Depression Level Assessment Based on Speech

Chaudhari Bhavana Jagdish[1] Akash Koli[2] More Chanchal Pandurang[3] Patel Pooja Chhotulal[4] Chaudhari Priyanka Kantilal[5]

[1,2,3,4,5]D.N.Patel College Engineering Shahada Tal Shahada Dist Nandurbar, India

Abstract— Depression is a serious mental health disorder that affects mood, thoughts, feelings, and the ability to function in everyday life. Depression is considered as a psychosomatic state associated with the soft biometric features. People who are suffering from depression always behave abnormal. The system accepts speech from the user and outputs the depression level at which he is. I-Vector technique is used to detect the depression level. It starts with removing silence from the speech and then extracts features. The fuzzy logic approach is used to describe the level of depression.

Key words: Depression Calculation, Noise Reduction

I. INTRODUCTION

Depression is a serious mental health disorder that affects mood, thoughts, feelings, and the ability to function in everyday life. Some of its characteristics are prolonged feelings of extreme sadness, guilt and hopelessness, and thoughts of death. Major depression is the leading cause of disability and is the cause of more than two-thirds of suicides each year. Depression is a common mental disorder that presents persistent feelings of sadness, intrusive negative thoughts and, cognitive difficulties such as poor concentration, leading to functional impairment. A range of acoustic features have already been identified for suitability in the classification of depression. Speech production cues such as pitch and formant measures are useful due in part to the effects of increased tension in the vocal tract associated with depression. Spectral and energy based measures are also useful in classifiers, as depressive speech can contain more information in the higher energy bands when compared with neutral speech. Spectral centroid based methods including the sub-band spectral centroid features have recently shown promise in other applications, and other work shows that these newer measures potentially include information useful in the classification of depression.

II. PROBLEM IDENTIFICATION

The medical field has risen in all aspects of computationally analysing the diseases. The psychologists all over the world treat their patients with verbal communication and manually check if the patient is depressed or sad. If he is depressed there is no measure to calculate it in percentage.

The psychologists put lot of effort in identifying the mental state of the patient by questioning certain things. The depression is then predicted on the basis of answer that the patient gives. The psychologists detect the depression after numerous counseling sessions.

III. SYSTEM ARCHITECTURE

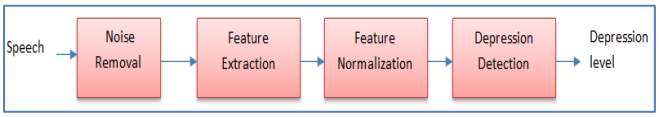

Fig. 1: Architecture Of System

In this system, the depression level of each speech signal using I-Vector technique. In the proposed approach first of all silence is removed from the speech signal then features are extracted from audio using I-Vector after that split overlapping function is applied to evaluate overlapped audio beats. In the end, depression is evaluated using relationship matrix. The system analyses the speech features like pitch and energy. On the basis of these features, the depression is evaluated.

IV. MODULES OF THE SYSTEM

- Silence Removal
- Silence Removal
- Depression Feature Extraction using I-Vector Approach
- Feature Normalization
- Level Calculation using Fuzzy Logic

V. PROCEDURE

- The user first speaks a sentence using mic.
- The speech recorded is pre-processed by removing noise and silence.
- The features such as pitch and energy are extracted.
- These features are used for assessing depression from the speech.
- Finally the depression level is given as output to the user.

VI. RELATED WORK

Daniel Povey et al[1]. describe the design of Kaldi, a free, open-source toolkit for speech recognition research. Kaldi provides a speech recognition system based on finite-state transducers (using the freely available OpenFst), together with detailed documentation and scripts for building complete recognition systems. Kaldi is written is C++, and the core library supports modeling of arbitrary phonetic context sizes, acoustic modeling with subspace Gaussian mixture models (SGMM) as well as standard Gaussian mixture models, together with all commonly used linear and affine transforms. Kaldi is released under the Apache License v2.0, which is highly nonrestrictive, making it suitable for a wide community of users.

Niels Rosenquist et al[2]. The etiology of depression has long been thought to include social environmental factors. To quantitatively explore the novel possibility of person-to-person spread and network-level determination of depressive symptoms, analyses were performed on a densely interconnected social network of 12 067 people assessed repeatedly over 32 years as part of the Framingham Heart Study. Longitudinal statistical models were used to examine whether depressive symptoms in one person were associated with similar scores in friends, coworkers, siblings, spouses and neighbors. Depressive symptoms were assessed using CES-D scores that were available for subjects in three waves measured between

1983 and 2001. Results showed both low and high CES-D scores (and classification as being depressed) in a given period were strongly correlated with such scores in one's friends and neighbors. This association extended up to three degrees of separation (to one's friends' friends' friends). Female friends appear to be especially influential in the spread of depression from one person to another.

Michel Valstar et al[3]. Mood disorders are inherently related to emotion. In particular, the behaviour of people suffering from mood disorders such as unipolar depression shows a strong temporal correlation with the affective dimensions valence and arousal. In addition, psychologists and psychiatrists base their evaluation of a patient's condition to a large extent on the observation of expressive facial and vocal cues, such as dampened facial expressions, avoiding eye contact, and using short sentences with flat intonation. It is in this context that we present the third Audio-Visual Emotion recognition Challenge (AVEC 2013). The challenge has two goals logically organised as sub-challenges: the first is to predict the continuous values of the affective dimensions valence and arousal at each moment in time. The second sub-challenge is to predict the value of a single depression indicator for each recording in the dataset. This paper presents the challenge guidelines, the common data used, and the performance of the baseline system on the two tasks.

Kua et al[4]. Most conventional features used in speaker recognition are based on spectral envelope characterizations such as Mel-scale filter bank cepstrum coefficients (MFCC), Linear Prediction Cepstrum Coefficient (LPCC) and Perceptual Linear Prediction. (PLP). The MFCC's success has seen it become a de facto standard feature for speaker recognition. Alternative features, that convey information other than the average subband energy, have been proposed, such as frequency modulation (FM) and subband spectral centroid features. In this study, we investigate the characterization of subband energy as a two dimensional feature, comprising Spectral Centroid Magnitude (SCM) and Spectral Centroid Frequency (SCF). Empirical experiments carried out on the NIST 2001 and NIST 2006 databases using SCF, SCM and their fusion suggests that the combination of SCM and SCF are somewhat more accurate compared with conventional MFCC, and that both fuse effectively with MFCCs. We also show that frame-averaged FM features are essentially centroid features, and provide an SCF implementation that improves on the speaker recognition performance of both subband spectral centroid and FM features.

ACKNOWLEDGEMENT

The Acknowledgement is just like a drop in the ocean of the deep sense of gratitude within our heart for people who helped us out of most embarrassing part of our life when we were standing at most difficult step towards our dream of life.

Many people have contributed to the success of this project work. Although a single sentence hardly suffices, we would like to thank Almighty God for blessing us with his grace. We extend our sincere and heartfelt thanks to Prof. V. S. Mahajan, Head of Department, Computer Engineering, for providing us the right ambience for carrying out this work.

We are grateful and sincerely appreciate the effort of our respected project in-charge Prof. R. A. Shaikh & Prof. D. B. Shukla & Prof. A. I. Pathan who acted as a fulcrum for us and supported us during the ups and downs of our project. We are profoundly indebted to our project guide Prof. A. B. Koli for innumerable acts of timely advice, encouragement and we sincerely express our gratitude to him.

We express our immense pleasure and thankfulness to all the teachers and staff of the Department of Computer Engineering and Information Technology for their cooperation and support.

REFERENCES

[1] D.Povey, A.Ghoshal, G.Boulianne, L.Burget, O.Glembek, N.Goel, M.Hannemann, P.Motlicek, Y.Qian, P.Schwarz, J.Silovsky, G.Stemmer, K.Vesely,"The Kaldi speech recognition toolkit" , in: IEEE Workshop on Automatic Speech Recognition and Understanding, IEEE Signal Processing Society, 2011.

[2] Niels Rosenquist, J., Fowler, J. & Christakis, N. "Social Network Determinants of Depression".Molecular Psy chiatry 16 (3): 273–281, (2011).

[3] M.Valstar, B.Schuller, K.Smith, F.Eyben, B.Jiang, S.Bilakhia, S.Schnieder, R.Cowie, M.Pantic, AVEC2013-"The continuous audio/visual emotion and depression recognition challenge", in: Proceedings of AVEC'13, 2013

[4] Kua, J. M. K., "Investigation of Spectral Centroid Magnitude and Frequency for Speaker Recognition." Proc. Odyssey: Speaker and Lang. Rec. Workshop, 2010, pp. 34 - 39.

