

Emotion and Gender Recognition using Fuzzy Support Vector Machine Through Speech Signals

B. Priya¹, E. Surya²

Department of Computer Science and Engineering
Rajalakshmi Engineering College,
Chennai

Abstract:- The temporal structure of the speech data is useful in analyzing the user emotions and gender. A system is proposed which recognizes the person's emotions and gender through audio signals. By identifying the emotions an application can be built to provide service in accordance to the user emotions. The system is composed of two sub-systems: 1) Gender Recognition and 2) Emotion Recognition. Independent component analysis (ICA) algorithm is used for emotion and gender recognition. Independent component analysis (ICA) is a recently developed method in which the goal is to find a linear representation of non-gaussian data so that the components are statistically independent, or as independent as possible. For classification Fuzzy Support vector machine (FSVM) is used. The proposed classification method enhances the Support Vector Machine (SVM) in reducing the effect of outliers and noises in data points. Hence the provided methods allows interaction among humans and computers and allows an effective human-computer intelligent interaction.

Index Terms—SVM, classification, ICA, linear, recognition, FSVM.

I. INTRODUCTION

Speech processing is the study of speech signals and the processing methods of the signals. The signals are usually processed in a digital representation, so speech processing can be regarded as a special case of digital signal processing, applied to speech signal. Aspects of speech processing include

the acquisition, manipulation, storage, transfer and output of speech signals. From the speech signals the emotions and gender of a person can be determined. The actual user emotion may help a system track the user's behavior by adapting to the inner mental state. Generally recognition of emotions is in the scope of research in the human-machine-interaction.

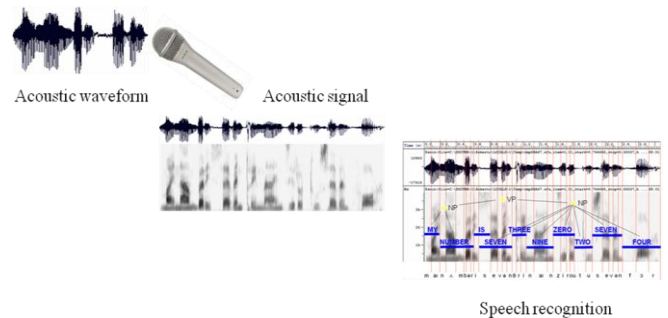


Fig : 1 Speech Recognition

Among other modalities like mimic speech is one of the most promising and established modalities for the recognition. There are several emotional hints carried within the speech signal. Attempts in detecting emotional speech analyze in general signal characteristics like pitch, energy, duration or spectral distortions. However, on semantically higher levels emotional clues can also be found. Speech emotion analysis refers to the use of various methods to analyze vocal behavior as a marker of affect (e.g., emotions, moods, and stress), focusing on the nonverbal aspects of speech. The basic assumption is that there is a set of

objectively measurable voice parameters that reflect the affective state a person is currently experiencing (or expressing for strategic purposes in social interaction).

II RELATED WORKS

Recognizing a person's emotional state starting from audio signal registrations. The system is able to recognize six emotions (anger, boredom, disgust, fear, happiness, and sadness) and the neutral state. This set of emotional states is widely used for emotion recognition purposes. It also distinguishes a single emotion versus all the other possible ones, as proven in the proposed numerical results. The system is composed of two subsystems: 1) gender recognition (GR) and 2) emotion recognition (ER). For classification of gender and emotions Support Vector Machine (SVM) is used. Principal Component Analysis (PCA) is used for recognition system. The experimental analysis shows the performance in terms of accuracy of the proposed ER system[1]. Iterative Feature Normalization (IFN) framework, an unsupervised

front-end, especially designed for emotion detection. The externalization of emotion is intrinsically speaker-dependent. A robust emotion recognition system should be able to compensate for these differences across speakers. A natural approach is to normalize the features before training the classifiers. However, the normalization scheme should not affect the acoustic differences between emotional classes. The IFN approach aims to reduce the acoustic differences, between the neutral speech across speakers, while preserving the inter-emotional variability in expressive speech.[4] proposed a systematic approach for recognition of human emotional state from audiovisual signals is explored. Machine recognition of human emotional state is an important component for efficient human-computer interaction. The majority of existing works address this problem by utilizing audio signals alone, or visual information only. The audio characteristics of emotional speech are represented by the extracted prosodic, Mel-

frequency Cepstral Coefficient (MFCC), and formant frequency features. A face detection scheme based on HSV color model is used to detect the face from the background.[5] proposed the role of acoustic measures related to the voice source in automatic gender classification, implemented using Support Vector Machines (SVMs). Differences of physiological properties of the glottis and the vocal tract are partly due to age and/or gender differences. Since these differences are reflected in the speech signal, acoustic measures related to those properties can be helpful for automatic age and gender classification. Acoustic measures of the vocal tract and the voice source were extracted from 3880 utterances spoken by 205 male and 160 female talkers (aged 8 to 39 years old). Formant frequencies and formant bandwidths were used as vocal tract measures, and open quotient and source spectral tilt correlates were used as voice source measures. Results show that the addition of voice source measures can help improve automatic gender classification results for most age groups. Acoustic measures from both the voice source and the vocal tract are used for automatic gender classification of 8 to 39 year old talkers.[2]

III.SYSTEM DESIGN

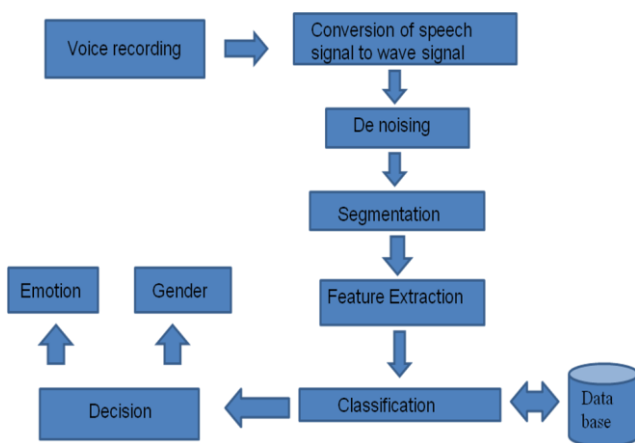


Fig : 2 System Architecture

Fig 2 shows the system architecture in which the speech is recorded into the system and is converted into wave signal. De noising of the wave signal is done by using a Haar wavelet filter. The signal is then segmented into number of samples using Independent Component Analysis (ICA) algorithm. Linear Discriminant Analysis (LDA) extracts the individual signals from mixtures. The extracted signals are to be classified by using a classification technique. Classification of emotions and gender is done by using two Fuzzy Support Vector Machines (FSVM).The gender and emotions are recognized by the system.

IV.IMPLEMENTATION WORK

A. Speech processing

In speech processing human speech is recorded to the system. The recorded speech signal has to be converted to a wave signal using Discrete Wavelet Transform (DWT) which will sample the signals discretely. De noising has to be done which is the process of removing noise from the wave signal. A Haar wavelet filter is used to remove the noise signal.

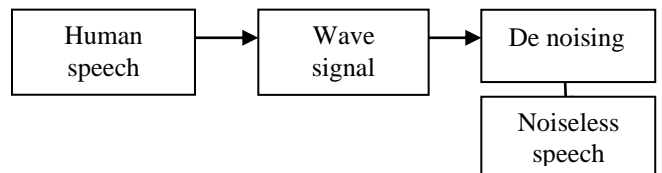


Fig : 3 Speech Processing

The human speech is recorded to the system through a microphone. It should be both male and female voices so that the gender and emotions of a person can be recognized. The recorded speech is converted to wave signal and a graph is plotted with respect to time and amplitude. The graph is plotted with the values generated. The wave signal is finally de-noised which will remove the background noises and a noiseless speech is obtained.

B. Segmentation And Feature Extraction

In segmentation and feature extraction the noise removed wave signal is segmented into number of samples. From the samples the features are to be extracted which will be effectively used for the gender and emotion recognition. For segmenting the samples Independent Component Analysis (ICA) algorithm is used which is used for revealing hidden factors that underlie sets of random variables, measurements or signals.

1) Independent Component Analysis

It is a recently developed method in which the goal is to find a linear representation of non-gaussian data so that the components are statistically independent, or as independent as possible. It is essentially a method for extracting individual signals from mixtures. Its power resides in the physical assumptions that the different physical processes generate unrelated signals. Independent Component Analysis is introduced as a widely used technique for solving

the noise and hidden source separation problem. ICA outputs can still contain strong residual components of the interfering speakers whenever noise or reverberation is high. In such cases, nonlinear post-processing can be applied to the ICA outputs, for the purpose of reducing remaining interferences.

2) Linear Discriminant Analysis

For feature extraction Linear Discriminant Analysis (LDA) is used which is essentially a method for extracting individual signals from mixtures. Linear Discriminant Analysis gives the discrimination within class and between class images. The LDA method constructs an optimal projection of the trained data by maximizing the ratio of the determinant of the between-class scatter matrix of the projected data to the within-class scatter matrix of the projected data.

C. Classification

Classification is done using Fuzzy Support Vector Machine (FSVM) and two Fuzzy Support Vector Machines are used for gender and emotion classification and the system recognizes the gender and emotion from the classified signals.

1) Fuzzy Support Vector Machine

FSVM-Fuzzy Support Vector Machine is a classifier that is used between human and computer interactive. The method of FSVM further improves the multi-classification method of supported vector machine. It deals with reasoning that is approximate rather than fixed and exact. Compared to traditional binary sets (where variables may take on true or false values), fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false. Furthermore, when linguistic variables are used, these degrees may be managed by specific functions. Fuzzy logic has been applied to many fields, from control theory to artificial intelligence.

V.EXPERIMENTAL RESULTS

Here are the experimental results of the above system done in MATLAB.

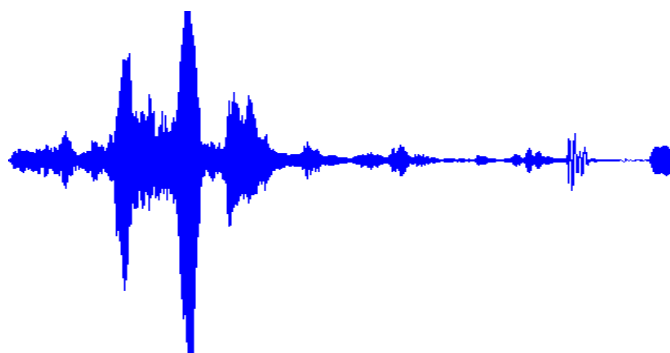


Fig : 4 Wave Signal

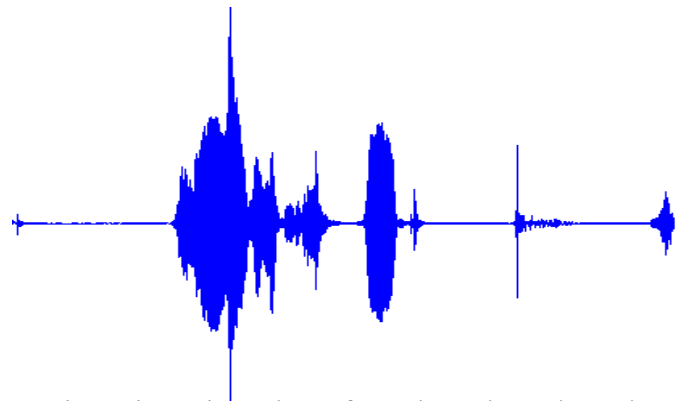


Fig : 5 De noised Signal

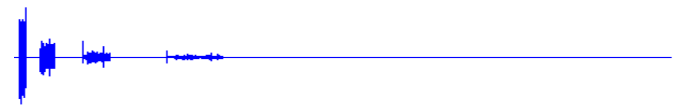


Fig : 6 Segmented Signal



Fig : 7 Featured Signal

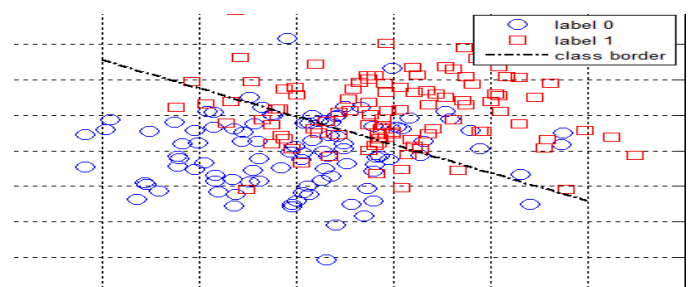


Fig : 7 Classification

VI. CONCLUSION

The system proposes the person's emotions and gender recognition system through audio signals. Initially the speech signal is converted to wave signal and is given as the input for De noising. A Haar Wavelet Filter is used for De noising. The De noised speech is segmented using an Independent Component Analysis (ICA) algorithm. The proposed methods enhances the SVM in reducing the effect of outliers and noises in data points and also allows interaction among humans and computers and allows an effective human-computer intelligent interaction.

VII. REFERENCES

- [1] Igor Bisio, Alessandro Delfino, Fabio Lavagetto, Mario Marchese, and Andrea Sciarra **“Gender-Driven Emotion Recognition Through Speech Signals for Ambient Intelligence Applications”** IEEE transactions on EMERGING TOPICS IN COMPUTING, 2014.
- [2] Yen-Liang Shue and Markus Iseli **“The Role Of Voice Source Measures On Automatic Gender Classification”** IEEE International Conference on SPEECH AND SIGNAL PROCESSING 2008.
- [3] Oscar T.-C. Chen, Jhen Jhan Gu, Ping-Tsung Lu and Jia-You Ke **“Emotion-Inspired Age and Gender Recognition Systems”** IEEE Conference on CIRCUITS AND SYSTEMS 2012.
- [4] Carlos Busso, Soroosh Mariooryad, Angeliki Metallinou, and Shrikanth Narayanan, **“Iterative Feature Normalization Scheme for Automatic Emotion Detection from Speech”** IEEE Transactions on AFFECTIVE COMPUTING 2014.
- [5] Yongjin Wang, and Ling Guan, **“Recognizing Human Emotional State From Audiovisual Signals”** IEEE Transactions on MULTIMEDIA 2008.
- [6] Andr'e Stuhlsatz, Christine Meyer, Florian Eyben, Thomas Zielke, G'unter Meier, Bj'orn Schuller **“Deep Neural Networks For Acoustic Emotion Recognition: Raising the Benchmarks”** IEEE International Conference on SPEECH AND SIGNAL PROCESSING rocessing 2011.
- [7] Carlos Busso Angeliki Metallinou and Shrikanth S. Narayanan **“Iterative Feature Normalization For Emotional Speech Detection”** IEEE International Conference on SPEECH AND SIGNAL PROCESSING 2011.
- [8] Konstantin Markovi, And Tomoko Matsui **“Music Genre and Emotion Recognition Using Gaussian Processes”** IEEE 2014.
- [9] Moataz M. H. El Ayadi, Mohamed S. Kamel, and Fakhri Karray **“Speech Emotion Recognition Using Gaussian Mixture Vector Autoregressive Models”** IEEE International Conference on SPEECH AND SIGNAL PROCESSING 2007.
- [10] Tauhidur Rahman and Carlos Busso **“A Personalized Emotion Recognition System Using An Unsupervised Feature Adaption Scheme”** IEEE International Conference on SPEECH AND SIGNAL PROCESSING 2012.
- [11] Stavros Ntalampiras and Nikos Fakotakis **“Modeling the Temporal Evolution of Acoustic Parameters for Speech Emotion Recognition”** IEEE Transactions on AFFECTIVE COMPUTING 2012.