Eth.J.Sci & Technol. 5(2):114-123, 2008 ISSN 1816-3378

Spectral Psychoanalysis of Speech under Strain

Durga Prasad Sharma [1], Prathap M[2] and Solomon Tadesse[3]
[1]*Professor, AMIT, under UN Development Program dp.shiv08@gmail.com*
[2]*Department of Computer Science and Engineering, Faculty of Engineering, Bahir Dar University, Ethiopia.*
[3]*Arbaminch University, Ethiopia*

Abstract
The non-verbal content of speech carries information about the physiological and psychological condition of the speaker. Psychological strain is a pathological element of this condition, of which one of the causes is accepted to be Exam-Strain. Objective, quantifiable correlates of strain are searched for by means of measuring the acoustic modifications of the voice brought about by Exam-Strain. Different voice features from the speech signal to be influenced by strain are: loudness, fundamental frequency, jitter, zero-crossing rate, speech rate and high-energy frequency ratio. To examine the effect of Exam-Strain on speech production an experiment was designed. Final Year students of age group 22 to 24 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 psychoanalysis of the test results are reported in this paper.

Keywords: Speech; Strain; Spectral Psychoanalysis; Quantifiable; Neural Network; pathological

1. Introduction
Feelings are the unique features in the creatures. Emotions have been long been recognized to be an important aspect of human beings. More recently, psychologists have begun to explore the function of emotions as a positive component in human cognition and intellect. Vocal language comes from our inside of feelings.

Feature such as emotions, mood, physical characteristics and further pragmatic information are contained in the speech signals. Many of these characteristics are also audible. An emotional speech with high content differs in some parameter from unbiased/neutral speech. In recent years, the interests for automatically detection and interpretation of emotions in speech have grown and vocal emotions have also tended to be studied in segregation. About 25-30% of information contents in clean speech signal refer to the speaker. These linguistically immaterial speaker characteristics make speech recognitions less effective but can be used for speaker recognitions and psychoanalysis of speaker's health state and emotional status.

With the increasing demand for voice based and speech technology systems, there is an increasing demand for processing of emotions and other pragmatic effects of human beings. In some cases, it is very important to detect the emotional state of person for an instance strain fatigue.

Standard speech may be regarded as speech made in a quiet room with no task obligations. Strain in speech, on the other hand, is a result of speech produced under emotional states, fatigue, environmental noise, heavy workload, and or sleep loss. Here some of the consequences of physiological strain are respiratory changes including increased respiration rate, irregular breathing and increased muscle tension of the vocal cords. These factors may result in irregular vocal fold movement and other vocal system modifications that ultimately affect the quality of the utterances. The presence of strain in speech causes changes in phoneme production with respect to glottal source factors, pitch, intensity, duration, and spectral shape. In linear acoustic theory,

speech production process is described in terms of source filter model. This model assumes plane wave propagation in the vocal tract and neglects nonlinear terms. Linear acoustic theory suggests that frequency in vocal tract filter; intensity and duration of glottal signal may be assumed to change due to strained speech production. In this paper, linear acoustic features and nonlinear features in frequency domain have been investigated in strain classification (Ververidis and Kotropoulos, 2006)

2. Speech Correlators of Strain

Features, which are usually applied for detecting the emotional strain, are:

Fundamental frequency: The fundamental tone, often referred to simply as the fundamental and abbreviated of, is the lowest frequency in a harmonic series. The fundamental frequency (*F*0) of a periodic signal is the inverse of its period, which may be defined as the smallest positive member of the infinite set of time shifts that leave the signal invariant. This definition applies strictly only to a perfectly periodic signal. The significance of defining the pitch period as the smallest repeating unit can be appreciated by nothing that two or more concatenated pitch periods from a repeating pattern in the signal. However, the concatenated signal unit obviously contains redundant information.

Formats: A format is a peak in the frequency spectrum of a sound caused by acoustic resonance. In phonetics, the word refers to sounds produced by the vocal tract. In acoustic, it refers to resonance in sound sources, notably musical instruments, as well as that of sound chambers. However, it is equally valid to talk about the format frequency of the air in a room, as exploited; Formats are the distinguishing or meaningful frequency components of human speech and of singing. By definition, the information that humans require to distinguish between vowels can be represented purely quantitatively by the frequency contents of the vowel sounds. Formats are the characteristics partials that identify vowels to the listener. Most of these formats are produced by the tube ad chamber resonance, but a few whistle tones derive form periodic collapse of venturi effects low-pressure zones. The format with the lowest frequency is called f1, the second f2, and the third f3. Most often the two first formants, f1 and f2, are enough to disambiguate the vowels. These two formats are primarily determined by the position of the tongue.

Duration: Duration is a property of a tone that becomes one of the bases rhythm. A tone may be sustained for varying lengths of time. For example, an event in the common sense has a duration greater than zero (but not very long), but in certain specialized senses (such as in the theory of relativity), a duration of zero. It is often cited as one of the fundamental aspects of music, see also rhythm. Durations, and their beginnings and endings, may be described as long, short, or taking a specific amount of time. Often duration is described according to terms borrowed from descriptions of pitch.

Zero Crossing: The zero-crossing rate is the rate of sign-changes along a signal, i.e., the rate at which the signal changes from positive to negative or back. This feature has been used heavily in both speech recognition and music information retrieval and is defined formally as:

$$zcr = \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{I}\{s_t s_{t-1} < 0\}$$

Equation (1)

Where S is a signal of length T and the indicator function $\mathbb{I}\{A\}$ is 1 if its argument A is true and 0 otherwise.

Power Spectral Density: Sinusoidal representation has been widely applied to speech modification, low bit rate speech and audio coding. Usually, speech signal is analyzed and synthesized using the overlap-add algorithm or the peak-picking algorithm. But the overlap-add algorithm is well known for high computational complexity and the peak-picking algorithm cannot track the transient and syllabic variation well. Peaks are picked in the curve of power spectral density for speech signal; the frequencies corresponding to these peaks are arranged according to the descending orders of their corresponding power spectral densities. These frequencies are regarded as the candidate frequencies to determine the corresponding amplitudes and initial phases according to the least mean square error criterion (Levitt and Rabiner, 1971; Johnstone and Scherer, 1999).

3. Methodology

In this research work the samples have been taken of the persons those who are in the age of 22 to 24 year and collecting all the samples of the speaker at the time of their viva examination before and after the examination for psychoanalysis of strain in speech. All the samples collected used as a database for the spectrum psychoanalysis of the speech under strain. In order to check the whether the speech have strain or not. For this checking the Multilayer Back Propagation Algorithm & the Matlab Tool box is used. The block diagram of method used is shown below (Rule, 1969; Schafer and Rabiner, 1971).

Neuron Model (logsig, tansig, purelin)
An elementary neuron with *R* inputs is shown below. Each input is weighted with an appropriate *w*. The sum of the weighted inputs and the bias forms the input to the transfer function *f*. Neurons can use any differentiable transfer function *f* to generate there output

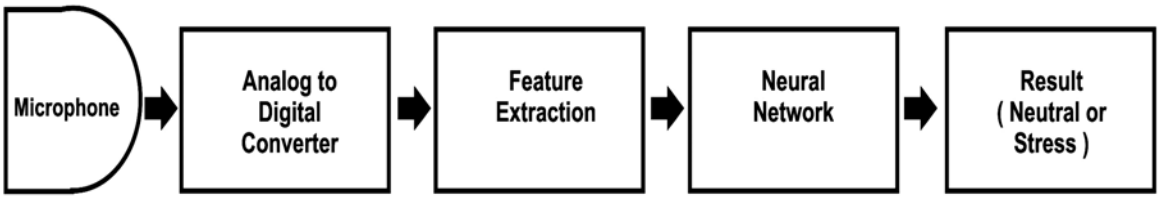

Fig .3.1. Block diagram of proposed method

.

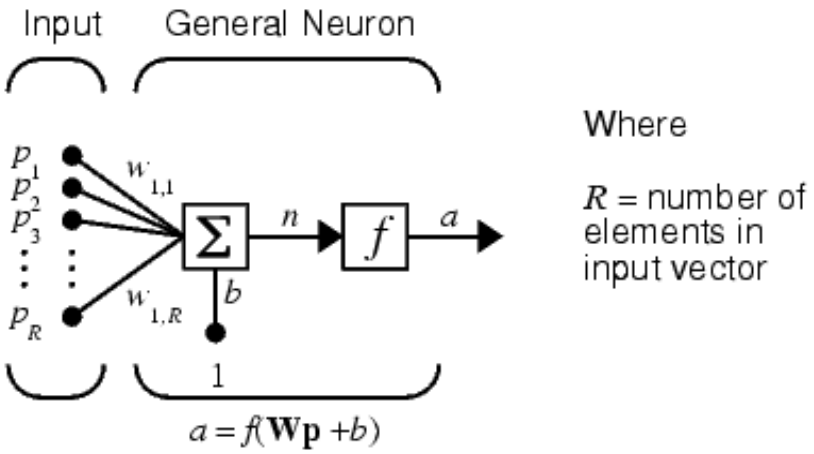

Fig.3.2. Multilayer networks often use the log-sigmoid transfer function `logsig`

4. Observations

The waveforms of features viz-fundamental frequency,

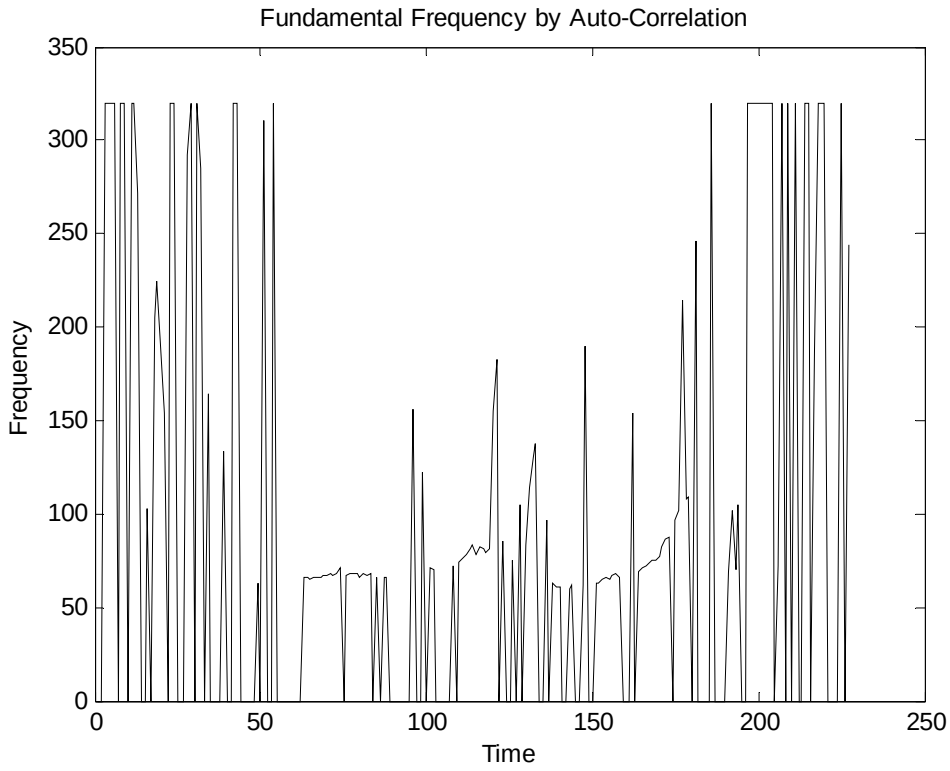

Fig. 4.1. Fundamental Frequency for Neutral Sample

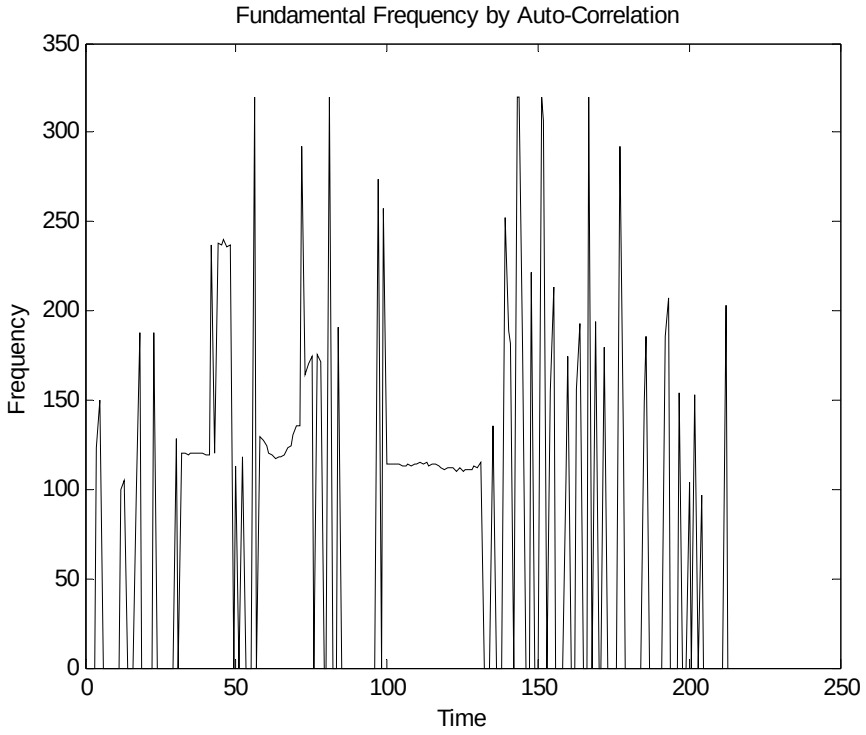

Fig.4.2. Fundamental Frequency for Strain Sample

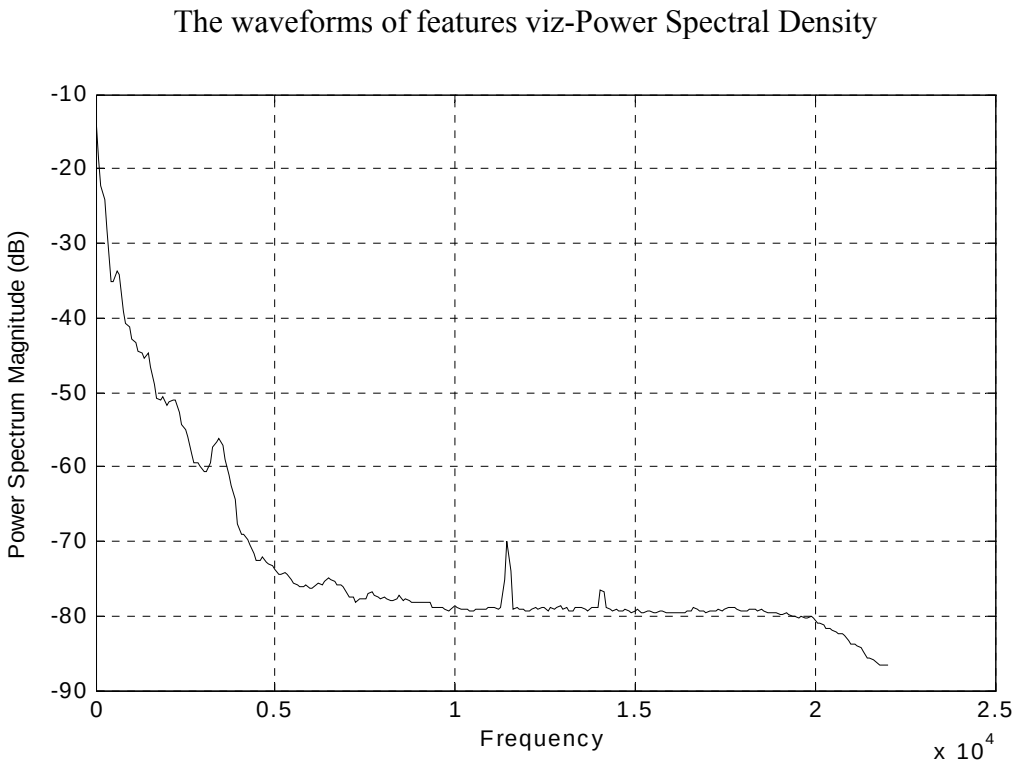

Fig. 4.3. Power Spectral Density for Neutral Sample

The waveforms of features viz-Short Time Average Energy

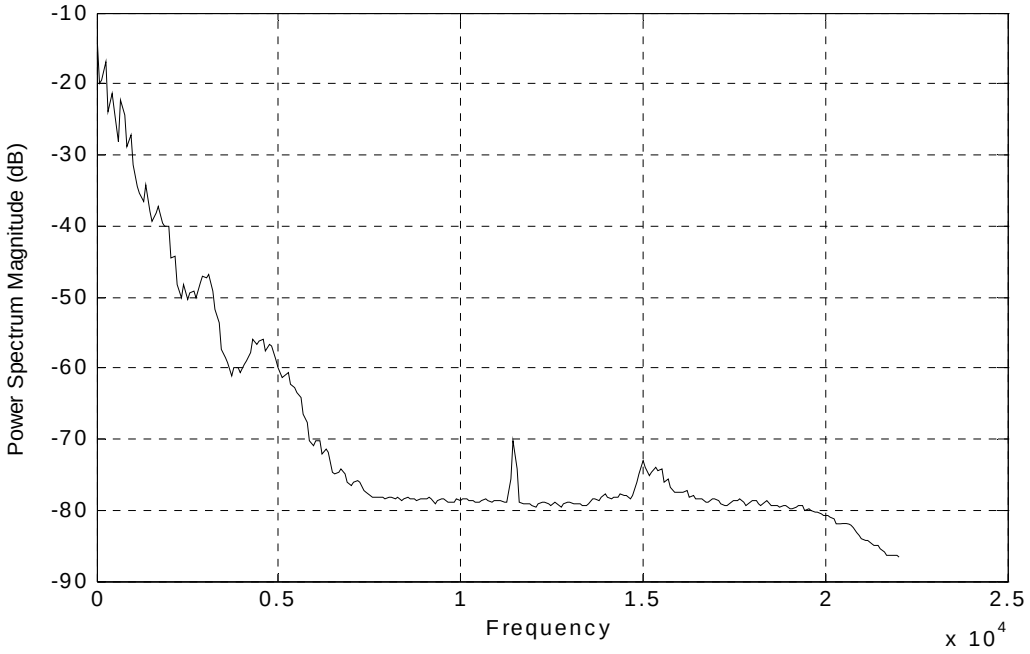

Fig. 4.4. Power Spectral Density for Strain Sample

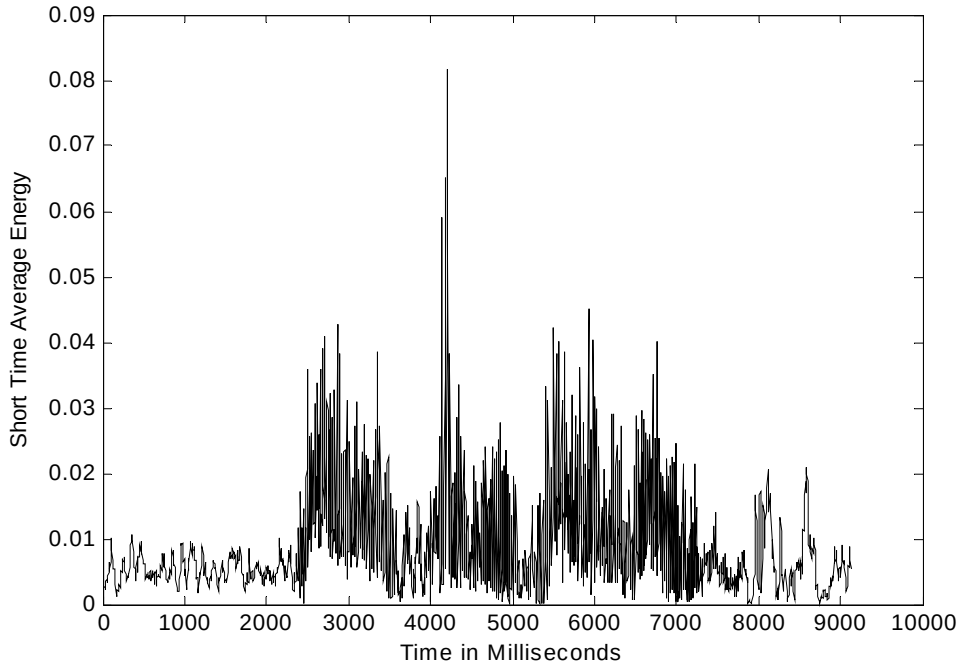

Fig 4.5. Short Time Average Energy for Neutral Sample

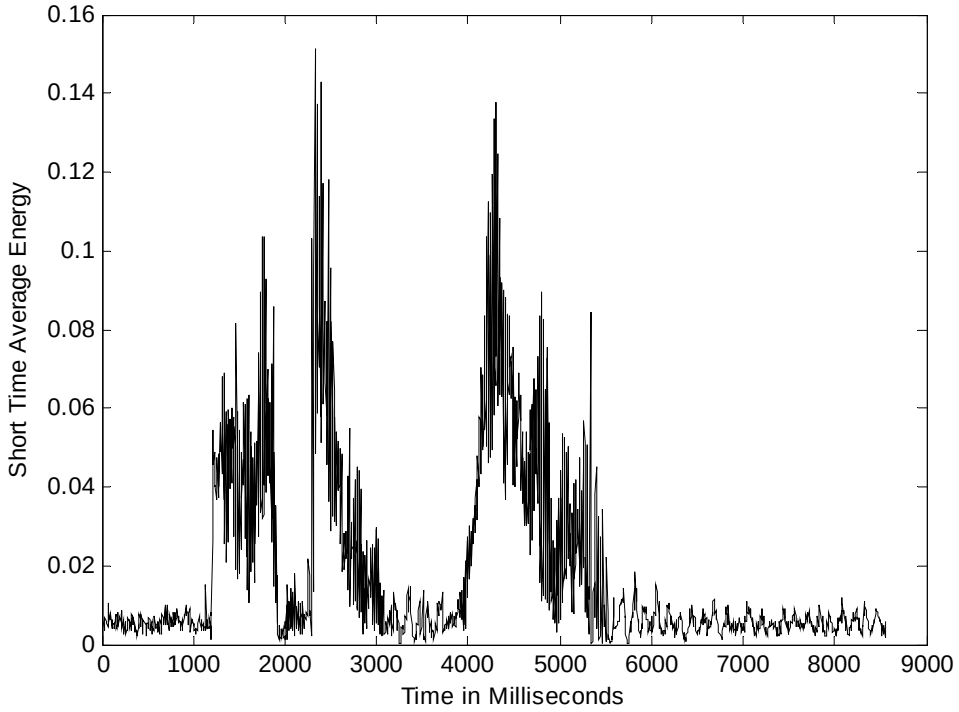

Fig.4.6. Short Time Average Energy for Strain Sample

The waveforms of features viz-Zero Crossing

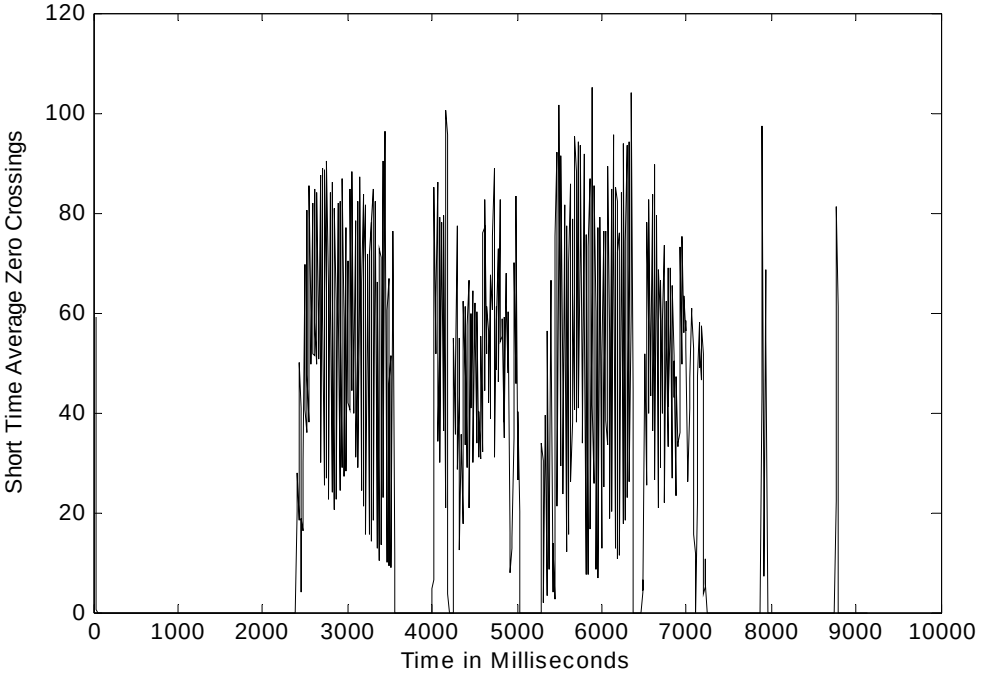

Fig 4.7. Short Time Average Zero Crossings for Neutral Sample

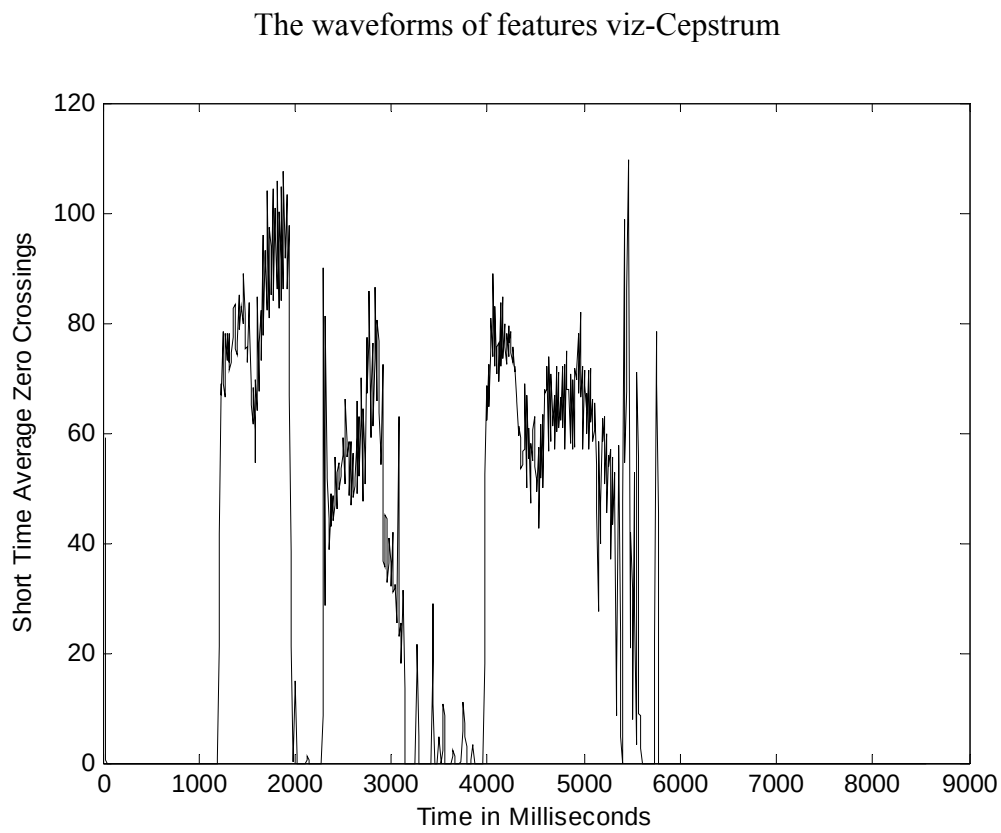

Fig 4.8. Short Time Average Zero Crossings for Strain Sample

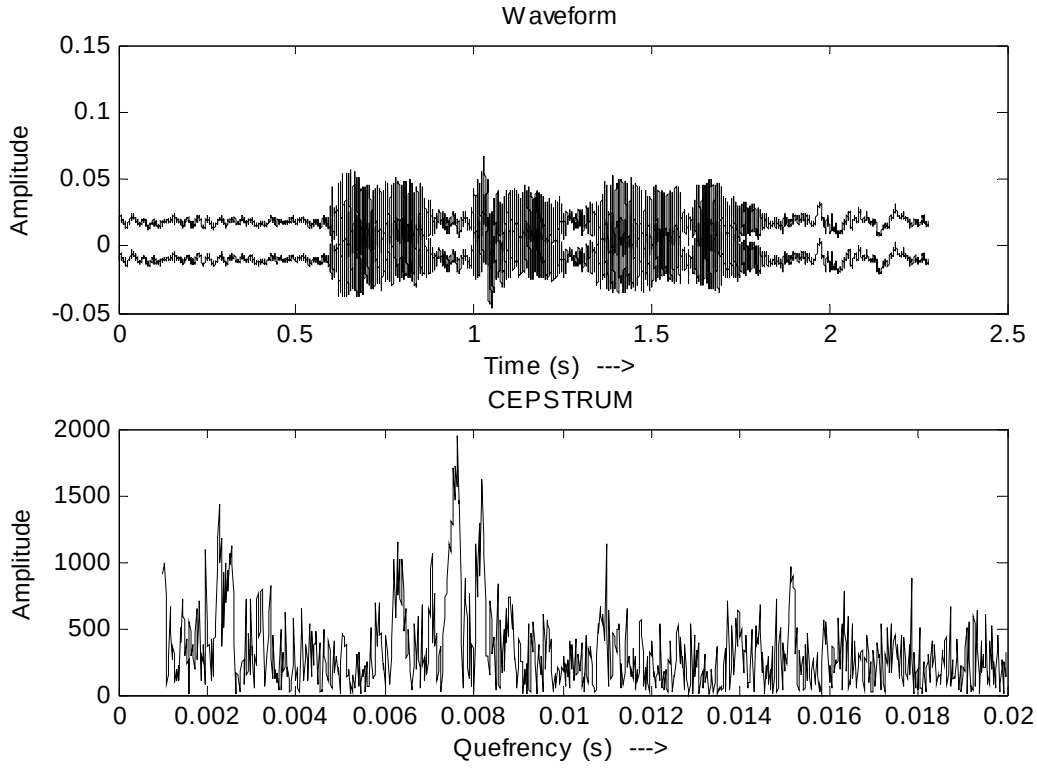

Fig.4.9. Cepstrum for Neutral Sample

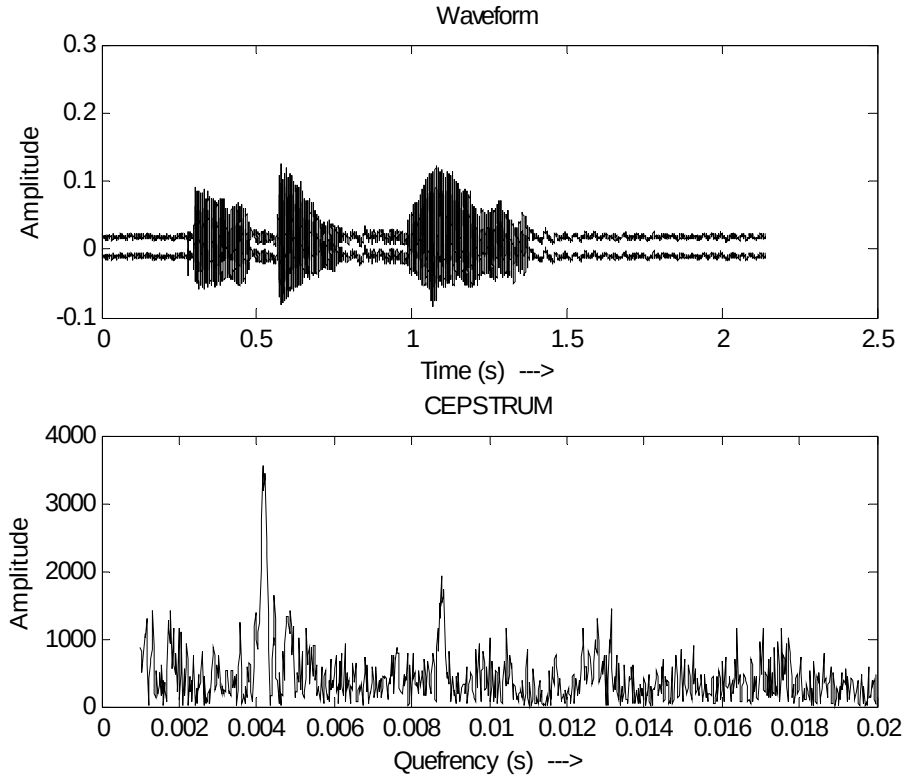

Fig.4.10. Cepstrum for Strain Sample

5. Results

This experiment uses phonetically rich sentences from the Exam Strain corpus for our psychoanalysis of strained speech. These sentences were automatically segmented into phoneme-like units. ANN does comparison of Neutral and Strain Sample. The results are shown below:

6. Conclusion

The Spectral Psychoanalysis of speech signal is aimed at extracting spectral features such as fundamental frequency, Short-time-energy, short-time-zero crossing, Spectrum, Cestrum, Spectral centroid etc. Changes in spectrum of speech signal have shown to be an indicator of the internal emotional state of a person. This research work, has extracted these spectral features of some speakers in neutral condition and under strain condition. This research has formed the feature matrix of the feature vectors obtained. For classification of the speech signal for strain Artificial Neural Network plays main role. The Standard Deviation of short –Time Energy is a reliable indicator of strain**.** Thus, the study concludes that spectral psychoanalysis is an efficient tool for detecting strain in speech in its various areas of applications. Spectral psychoanalysis can be used in the terrorism forensics.

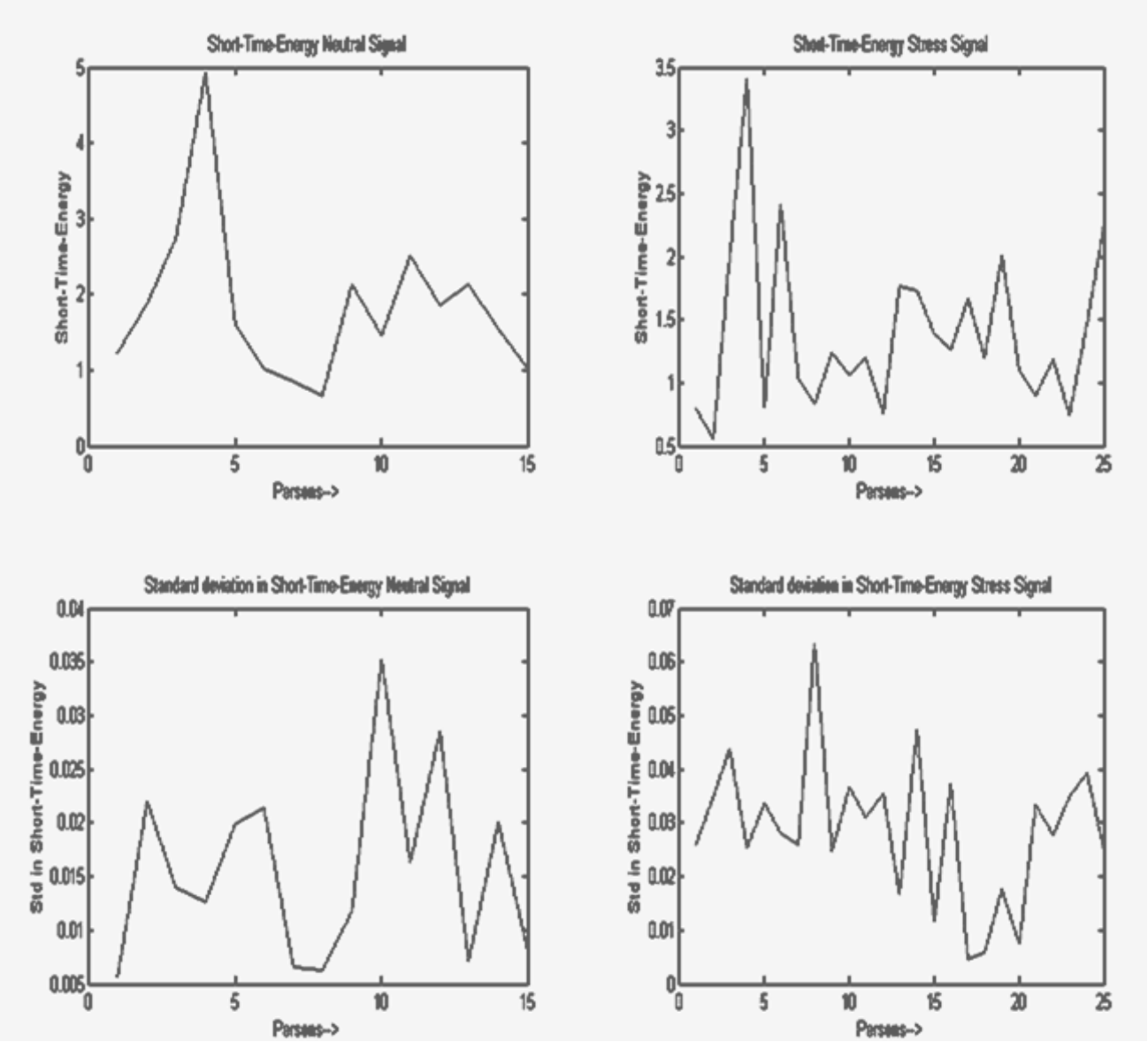

Fig. 4.11. Standard Deviation of short –Time Energy

7. References

Johnstone T and Scherer K (1999). The effects of emotions on voice quality. *Proceedings of 14th International Congress of Phonetic Science. San Francisco*. pp.2029-2032.

Levitt H and Rabiner LR (1971). Analysis of Fundamental Frequency Contours in Speech. *Journ. Acoust. Soc. Amer.* 49(2): 569-582.

Rule R (1969). Investigation of Strain Patterns for Speech Synthesis. *Rosenberg Journal Acoust. Soc. Amer.* 45(1): 92-101.

Schafer RW and Rabiner LR (1971). Design of Digital Filter Banks for Speech Analysis, *Proceedings of the Fifth Annual Princeton Conference on Information Sciences and Systems.* pp.40-47.

Ververidis D and Kotropoulos C (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication* 48(9): 1162-1181.

www.archive.org/stream/.../scopeofpsychoana012068mbp_djvu.txt

www.jkp.com/catalogue/tag/autism

www1.chapman.edu/~ruppel/ConferenceAbstracts.htm

www.humanities.uci.edu/critical/publications.html

www.eoneill.com/references/cotsell/chapter11.htm