ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 9, September 2018

Methodology for Gender Identification and Age Recognition through Multiple Indian Language Speech Signal

Manisha [1], Vinay Jain [2]

M. Tech Scholar, Communication System, Shri Shankaracharya Technical Campus (SSGI), Bhilai, CG, India[1]

Associate Professor, Department of Electrical & Electronics Shri Shankaracharya Technical Campus (SSGI), Bhilai, CG, India[2]

ABSTRACT: The human voice is comprised of sound made by a human being using the vocal cord for talking, singing, laughing, crying and shouting. It is particularly a piece of human sound creation in which the vocal cord is the essential sound source, which plays an important role in the conversation. The applications of speech or voice processing technology play a crucial role in human computer interaction. The system improves gender identification, age group classification, age and emotion recognition performance. The research work uses new and efficient methods for feature extraction of speech or voice and classification of standard method on the various audio datasets. Mel Frequency Cepstral Coefficients feature extraction and selection is performed to find a more suitable feature set for building speaker models. The proposed system uses Gaussian Mixture Model is a super vector for system feature selection and feature modelling. Support Vector Machine classification and feature matching technique is used to classify the feature for different age groups like child, teenage, young, adult and senior to increase the resultant performance and accuracy for multiple languages. The database is created using the audio files for each age group of speakers and for each emotion as an input, performs feature extraction and identifies the gender, classify age group, and recognize age and emotion.

KEYWORDS: Mel Frequency Cepstral Coefficient (MFCC), Gaussian Mixture Model (GMM), Expectation-Maximization (EM), Maximum a Posteriori (MAP), Hidden Markov Models (HMMs), Suprasegmentally Hidden Markov Models (SPHMMs), Interactive Voice Response System IVRs).

I. INTRODUCTION

Human interaction with computers in done in many ways and the interface between human and the computer is crucial to facilitate this interaction. Maximum desktop applications, internet using browsers like Firefox, chrome and internet explorer. The computers make use of the prevalent Graphical User Interfaces (GUI). Voice User Interfaces (VUI) are used for speech recognition and synthesizing systems. Human Computer Interaction (HCI) aims to improve the interface between users and computers by making computers more usable and receptive to users need. There are many speaker characteristics that have useful applications. The most popular include gender, age, health, language, dialect, accent, socialist, idiolect, emotional state and attention state. These characteristics have many applications in dialogue systems, speech synthesis, forensic, call routing, speech translation, language learning, assessment systems, speaker recognition, meeting browser, law enforcement, human robot interaction and smart workspaces. For example, the spoken dialogue system provides services in the domains of finance, travel, scheduling, tutoring. The systems need to gather information from the user automatically in order to provide timely and relevant services.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 9, September 2018

II. LITERATURE REVIEW

In presents a dimension reduction technique which aims to improve greater efficiency and the accuracy of speaker's age group and precise age estimation systems based on the human voice signal. Two different genders-based age estimation approaches studied, the first is the age group (senior, adult, and young) classification and the second is an accurate age estimation using regression technique. These two approaches use the GMM super vectors as features for a classifier model. Age group classification assigns an age group to the speaker and age regression estimates the speaker's precise age in years. In paper [5] presents gender detection is an extremely useful task for an extensive variety of voice or speech-based applications. In the spoken language systems INESC ID, the gender identification component is initial and the basic component of our voice processing system, where it is utilized prior to speaker clustering, in order to avoid mixing speakers between male and female gender in the same cluster. Gender information (male or female) is also used to create gender dependent acoustic module for speech recognition.

In introduce new gender detection and an age estimation approach is proposed. To develop this method, after deciding an acoustic features model for all speakers of the sample database, Gaussian mixture weights are extricated and connected to build a super vector for each speaker. Then, hybrid architecture of General Regression Neural Network (GRNN) and Weighted Supervised Non-Negative Matrix Factorization (WSNMF) are developed using the created super vectors of the training data set. The hybrid method is used to detect the gender speaker while testing and to estimate their age. Different biometric features can be used for forensic identification. Choosing a method depends on its use and efficient reliability of a particular application and the available data type. In some crime cases, the available evidence or proof might be in the form of recorded voice. Speech patterns can include unique and important information for law enforcement personnel.

In mainly focused on enhancing emotion recognition and identification performance based on two stages that is combination of gender recognizer and emotion recognizer. The system work is a gender dependent, text independent and speaker independent emotion recognizer. Both Hidden Markov Model (HMM) and Supra segmental Hidden Markov Model (SPHMM) have used as classifiers in the two-stage architecture. This architecture has been evaluated on two different and separate speech databases. The two databases are emotional prosody speech and transcripts database and human voice collected database.

In explores the detection of specific type emotions using discourse information and language in combination with acoustic signal features of emotion in speech signals. The main focus is on a detecting type of emotions using spoken language data obtained from a call center application. Most previous work in type emotion recognition has used only the acoustic features information contained in the speech. The system contains three sources of information, lexical, acoustic and discourse is used for speaker's emotion recognition.

In develop models for detecting various characteristics of a speaker based on spoken the text alone. These characteristics or attributes include whether the speaker is speaking native language, the speakers age and gender, the regional information reported by the speakers. The research explores various lexical features information as well as features inspired by linguistic (a language related) information and a number of word and dictionary of affect in language. This system suggests that when audio or voice data is not available, by exploring effective audio feature sets only from uttered text and system combinations of multiple classification algorithms, researcher build statistical models to detect these attributes of speakers, equivalent to frameworks that can explore the audio information.

In present speaker characteristic recognition and identification field has made extensive use of speaker MAP adaptation techniques. The adaptation allows speaker model feature parameters to be estimated using less speech data than needed for Maximum Likelihood (ML) training method. The Maximum Likelihood Linear Regression (MLLR) and Maximum a Posteriori (MAP) techniques have typically been used for speaker model adaptation. Recently, these adaptation techniques have been incorporated into the feature extraction stage of the SVM classifier-based speaker identification and recognition systems. In [15] humans, emotional speech recognition contributes much to create harmonious human to machine interaction, additionally with many potential applications. Three approaches to augment parallel classifier are compared for recognizing emotions from a speech by the speech database. Classifier applied on prosody, spectral, MFCC and other common features. One is standard classification schemes (one versus one) and two methods are directed a cyclic Graph (DAG) and Unbalanced Decision Tree (UDT) that can form a binary decision tree classifier. The hierarchical classification technique of feature driven hierarchical SVMs classifiers is

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 9, September 2018

designed, it uses different feature parameters to drive each layer and the emotion can be sub divided layer by layer. Finally, analysis of the classification rate of those three extends binary classification, DAG system performs the best for testing database and standard classifier is not far behind, the UDT is the poorest because of relying on upper layer order processing. In [16] the extraction and matching process is implemented after the signal pre-processing is performed. No parametric method for modelling the human voice processing System. The nonlinear sequence alignment called as Dynamic Time Warping (DTW) used as features matching techniques. This paper presents the technique of MFCC feature extraction and wrapping technique to compare the test patterns.

III. PROBLEM STATEMENT

This helps to identify the gender, then classify the speaker age group belong to a certain category, then further system will process to recognize exact age and also system display emotional state of the speaker with his/her profile detail which is stored in the database. The objective of the system is to extract the feature and compare with database to identify the gender and also it classifies the certain speaker age group, these two tasks helps to get increase the system performance and accuracy. On the other side, perform feature selection for speaker classification and matching using popular classification techniques, so efficient classifier classifying speaker characteristics. This system applies techniques like MFCC feature extraction algorithm, GMM modelling technique, SVM classifier and matching technique. The main issue in voice or speech processing research to achieve high efficiency and performance of different age group and different language dependent speaker and to reduce the large size of the dimension of feature matrix using many techniques.

IV. METHODOLOGY

Feature Extraction

The extraction of the best parametric representation of the acoustic signals of the human voice is an important task to produce a letter recognition performance. The result efficiency of feature extraction phase is important for the next phase like modelling, classification and feature matching since it affects its behaviour.

A. Gaussian Mixture Model

A GMM model is a probability density function represented using a weighted sum of all Gaussian component densities. Modelling technique is commonly used parametric model of the probability distribution of features in a proposed system, such as voice tract related spectral features of signal in a speaker recognition system. The parameters are estimated from training sample voice data using the iterative EM algorithm or MAP estimation from a well-trained prior modelling approach is a well-known modelling technique in text independent speaker recognition systems for frame-based features.

Each component density is a variant Gaussian function of the form, with mean vector and covariance matrix, the complete Gaussian mixture model is parameterized by the covariance matrices, mixture weights and mean vectors from all component densities.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 9, September 2018

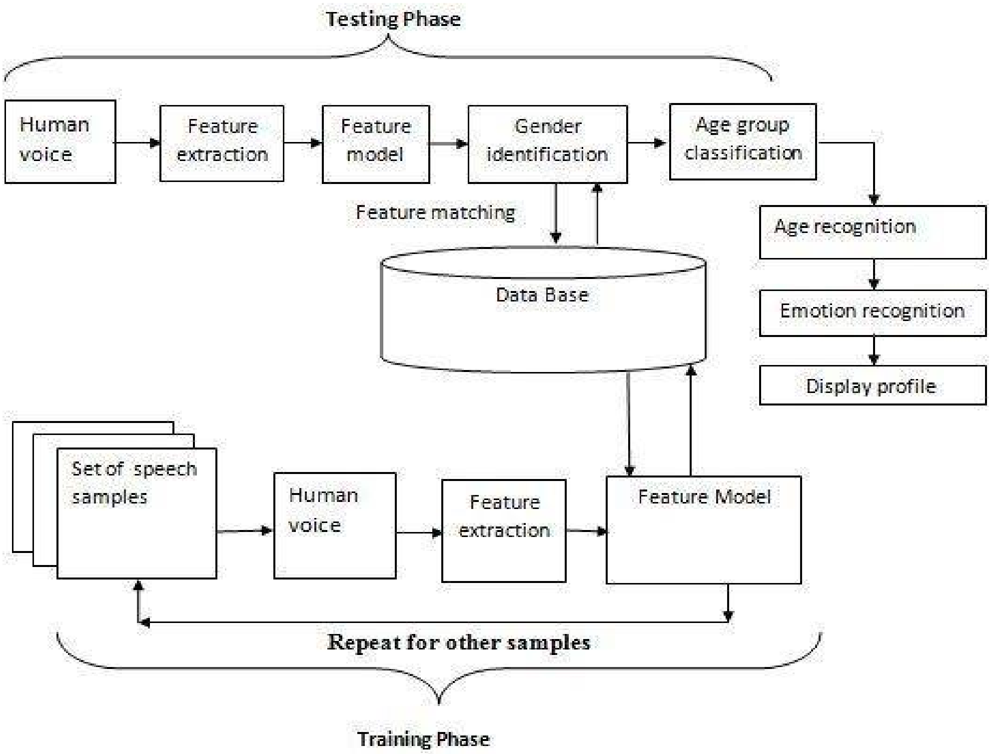

Figure1: Age and Gender detection block diagram

B. Hidden Markov Model (HMM)

In the Markov model each state corresponds to one observable event. But this model is too restrictive, for a large number of observations the size of the model explodes, and the case where the range of observations is continuous is not covered at all. The Hidden Markov concept extends the model by decoupling the observation sequence and the state sequence. For each state a probability distribution is defined that specifies how likely every observation symbol is to be generated in that particular state. As each state can now in principle generate each observation symbol it is no longer possible to see which state sequence generated a observation sequence as was the case for Markov models, the states are now hidden, hence the name of the mode.

V. CONCLUSION

Thus, the proposed system helps to identify, classify and recognize exact speaker age with emotion and displaying profiles of speaker using the trained database. The speaker profile is helpful in many applications like for advertisement, targeting to particular people, automatically identification of this feature, age, emotion to provide facility and service to customer in a call centre, in some field speaker's voice can be used as the biometric security because each human has a unique voice pattern and unique feature. The result in the feasible way to increase the accuracy and efficiency of system output.
The future enhancement of the system can be extended to recognize for more complicated noise sample (.wav file). The health condition of the speaker can also identify separate the individual speaker classification and age also possible to detect for mix mode gender speaker.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 9, September 2018

VI. RESULT

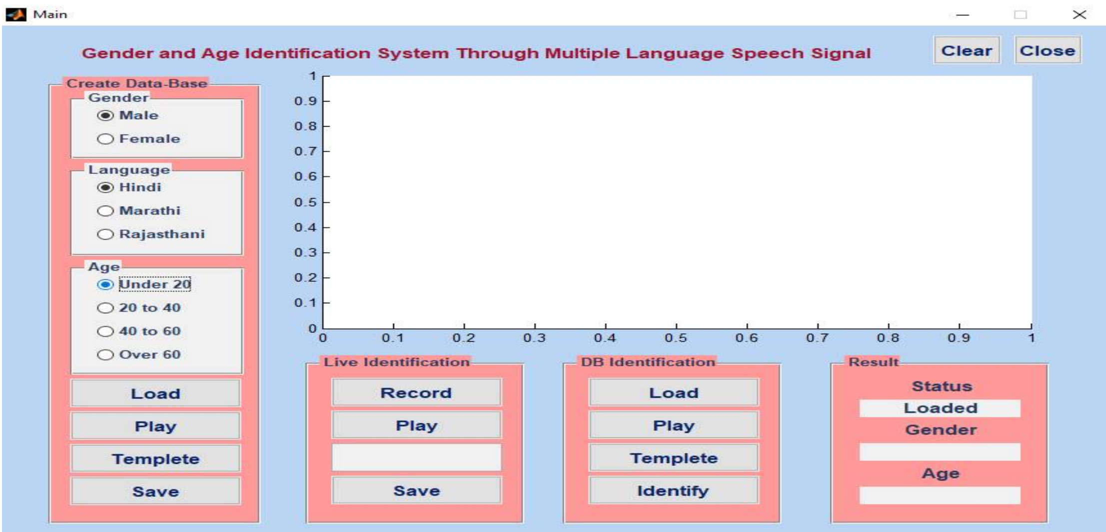

Figure2: Main Project GUI figure

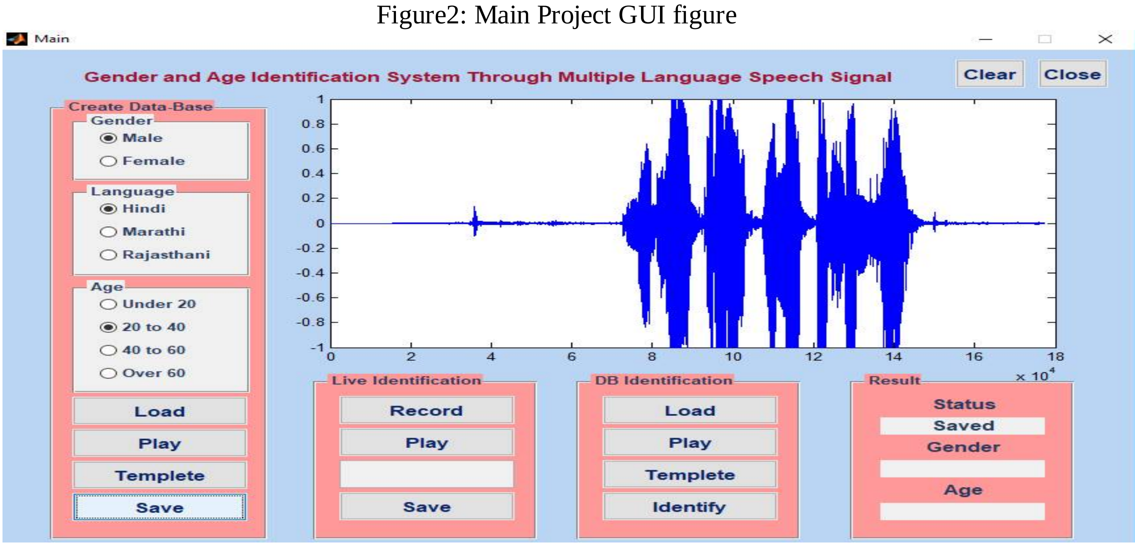

Figure3: Add Record into Database

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 9, September 2018

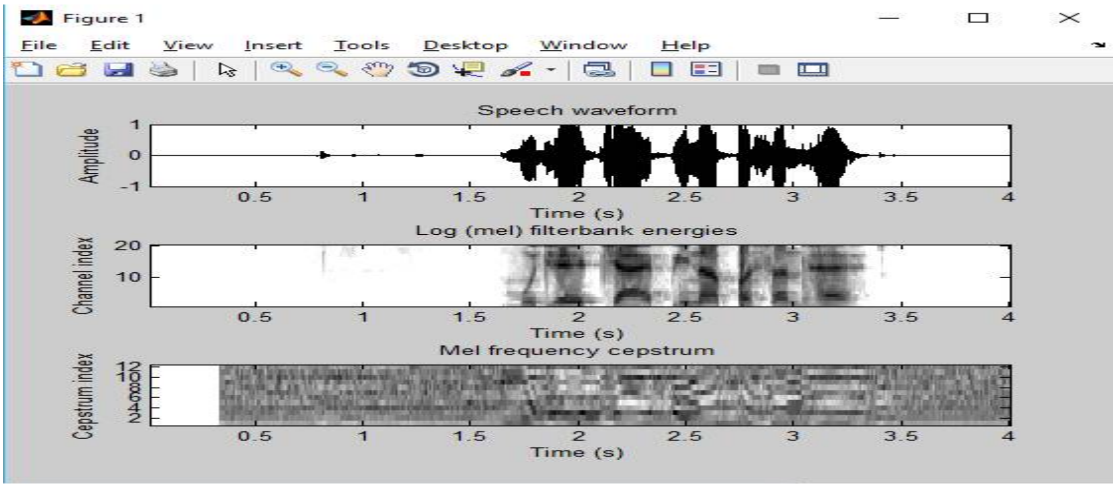

Figure4: Voice Parameters

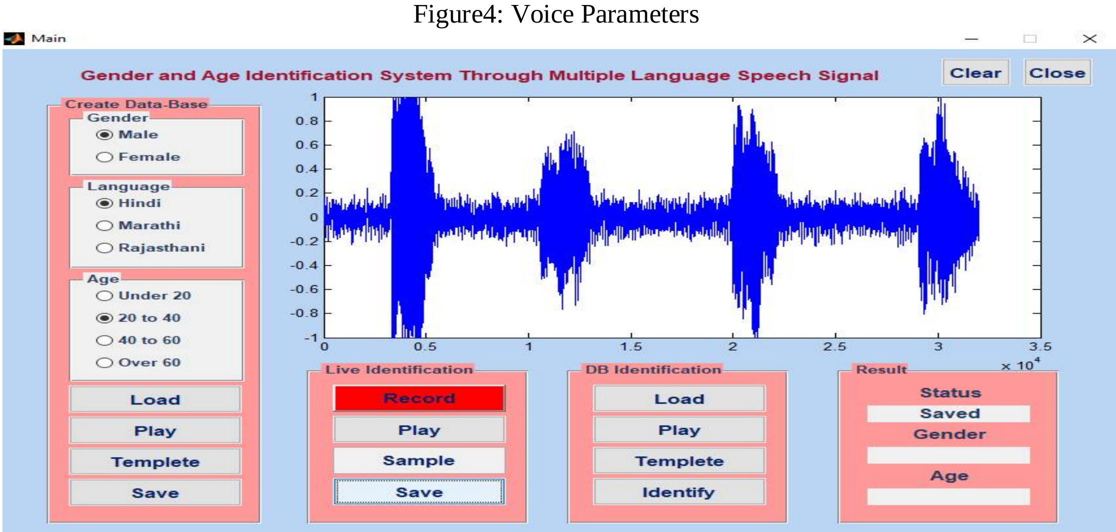

Figure5: Recording Sample by GUI

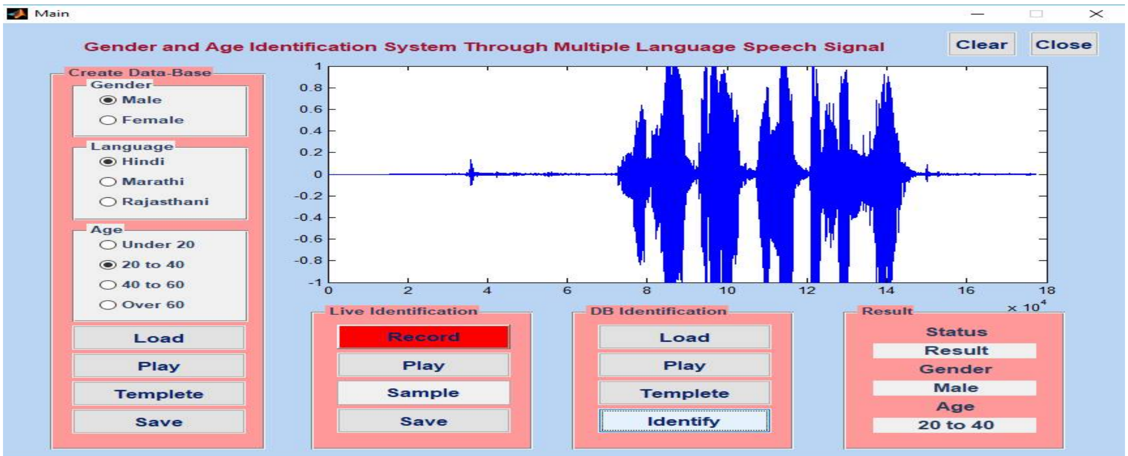

Figure6: Final Result

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 9, September 2018

REFERENCES

[1] Gil Dory, Ron M. Hecht, Mireille Avital and Yaniv Z, SEPTEMBER, 2011. Supervector Dimension Reduction for Efficient Speaker Age Estimation Based on the Acoustic Speech Signalise Transaction V.19, NO. 7.
[2] Hugo Meinedo1 and Isabel Trancosa, 2008Age and Gender Classification using Fusion of Acoustic and Prosodic Features, Spoken Language Systems Lab, INESC-ID Lisboan, Portugal, Instituto Superior Technicon, Lisboan, Portugal.
[3] Ismail Mohr Adnan Shahin, 2013Gender-dependent emotion recognition based on HMMs and Spamming J Speech Techno, Springer 16:133141.
[4] Mohamad Hasan Bihari and Hugo Van h, ITN2008 Speaker Age Estimation and Gender Detection Supervised Nonnegative Matrix Factorization, Centre for Processing Speech and Images Belgium.
[5] Shivaji J Chaudhari and Ramesh M Kayaker, May 2015 Automatic Gender Dependent Emotion Recognition, International Journal of Computer Applications(IJCA) (0975 - 8887), Volume 117 No. 17.
[6] Shivaji J. Chaudhari and Ramesh M. Kagalkar, July 2015 A Methodology for Efficient Gender Dependent Speaker Age and Emotion Identification System, International Journal of Advanced Research in Computer and Communication Engineering(IJARCCE) ISSN 2319-5940, Volume 4, Issue 7.
[7] Chul Min Lee and Srikanth S. Narayanan, 2005 Toward Detecting Emotions in Spoken Dialogs, IEEE transaction 1063-6676.
[8] Tetsuya Takiguchi and Yasuo Ariki, 2006 Robust feature extraction using kernel Compartment of Computer and System Engg Kobe University, Japan, ICASSP 14244-0469.
[9] Michael Feld, Felix Burkhardt and Christian Muller, 2010 Automatic Speaker Age and Gender Recognition in the Car for Tailoring Dialog and Mobile Services,German Research Center for Artificial Intelligence, INTERSPEECH.
[10] M A. Hossan, Sheeraz Memon and Mark A Gregory, A Novel Approach for MFCC Feature extraction, RMIT university, Melbourne, Australia, IEEE, 2010.
[11] Ruben Solera-Urea, 2008 Real-time Robust Automatic Speech Recognition Using Compact Support Vector Machines, TEC 2008-06382 and TEC 2008-02473.
[12] Wei HAN and Cheong fat CHAN, 2006 An Efficient MFCC Extraction Method in Speech Recognition, Department of Electronic Engineering, The Chinese University of Hong Kong Hong Kong, 78039390-06/IEEE ISCAS.
[13] AU Khan and L. P. Bhaiya, 2008 Text Dependent Method for Person Identification through Voice Segmenting- 2277-1956 IJECSE.