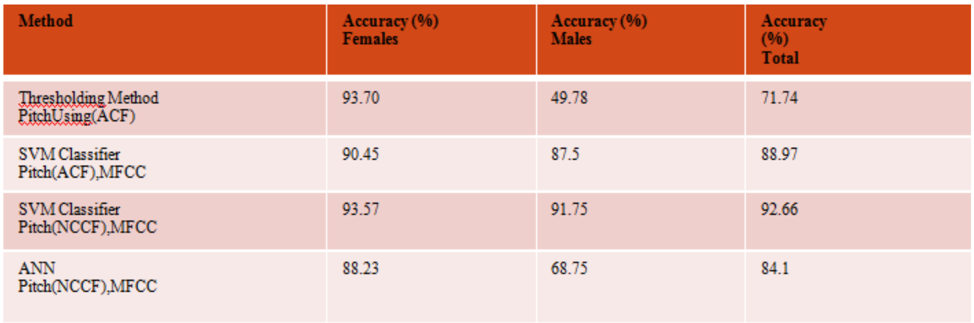


Fig.9 Performance in terms of Confusion matrix

The confusion matrix shows an overall rate of 84.1% accuracy.Green cells represent correct answers and red cells represent all types of incorrect answers.



Method	Accuracy (%) Females	Accuracy (%) Males	Accuracy (%) Total
Thresholding Method PitchUsing(ACF)	93.70	49.78	71.74
SVM Classifier Pitch(ACF),MFCC	90.45	87.5	88.97
SVM Classifier Pitch(NCCF),MFCC	93.57	91.75	92.66
ANN Pitch(NCCF),MFCC	88.23	68.75	84.1

Fig.10 Comparison Of Various Gender Recognition Methods.

Among all the gender recognition methods, the classifier SVM using Pitch calculated using normalized cross-correlation and MFCC exhibits a much better accuracy of about 92.66%.

EMOTION RECOGNITION: WITHOUT GENDER INFORMATION

In this subsection, the accuracy of a traditional approach, without having any ``a priori" information on the gender of the speaker, is shown. In this case, a single SVM has been trained with both male and female speeches. In more detail, the SVM has been trained and tested, considering the overall BES signals, by the k-fold cross-validation approach. The original BES signals are randomly partitioned into k equal size subsets. Among the k subsets, a single subset is retained to test the SVM, and the other k - 1 subsets are employed to train it. The cross-validation process is then repeated k times, with each of the k subsets used once as validation set. The obtained k resultsare then averaged to produce a single result. In this paper, in all considered cases, k = 10 has been employed and an accuracy of 74.28% is obtained.

EMOTION RECOGNITION: WITH GENDER INFORMATION

Evaluated the system performance when the "a priori'' information on the gender of the speaker is used. This information has been obtained by exploiting, in the testing phase, the Gender Recognition. Two SVMs one for each gender, have been trained: the first SVM through male speeches signals, the second through female ones. Also in this case, SVM training and testing phases have been carried out by two k-fold (k = 10) cross validations and, again, the overall BES signals have been employed by dividing male speech from female speech. The system shows an improved accuracy of 82.85 % when gender recognition was done using artificial neural networks and a much better accuracy of 88.57 % when gender recognition was done using support vector machine.

VI. CONCLUSION

The proposed system, able to recognize the emotional state of a person starting from audio signals registrations, is composed of two functional blocks: Gender Recognition (GR) and Emotion Recognition (ER). The former has been implemented by 3 methods, a Pitch Frequency Estimation method and thresholding to perform gender recognition, gender recognition using SVM and ANN, the latter by two Support Vector Machine (SVM) classifiers (fed by properly selected audio features), which exploit the GR subsystem output.The performance analysis shows the accuracy obtainedwith the adopted emotion recognition system in terms of recognition rate and the percentage of correctly recognized emotional contents. The experimental results highlight that theGender Recognition (GR) subsystem allows increasing the overall emotion recognition accuracy from 74.28 % to 88.57% due to the a priori knowledge of the speaker gender.

REFERENCES

[1] Igor Bisio, Alessandro Delfino, Fabio Avagetto, Mariomarchese, and Andrea Sciarrone “Gender-Driven Emotion Recognition Through Speech Signals for Ambient Intelligence Applications” corresponding author: i. Bisio.

[2] M. El Ayadi, M. S. Kamel, and F. Karray, ``Survey on speech emotion recognition: Features, classification schemes, and databases,'' Pattern Recognition., vol. 44, no. 3, pp. 572_587, 2011.

[3] H. Ting, Y. Yingchun, and W. Zhaohui, ``Combining MFCC and pitch to enhance the performance of the gender recognition,'' in Proc. 8th Int. Conf. Signal Process., vol. 1. 2006.

[4] D. Deepawale and R. Bachu, ``Energy estimation between adjacent formant frequencies to identify speaker's gender,'' in Proc. 5th Int. Conf. ITNG 2008, pp. 772_776.

[5] D. Gerhard, ``Pitch extraction and fundamental frequency: History and current techniques,'' Dept. Comput. Sci., Univ. Regina, Regina, SK, Canada,Tech. Rep., 2003.

[6] R. Snell and F. Milinazzo, ``Formant location from LPC analysis data,'' IEEE Trans. Speech Audio Process., vol. 1, no. 2, pp. 129_134, Apr. 1993.

[7] Fan Yingle, Yi Li and Tong Qinye “Speaker Gender Identification Based on Combining Linear and Nonlinear Features” Proceedings of the 7th World Congress on Intelligent Control and Automation.

[8] Yixiong Pan, PeipeiShen and LipingShen“Speech Emotion Recognition Using Support Vector Machine” Department of Computer Technology International Journal of Smart Home Vol. 6, No. 2, April, 2012.

[9] Y.-L. Shue and M. Iseli, ``The role of voice source measures on automatic gender classification,'' in Proc. IEEE ICASSP, Mar./Apr. 2008, pp. 4493_4496.

[10] G H Patel College of Engineering, Gujarat Technology University, India, `` Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition ',, Volume 1, Issue VI, July 2013

[11] O. Pierre-Yves, ``The production and recognition of emotions in speech: Features and algorithms,'' Int. J. Human-Comput. Stud., vol. 59, no. 1, pp. 157_183, 2003.

[12] Xia Mao, Lijiang Chen, Liqin Fu, “Multi-Level Speech Emotion Recognition based on HMM and ANN” 2009 World Congress on Computer Science and Information Engineering.

[13] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, ``A database of German emotional speech,'' in Proc. Interspeech, 2005,pp. 1517_1520.

[14] Aamir Khan, Muhammad Farhan, Asar Ali “Speech Recognition: Increasing Efficiency Support Vector Machine”. International Journal of Computer Applications (0975 – 8887) Volume 35– No.7, December 2011

[15] BhoomikaPanda ,DebanandaPadhi, Kshamamayee Dash, Prof. SanghamitraMohanty, “Use of SVM Classifier & MFCC in Speech Emotion Recognition System”. Volume 2, Issue 3, March 2012ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering.

[16] Yakun Hu, Dapeng Wu, and Antonio Nucci, “Pitch-based Gender Identification with Two-stage Classification” ,Department of Electrical and Computer Engineering, University of Florida, Gainesville,FL 32611. Correspondence author: Prof. Dapeng Wu, u@ece.ufl.edu, http://www.wu.ece.ufl.edu. Antonio Nucci is with Narus, Inc., 570 Maude Court, Sunnyvale, CA 94085.

[17] Kunjithapatham Meena1, KulumaniSubramaniam, and MuthusamyGomathyShrimathi “Gender Classification in Speech Recognition using Fuzzy Logic and Neural Network”. The International Arab Journal of Information Technology, Vol. 10, No. 5, September 2013.

[18] www.expressivespeech.net

[19] R. Rojas: Neural Networks, Springer-Verlag, Berlin, 1996

[20] MathWorks India

BIOGRAPHY

NishaChandran is currently pursuing M.Tech in Electronics with specialisation in Signal Processing from College Of Engineering, Kallooppara (CUSAT), Kerala, India. She received her B.Tech degree in Electronics and Communication from College Of Engineering,Kidangoor (CUSAT), Kerala, India. Her areas of interest are Speech Signal Processing, Digital Image Processing, Digital Communication, Instrumentation and Control Engineering.

Mahesh.B.S is currently working as Assistant Professor in College Of Engineering, Kallooppara, Kerala, India. He received hisM.Tech in Electronics with specialisation in Wireless Technology from Toch Institute Of Science And Technology, Kerala, India and B.Tech degree from College of Engineering,Cherthala, Kerala, India. Her areas of interest are Digital Image Processing, Biometrics, Digital Communication, Wavelet, Wireless Technology and Embedded Design.

DOI:10.15662/IJAREEIE.2016.0506002