DETECTION OF SPEECH UNDER STRESS USING SPECTRAL ANALYSIS

Neha P. Dhole,

M.E. 2nd year (Digital Electronics), ***neha_p43@rediffmail.com***

GUIDE BY:

Professor: - Dr. AJAY A.GURJAR

Department of Electronics & Telecommunication Engineering,
Sipna's C.O.E. &T Sant Gadge Baba Amravati University, Amravati, Maharashtra, India

Abstract

This paper deals with an approach to the detection of speech in English language. The stress detection is necessary which provides real time information of state of mind of a person. Voice features from the speech signal is influenced by stress is MFCC is considered is this paper. To examine the effect of Exam-Stress on speech production an experiment was designed. First Year students of age group 18 to 20 were selected and assignment was given to them and instructs them that have viva on that assignment and their performance in the viva will decide their final internal marks in the examination. The experiment and the analysis of the test results are reported in this paper.

Keywords: *Speech, Stress, Spectral Analysis, Discrete Wavelet Transform, Artificial Neural Network.*

----------***----------

1. INTRODUCTION

Emotions have long been recognized to be an important aspect of human beings. More recently, psychologists have begun to explore the role of emotions as a positive component in human cognition and intelligence. Spoken language comes from our inside. Factors such as mood, emotion, physical characteristics and further pragmatic information are contained in speech signals. Many of these characteristics are also audible. An emotional speech with high content differs in some parameters from a neutral speech. In recent years, the interest for automatically detection and interpretation of emotions in speech has grown and vocal emotions have also tended to be studied in isolation. About 25% of information contained in a clean speech signal refers to the speaker. These linguistically irrelevant speaker characteristics make speech recognition less effective but can be used for speaker recognition (ca. 15% of information) and analysis of the speaker's emotional and health state (ca. 10% of information).

With increasing demand for speech technology systems, there is an increasing need for processing of emotion and other pragmatic effects (simulation in synthetic speech, elimination in robust speech recognition). In some cases, it is very important to detect the emotional state of a person (e.g., stress, fatigue or use of alcohol) from his/her voice.

2. METHODOLOGY USED

In this research work we have taken samples of the persons those who are the age of 18 to 20 year, and collecting all the samples of the speaker at the time of their viva examination before and after the examination for analysis of stress in speech. All the samples collected used as a database for the spectrum analysis of the speech under stress. In order to check the whether the speech have stress or not. For this checking we are using the neural network, and & the Matlab Tool box. The block diagram of method used is shown below.

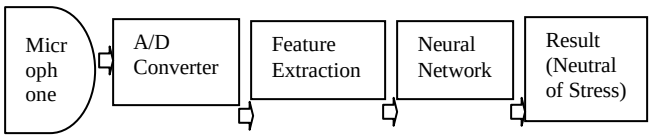

Fig. 1.1 Block diagram of proposed method

2.1 Discrete Wavelet Transform (DWT)-

Discrete Wavelet Transforms (DWT) is the process of transformation of a signal to the high frequency and low frequency components by using digital filtering techniques. In Discrete Wavelet Transform (DWT) we are taking into account only the low frequency components of the signal under consideration because low frequency components

characterize a signal more than its high frequency components.

2.2 ANFIS: Artificial Neuro Fuzzy Inference Systems

ANFIS uses two neural network and fuzzy logic approaches. When these two systems are combined, they may qualitatively and quantitatively achieve an appropriate result that will include either fuzzy intellect or calculative abilities of neural network. As other fuzzy systems the ANFIS rules, we may recognize five distinct layers in the structure of ANFIS network which makes it as a multi-layer network. A kind of this network, which is a Sugeno type fuzzy system with two inputs and one output, is indicated.

2.3 MFCC

Speech is usually segmented in frames of 20 to 30 ms, and the window analysis is shifted by 10 ms .each frame is converted to 12 MFCCs plus a normalized energy parameter. The first and second derivatives (A's and AA's) of MFCCs and energy are estimated, resulting in 39 numbers representing each frame. Assuming a sample rate of 8 kHz, for each 10 ms the feature extraction module delivers 39 numbers to the modeling stage. This operation with overlap among frames is equivalent to taking 80 speech samples without overlap and representing them by 39 numbers. In fact, assuming each speech sample is represented by one byte and each feature is represented by four bytes (float number), one can see that the parametric representation increases the number of bytes to represent 80 bytes of speech (to 136 bytes). If a sample rate of 16 kHz is assumed, the 39 parameters would represent 160 samples. For higher sample rates, it is intuitive that 39 parameters do not allow reconstructing the speech samples back. Anyway, one should notice that the goal here is not speech compression but using features suitable for speech recognition.

3. CLASSIFICATION AND RECOGNITION

Artificial neural network can learns from examples for a defined task, something which cannot be done using a conventional digital computer. Neural network is a complex pattern classifier composed of interconnected processing units called nodes, which can perform mathematical operations in a similar way as the human brain does . A neural network solves problems by self learning and self organization and is characterized by their topologies, activation function and weight vectors which are used in their hidden and output layers for processing simple mathematical operations. Neural networks can perform computations in a more effective way because of their massively parallel computational structure, fault tolerance, ability for generalization and inherently adaptive mechanism of learning.

4. RESULT

In our experiments, we used phonetically rich sentences from the Exam Stress corpus for our analysis of stressed speech. These sentences were automatically segmented into phoneme-like units. ANN does comparison of Neutral and Stress Sample. The results are shown below:

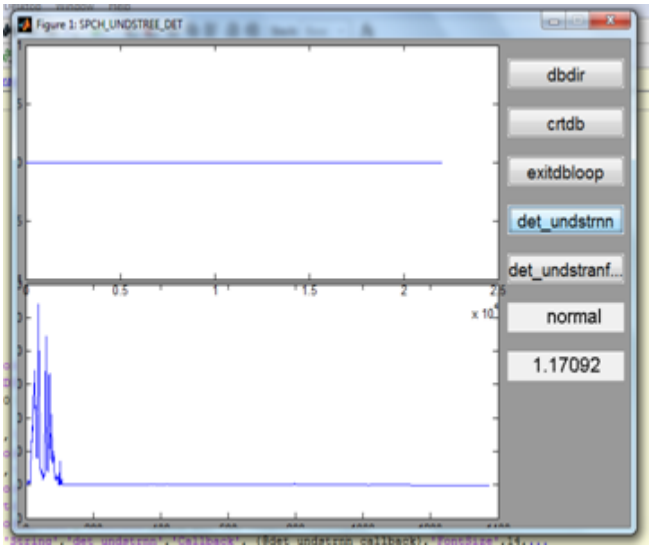

Fig. 5.1 For Normal Speech

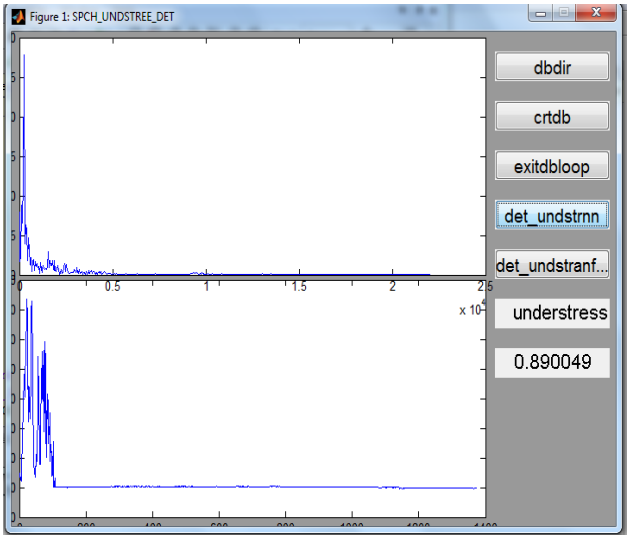

Fig.5.2 For Under Stressed Speech

CONCLUSIONS

The Spectral Analysis of speech signal is aimed at extracting spectral features such as MFCC Changes in spectrum of speech signal have shown to be a indicator of the internal emotional state of a person. In this research work, we have extracted these spectral features of some speakers in neutral condition and under stress condition. We have formed the feature matrix of the feature vectors obtained. For classification of the speech signal for stress Artificial Neural Network and ANFIS plays main role.. Thus, we could

conclude that spectral analysis is an efficient tool for detecting stress in speech.

REFERENCES

[1]. T. Johnstone and K. Scherer, "The effects of emotions on voice quality," Proceedings of 14th International Conference of Phonetic Science. San Francisco, pp. 2029-2032, 1999.
[2] D. Ververidis and C. Kotropoulos, "Emotional speech recognition: Resources, features, and methods," Speech Communication, vol. 48, No. 9, pp. 1162-1181, 2006.
[3] Analysis of Fundamental Frequency Contours in Speech, H. Levitt and L. R. Rabiner, Journ.Acoust. Soc. Amer., Vol. 49, No. 2, pp. 569-582, February 1971.
[4] Design of Digital Filter Banks for Speech Analysis, R. W. Schafer and L. R. Rabiner, Proc.of the Fifth Annual Princeton Conference on Information Sciences and Systems, pp. 40-47, March 1971.
[5] Investigation of Stress Patterns for Speech Synthesis by Rule, H. Levitt, L. R. Rabiner and A. E. Rosenberg, Journ. Acoust. Soc. Amer., Vol. 45, No. 1, pp. 92-101, January 1969.

BIOGRAPHIE:

Neha P. Dhole received a Bachelor degree in 2007 in Electronics & Telecommunication from Sant Gadge Baba Amravati university and appeared for M.E. (Digital Electronics) from Sant Gadge Baba Amravati university, her papers was published in namely following journals 1) "Automatic Induction Motor Starter with programmable Timmer", International journal of advance management technology & engineering, science 2011, No. 2249-7455. 2) "Reconstruction of Watermark from digital images", "Research link, 2011 ISSN-0973-1628"