International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 5, ISSN No. 2455-2143, Pages 163-165
Published Online March – April 2016 in IJEAST (http://www.ijeast.com)

IMPROVE HCII BY USING GENDER AND EMOTION RECOGNITION THROUGH SPEECH SIGNAL: A SURVEY

Prof. D. Varpe
Professor of Computer Engineering and Information Technology, PVG's College of Engineering and Technology, Pune, Maharashtra, India

Neha Sawant, Shweta Shah, Snehal Shivarkar
Department of Computer Engineering and Information Technology,
PVG's College of Engineering and Technology, Affiliated to Savitribai Phule Pune University,
Pune, Maharashtra, India

***Abstract*- A Human Computer Intelligent Interaction (HCII) is a field of science intended to provide natural ways for human to use computer as aid. HCII enable the world to communicate and interact effortlessly and intuitively. If machines have mental state of user it can interact more effectively with user. That is why, Human Computer Intelligent Interaction (HCII) plays important role. There are many techniques present to improve HCII. One such area is, gender and emotion recognition through speech signal. There are different techniques which implement gender and emotion recognition. This paper presents an analytical survey of recent published studies.**

Keywords- **HCII, Feature extraction, Gender Recognition, Emotion Recognition,**

I. INTRODUCTION

We have entered an era of enhanced digital connectivity. Now a day, in daily life computers are concerned in almost everything. People use technology to work, to communicate, to shop, to learn new information. It is believed that technology is embedding computers into homes, transportation means and working spaces. In future humans will be surrounded by intelligent, yet invisible computing devices that can anticipate in every need. That is why, it is greatly challenging issue to make the interaction between human and computer intelligent one (HCII).

HCII is approaching human computer interaction in different way, by moving their focus from computer centered designs to human centered designs. Human centered HCII has the ability to detect changes in communicative behavior of user. To achieve effective HCII computers should be able to interact naturally like human-human interactions. There are various technologies available to enhance HCII such as Image processing, Virtual Reality, Natural Language Processing,

Natural Language Processing involves recognizing gender and emotional state of user to enhance HCII. Emotion is medium of expression of mental state of oneself. Usually human recognize emotion of the speaker easily. They have achieved this intelligence by years of practice. The human mind learns from experience that is, it captures all kinds of emotions since childhood and learns to differentiate between the emotions based on its observations. Human recognize gender of speaker by observing appearance of speaker. By observing speaker's expressions and gestures human can detect emotional state of speaker's mind. For instance, when a person is sad, he speaks in very dull tone, his expressions changes to gloomy. Similarly when person is happy, he speaks in musical tone and there is look of delight in his face.

The survey of this study elaborates analysis of different techniques used for detection of gender and emotion of human. Main contribution to this paper concerns different possible solutions for gender and emotion detection. This survey paper is composed of i) Feature extraction ii) Gender detection iii) Emotion detection. It briefly explains various techniques available for gender and emotion detection through speech signal such as MFCC, autocorrelation method and different classification methods. It also includes figure analysis and summarizing sections.

II. EXISTING SYSTEMS FOR FEATURE EXTRACTION

Paper titled "Feature Extraction from Speech Data for Emotion Recognition" authored by S. Demircan and H. Kahramanlı published in 2014 describes which features are best suited for gender and emotion recognition. It implements extraction of Mel Frequency Cepstral Coefficients (MFCC)

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 5, ISSN No. 2455-2143, Pages 163-165
Published Online March – April 2016 in IJEAST (http://www.ijeast.com)

from the signals and classification with k-NN algorithm to recognize emotion. Statistics of Mel-Frequency Cepstral Coefficients which are computed over three phoneme type classes of interest are stressed vowels, unstressed vowels and consonants in the speech. The paper clearly indicates that indeed both the richer set of spectral features and the differentiation between phoneme type classes are helpful for the task. The method uses short time log frequency power coefficients (LFPC) to represent a speech signals and a discrete hidden Markov model (HMM) as the classifier. Performance of the LFPC feature parameters is compared with performance of linear prediction Cepstral coefficients (LPCC) and mel-frequency Cepstral coefficients (MFCC) feature used in speech recognition systems. This paper suggests that LFPC is a better choice as feature parameters for gender and emotion classification.

Paper titled " Improved MFCC-Based Feature for Robust Speaker Identification" ,authored by WU Zunjing , CAO Zhigang, published in 2005 describes, the Mel-frequency cepstral coefficient (MFCC) is the most widely used feature in speaker recognition. It presents a method which combines robust representations in the feature space sample and speech enhancement technique in signal space where the important aim for feature space is to extract the acoustic features of the input speech. The paper also shows that standard MFCC feature analysis is very successful except when the noise is present. The log function in MFCC is sensitive to noise, so it replaced the logarithmic transformation in the MFCC analysis by a combined function to improve the noise.

III. EXISTING SYSTEMS FOR GENDER DETECTION

The paper titled "Gender classification by pitch analysis" ,authored by BhagyaLaxmi Jena & Beda Prakash Panigrahi studies about developing a gender classifier using speech signal. It mainly concentrates on pitch analysis of speech signal for gender classification. It consist analysis of pitch values of male and female voice samples. It implements pitch determination through autocorrelation method. The paper states that there is sufficient difference between the average pitch value of male and female voice sample value obtained by auto-correlation method. The table1 shows the experimental average pitch values of male and female voice samples. It uses difference in pitch value to develop gender classifier by setting threshold pitch value for male and female voice samples.

Table1. Pitch comparison between male and female voice Sample

Female Voice	Average pitch (in Hz)	Male Voice	Average pitch (in Hz)
Female 1	236.89	Male 1	154.68
Female 2	273.7	Male 2	184.22
Female 3	310.01	Male 3	192.92
Female 4	258.98	Male 4	190.71
Female 5	178.40	Male 5	137.35

The paper titled, "Gender classification using pitch and formants" is presents information about gender detection using speech signals. It describes methods which are based on pitch, formants and combination of both. It briefly explains Pitch Detection Algorithms based on the autocorrelation function, the average magnitude difference function, cepstral analysis and formants extraction. The paper concludes that autocorrelation method shows better results for pitch estimation as compared to other techniques.

On the basis of above survey of gender detection techniques, it can be concluded that pitch calculation is very important factor in finding gender through speech signal. For more accuracy we can combine it with formant, MFCC.

IV. EXISTING SYSTEMS FOR EMOTION DETECTION

In this paper, titled "Study of speech emotion recognition method", authored by Aastha Joshi, Rajneet Kaur published in 2013, different emotion recognition methods are discussed for extracting audio features from speech sample. It also studies various classifier algorithms briefly. The paper states that Speech Emotion Recognition has a promising future and accuracy of the system depends upon the combination of features extracted, used classification algorithm and emotional speech database. This study provides a simple guide to carry out research study in the speech emotion recognition systems.

Paper titled "Recognizing emotion in speech" authored by Frank Dellaert, Thomas Polzin and Alex Waibel describes statistical pattern recognition techniques to classify speech signal according to their emotional content. They have conducted a small and informal experiment in order to assess how well a human does in classifying emotions. In this paper for classification they have used only pitch information extracted from speech. The paper compares classification methods such as Maximum Likelihood Bayes classifier (MLB), Kernel Regression (KR) and K-nearest neighbors (KNN) for emotion recognition. It studies feature selection methods like population hillclimbing and forward selection which gives better result with above mentioned classification methods. It interprets the features selected by these methods as the likely correlates of emotion.79

Paper titled, "Emotion and Gender Recognition of Speech Signals Using SVM", presents implementation of emotion detection through speech signal system. It describes two support vector machines (SVM'S) which are distinctly used for the male and female speaker's emotion recognition such as anger, fear, happiness, sad, neutral. It highlight that the knowledge of the speaker's gender allows a performance increase. The paper focuses on the evaluation of the system capability to recognize the single emotion with and without

Gender recognition (GR)**.** It also shows that the features selection technique gives a satisfying recognition rate and also allows reduction in employed features.

From above survey, it is clear that for emotion detection there are various classification technologies available like Maximum Likelihood Bayes classifier (MLB), Kernel Regression (KR) and K-nearest neighbors (KNN), Support Vector Machine (SVM). But SVM is more efficient among all as it has high accuracy.

V. PROPOSED SYSTEM

The proposed system composed of two functional blocks: First Gender Recognition (GR) and second Emotion Recognition (ER) which is able to recognize the gender and the emotional state of a person starting from audio signal. It implements a Pitch Frequency Estimation method and two Support Vector Machine (SVM) classifiers.
Output of Gender Recognition subsystem is given to Emotion Recognition subsystem as input. The Gender Recognition (GR) subsystem allows increasing the overall emotion recognition accuracy. It also implements a features selection algorithm which reduces number of employed features. Our system can be efficiently used to identify a single emotion, or emotion category, versus all the other possible ones.

VI. CONCLUSION

This survey gives brief overview about different algorithms which are used for gender and emotion recognition. Accuracy measures depend on features extracted, used classification algorithm and emotional speech database. From this survey it can be stated that there are many systems which implements either gender recognition or emotion recognition but few systems which implements both. If emotion detection system is gender driven the system becomes more efficient. If we provide detected gender as input to emotion detection subsystem along with other features the result obtained are more accurate. Still, the ongoing work presents a great hope for gender and emotion recognition.

VII. FUTURE SCOPE

It is suspected that the voice based gender and emotion detection system can be implemented over smart phone platforms. In future it can be used in smart home, smart office and virtual reality, and it may acquire importance in all aspects of future people's life.

VIII. REFERENCCES

[1] S. Demircan and H. Kahramanlı, "Feature Extraction from Speech Data for Emotion Recognition", 2014

[2] WU Zunjing , CAO Zhigang, "Improved MFCC-Based Feature for Robust Speaker Identification", 2005

[3] BhagyaLaxmi Jena & Beda Prakash Panigrahi, "Gender classification by pitch analysis", 2012

[4] Pawan Kumar, Nitika Jakhanwal, Anirban Bhowmick, and Mahesh Chandra, "Gender classification using pitch and formants", 2011

[5] Aastha Joshi, Rajneet Kaur, "Study of speech emotion recognition method", 2013

[6] Frank Dellaert, Thomas Polzin and Alex Waibel, "Recognizing emotion in speech"

[7] S.Sravan Kumar, T.RangaBabu, "Emotion and Gender Recognition of Speech Signals Using SVM",2015