International Journal of Engineering & Technology, 7 (1.2) (2018) 66-70

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

An approach to spectral analysis of psychologically influenced speech

Bhagyalaxmi Jena *, Sudhansu Sekhar Singh

Department of Electronics and Communication Engineering, Silicon Institute of Technology, Bhubaneswar
**Corresponding author E-mail: bjena@silicon.ac.in*

Abstract

The significant part of any speech signal lies in the information content and the emotions contents like stress or fatigue at a particular period of time. The classification of various types of stress and their effects are defined here. To analyze the changes in stressed speech than that of the normal speech, a database has been created which has investigated the stress among students during the examination in our college. In this paper, the spectral analysis of speech is done where emphasis has been given in the parameters like Fast Fourier Transform (FFT), spectrogram and Power Spectral Density (PSD). These parameters have been simulated using MATLAB codes. The comparison of the mentioned parameters is also done between a normal speech and a psychological stressed speech.

Keywords: *Speech Signal; Stress; Fast Fourier Transform; Spectrogram; Power Spectral Density.*

1. Introduction

Speech is considered as an important element in the arena of communication. However, the speech is not only conferred with communication rather it also represents the emotion, the reactions of the speaker in a particular environment. Factors like mood, physical characteristics are also available with the speech signal. One of the factors considered here is stress. Stress has been classified as physiological and psychological. These factors present in a speech signal distinguishes it in the neutral conditions.

This paper is based on finding the difference in pattern of normal speech and the stressed speech. This is accomplished by using different analysis method like the time domain analysis and frequency domain analysis [1].In analysis in time domain, the normal energy function, autocorrelation function and the zero crossing rate parameter can be used to study the difference in patterns for normal speech and stressed speech. Likewise, in frequency domain, this paper used the Fast Fourier Transform (FFT), spectrogram and power spectral density analysis (PSD). With the waveforms obtained, corresponding average values were calculated and the results were tabulated for different speakers having spoken the same sentence and in the same stress conditions.

2. Stress in the speech

2.1. The speech

Speech is the vocalized aspect of communication in human. It is totally based on the syntactic combination of lexical and names that are taken from large vocabularies. The main purpose of speech is to communicate. There are various ways in which potential of a speech communication can be characterized. A highly quantitative approach is made in information theory which was put forward by Shannon. As per the information theory, we can represent speech in terms of its message content. Another way to characterize speech is from the signal carrying the information of the message, i.e. the acoustic form of the signal [2].

When the air is pressurized from the lungs through the vocal cords, then the Speech is produced along the vocal tract. The vocal tract starts from the opening in the vocal cord (known as the glottis) till the mouth, and in an average person it is about 17 cm in length [3].

A very important part present in many speech codes is the designing of the voice channel like a short-term filter. And because the shape of the voice tract varies relatively slow, the transfer function of the filter modeling the tract needs to be updated frequently (generally in every 20 millisecond). Speech sound can be divided into three sub-classes depending on their mode of excitation.

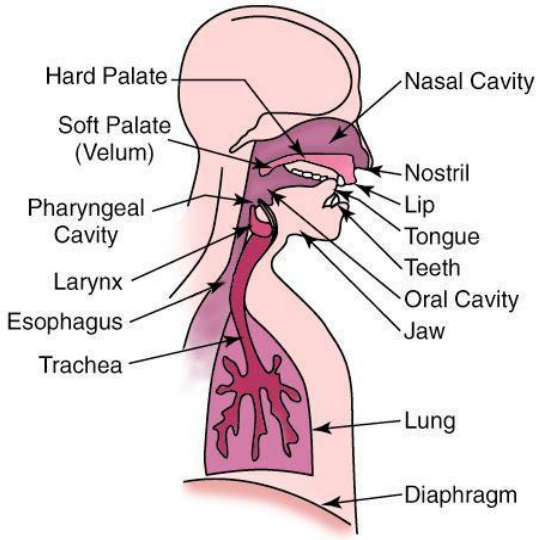

Fig. 1: Anatomy of Human Speech Production System.

2.2. Stress in speech

Stress is defined as the condition that causes the speaker to change the speech produced from normal situations. The stress can be broadly classified as two types i.e., psychological or perceptual stress and physiological stress.
Perpetual stress comes into picture when a person feels that the environment differs from normal such that his intention of production of speech changes from the normal conditions. The agents which cause the perceptual stress includes emotion, noise in the environment and actual work pressure.

2.3. Stressed speech

Stress is defined as the psychological state which is a response to a threat or a demanded task and is usually accompanied by emotions like fear, anger, sorrow, etc. Stress can also be brought about by external factors (such as workload, noise, vibration, sleep loss, etc.) and also by internal factors (such as emotions or fatigue state) [3]. The changes in emotion can affect the behavior of the speech involuntarily. Thus, "stressed speech" can be defined as any deviation in speech with respect to that of normal speech. Stress is defined as the relative emphasis that is given to particular syllables in a word, or to some words in a sentence. Stress is basically characterized by the properties such as high loudness and length of vowels, complete articulation of vowels, and variations in pitch.

3. Psychological effects of the stress

Stress is the psycho-physiological state which is signaled by subjective strain, dysfunctional physiological activities and degradation of performance. Psychological stress exactly refers to the emotional as well as physiological reactions which are experienced when an individual faces a situation in which they find the situation to be not normal or unperceivable. Major examples of stressful scenario include health hazards, and financial crisis, which affects the individual emotionally [4].
According psychology, stress is considered as the stimulus of pressure and strain. Minor amounts of stress could be desired, beneficial, and even healthy. Moreover, the positive stress helps in improving unpersons performance. It also plays an important role in motivation, adaptation, and reaction to the environment as required by it. However, high amounts of stress may harm the health of the person. Stress increases the hazard of strokes, cardiac arrest, and mental disorders like depression. Stress may be external and environment related, but internal perceptions can be created that may cause an individual to feel the anxiety or different types of negative emotions surrounding a situation, like pressure, discomfort, etc., which the individual find full of stress.
The effects of psychological stress can be inferred from the figure depicted below:

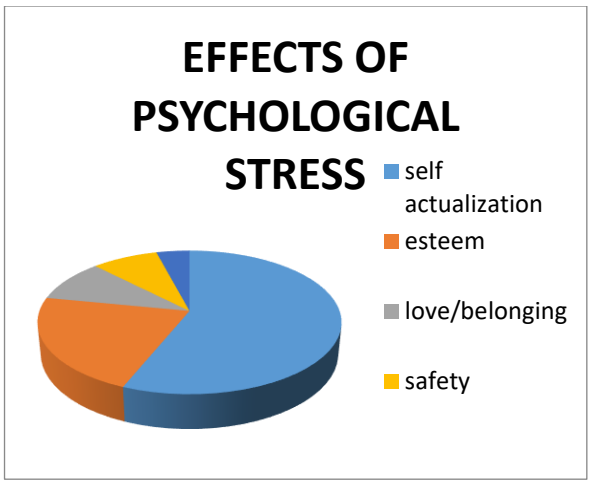

Fig. 2: Pie-Chart Representation.

4. Exam stress database-'SUNAS'

A sound database has been created using the software called WavePad Sound Editor. The database is here named as 'SUNAS' (speech under normal and actual stress). During the examinations, the database has been created by recording the speech of the students once before one hour of the commencement of the exam and again after an hour of the exam. Speech of 10 males and 10 females has been recorded. The speech signal pattern changes with the contents of the utterance of speech as well as with the emotions associated with it, thus the sentence "The weather is too hot today." has been taken into consideration so as to make the analysis more precise and distinguishable. Before the exam the speech recorded is considered as stressed speech as students are under the pressure of the exam and similarly after the exam the students are relaxed and the speech in that situation is considered as the normal speech. The complete database has been created in the same manner. Thereafter the changes in the pattern of stress speech are studied over the normal ones. Some parameters were tabulated.Some important information about the database can be inferred from the following table:

Table 1: Parameters of Stress and Normal Speech

.	Normal Speech	Stress Speech
Stress	Exam News	Exam News
Language	English	English
Speech Type	Read Sentence	Spontaneous
Speaker	20	20
Gender	10 Male 10 Female	10 Male 10 Female
Sampling Rate	44100	44100
Quality	Very Good	Good

5. Analysis of speech

5.1. Fast Fourier transform

Fast Fourier Transform is an algorithm used to compute the Discrete Fourier Transform (DFT) of a signal. DFT is a tool which converts a time domain signal into its respective frequency domain representation [5]. Thus, for analyzing the spectral parts of the speech signal under various stress conditions, this method was used.
DFT is defined as:

$$X_k = \sum_{n=0}^{N-1} x(n)\, e^{-i2\pi kn/N}$$

Where, x (n) is the input signal

k = 0, 1 … N-1.

N = total number of samples.

The FFT is considered as an efficient algorithm while computing the DFT of a given sequence. Typical all FFT algorithms is the periodicity and symmetry of the exponential term and the possibility of breaking down a transform into a sum of smaller transforms for subsets of data. Since n and k are integers, the exponential term is present and will be periodic with period N. And this is commonly known as twiddle factor and is represented by,

$$W_N = e^{-j2\pi/N}$$

FFT algorithms are completely based on the principle of
Decimation-in-time
Decimation-in-frequency
Most of the FFT algorithms are based on the principle of decimation-in-time, which involves the decomposition of original time or frequency sequence to its smaller subsequences. The total number

of complex multiplication and addition present in, DFT is $N^2$and N (N-1) respectively. Whereas in FFT total number of complex multiplication and addition is reduced to $(N/2)\log_2 N$ and $N\log_2 N$ respectivily. The FFT algorithms usually find their application in a wide range of areas, including linear filtering, correlation and spectrum analysis. Basically, the FFT algorithm is used as the most efficient way to compute the DFT and IDFT. FFT decomposes an N point time domain signal into N time domain signals each composed of a single point. The calculation of the N frequency spectra is the second step and lastly, these are synthesized into a single frequency component.

The fundamental query of longstanding theoretical interest is to prove lower bounds on the complexity aspect and the exact operation counts of the fast Fourier transforms, and many open problems remain [6]. But it is not even rigorously whether DFTs exactly require Ω (N log N) (i.e., order N log N or greater) operations, even for the simple case of power of two sizes, although no other algorithms with the lower complexity are still known. To be particular, the arithmetic operations count is basically the main focus of such type of questions, although the actual performance on present-day computers is calculated by various other factors such as CPU pipeline optimization or cache.

Following pioneering work by Winograd (1978), a tight Θ (N) lower bound is known for the total number of real multiplications that an FFT requires. It can be exhibited that only irrational real multiplications are required to compute a DFT of power-of-two length, $N=2^m$. Moreover, explicit algorithms that achieve this count are known. Unfortunately, these algorithms require a lot of additions to be practical, at least on modern computers with hardware multipliers.A tight lower bound cannot be known on the number of required additions, although lower bounds have been proven under some restrictive assumptions on the algorithms. However, this result applies only to the un normalized Fourier transforms and it does not explain why the Fourier matrix is difficult to compute than any other unitary matrix under the same scaling.

Here are some of the important applications of the FFT:

- Faster big integer and polynomial multiplication
- Computational efficient matrix-vector multiplication for Toeplitz, circulant and other structured matrices
- Filtering algorithms
- Fast algorithms which is used for discrete cosine or sine transforms (e.g. Fast DCT used for JPEG, MP3 encoding)
- Fast Chebyshev approximation
- To solve difference equations

5.2. Spectrogram

The spectrogram is defined as the visual representation of different frequency bands present in a signal in the given time intervals or some other variables. In this paper, we have considered the time as the independent variable. Spectrograms are created using the corresponding computed FFT of the given signal. Here, in every time interval, the spectral components present in that time interval are created and are represented in horizontal line while the vertical line separating these bands are of time intervals. Different shades in the spectrogram represent different energy densities for the corresponding frequencies in that time interval. The lighter shades represent lower energy density while the darker ones represent higher energy densities. Spectrogram is also called as spectral waterfalls, voiceprints, or voicegrams. Spectrogram can also be used for the identification of spoken words phonetically, and for analysis of the various voices of animals. Extensively they are used in the development of the field of SONAR, music, RADAR, and seismology,speech processing etc.

The most common format is the graph with two geometric dimensions: the horizontal axis will represent time or rpm, the vertical axis will represent frequency; a third dimension indicate the amplitude of a particular frequency present at a particular instant of time is represented by the color or intensity of each point in the image.Spectrogram is basically a two-dimensional graph, with a third dimension present and it is represented by colors. Time runs from left to right alongthe horizontal axis. Volcano and earthquake sub-groups of spectrograms shows around 10 minutes of data with the tic marks in the horizontal axis which corresponds to one minute interval. Our tremor sub-groups of spectrogram show around one hour of data. The vertical axis of the spectrogram represents the frequency, which can also be thought of as pitch or tone, with the lowest frequencies at the bottom and the highest frequencies at the top. The amplitude or energy or loudness of a particular frequency at a particular time is represented by the third dimension, color, with dark blues corresponding to low amplitudes and brighter colors up through red corresponding to progressively stronger amplitudes.

There are many variety of formats: sometimes the vertical axis and horizontal axis are switched, so time will run up and down. The amplitude is sometimes represented as the height of the 3D surface instead of intensity or color. We can make the frequency and amplitude axes either linear or logarithmic, depending on what the graph is being used for. The audio would usually represented with a logarithmic amplitude axis (probably in dB), and frequency will be linear to focus on harmonic relationships, or logarithmic to focus on musical, tonal relationships.

Spectrograms can be created in one of two ways: it may be approximated as a filter bank that results due to a series of band-pass filters (this was the one and only way which was used before the advent of modern digital signal processing), or calculated from the time signal using the Fourier transform. These two methods actually form two different time–frequency representations, but are equivalent under some conditions.

The bandpass filter method normally uses the method of analog processing in order to divide the input signal into the frequency bands; the magnitude of each filter's output controls a transducer that records the spectrogram as an image on paper. Creating a spectrogram using the FFT is fundamentally a digital process. Digitally sampled data in the time domain are broken up into chunks and it usually overlaps, and Fourier transformed to calculate the magnitude of the frequency spectrum for each chunk. Each chunk, then corresponds to a vertical line in the image; a measurement of magnitude versus frequency for a specific moment in time. These spectrums or time plots are then "laid side by side" to form the image or a three-dimensional surface, or slightly overlapped in various ways, i.e. windowing [7].

There are various applications of spectrogram which includes the study of speech synthesis and phonetics, steganography, audio timescale pitch modification, phase vocoder etc. Spectrograms is basically used to analyze the results of passing a test signal through a signal processor such as a filter in order to check its performance. High definition (HD) spectrograms are mostly used in the development of RF and microwave system. The spectrogram is now used to display S-parameters measured with vector network analyzers.The US Geological Survey (US-GS) now provides real-time (RT) spectrogram displays from a seismic station.

5.3. Power spectral density

Power spectral density is defined as the measure of distribution of power of the signal over the range of frequencies present in the signal. The PSD can be calculated from FFT too. This can be done by using the power formula on the FFT of the input signal. A different way of finding the PSD is by calculating the equivalent Fourier Transform of the signal and after that squaring the Fourier coefficients will give us the required PSD of the signal. In this paper, we have used only the FFT power calculation method. From the calculated PSD, we can get a brief idea of the variation of the signal power for different frequency components. The power of a signal is given by:

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} x(t)^2 dt$$

Where, $x(t)$ is input signal?

The above formula is now applied to FFT of the input signal to get the required PSD.

Hence, it is considered as a measure of a signal's power magnitude in the frequency domain. In a practical scenario, the PSD is computed from the FFT spectrum of the signal. The above equation can also be dedicated as a Parseval's theorem.

Any signal which can be represented as an amplitude that varies with the time always has a particular corresponding frequency spectrum. When these signals are considered in the form of frequency spectrum, few aspects of the received signals or the underlying processes producing them are disclosed. And additionally there are chances of presence of peaks corresponding to the harmonics of a fundamental peak, which indicates a periodic signal which is not simply sinusoidal. Continuous spectrum shows narrow frequency intervals which are strongly enhanced corresponding to resonances. [8-10].

6. Results

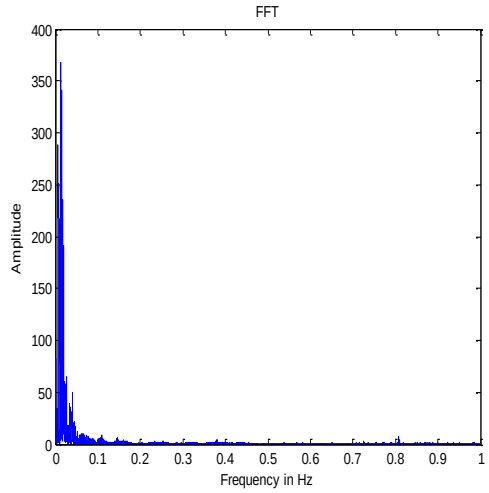

Fig. 3: FFT of the Input Normal Speech.

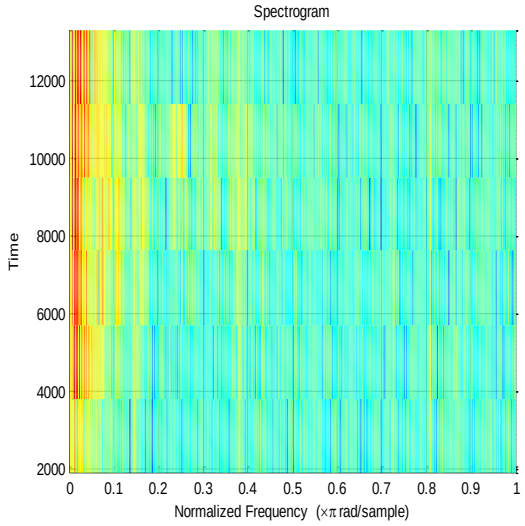

Fig. 4: Spectrogram of the Input Normal Speech.

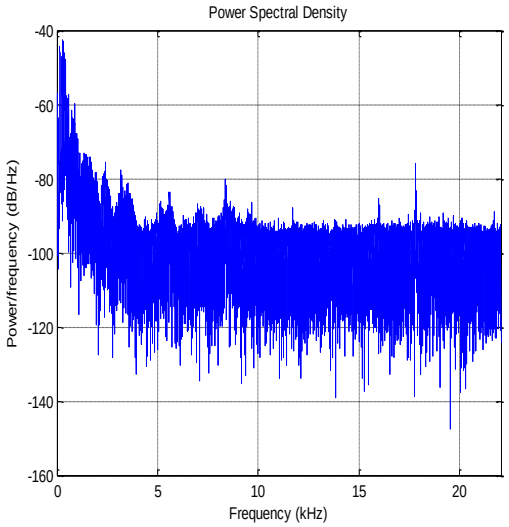

Fig. 5: Power Spectral Density of the Input Normal Speech.

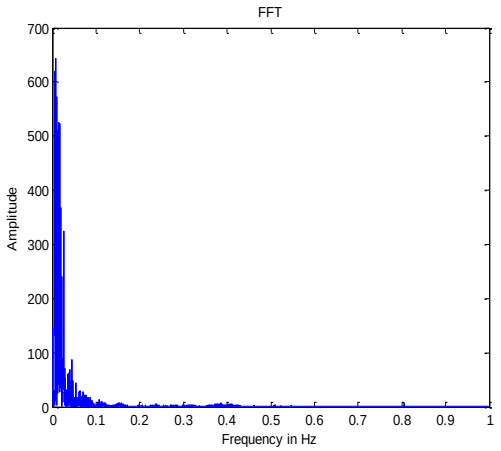

Fig. 6: FFT of Stressed Speech Signal.

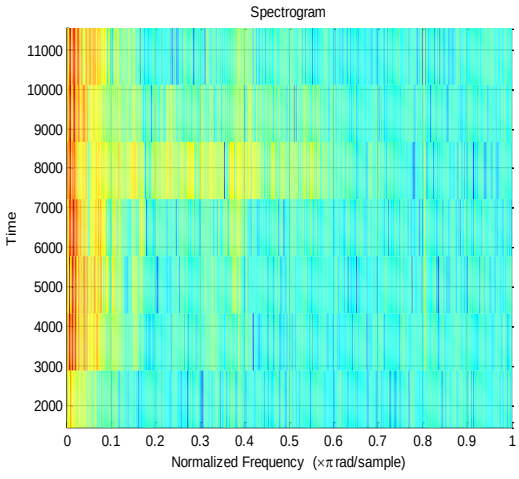

Fig. 7: Spectrogram of Stressed Speech Signal.

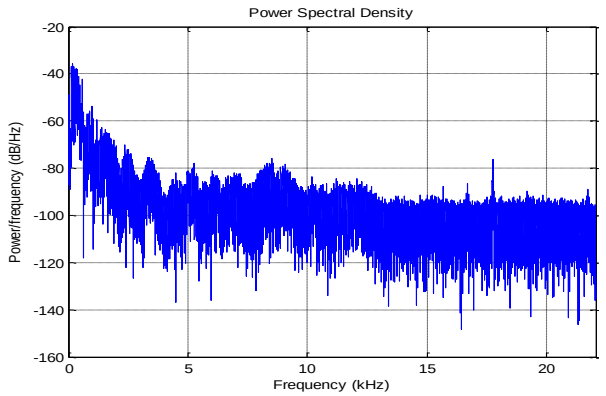

Fig. 8: Power Spectral Density of Stressed Speech.

7. Conclusion

In this study, we have tried to distinguish between the normal speech and the stressed speech using the parameters of the frequency domain analysis, we analysed the FFT, spectrogram and PSD of the speech signal. The amplitude and also the frequency content of the stressed speech is much greater than the normal speech, as it is clearly depicted by the FFT plot [11-13]. Contrary to this, the normal speech had the energy intensity of higher frequency term higher for only a short duration of time. And the last parameter, power spectral density (PSD), showed that the power of lower frequency terms was somewhat similar for both the normal speech and the stressed speech. The difference in them arises when we consider the higher frequency terms. The power of higher frequency terms in the stressed speech was way greater than the corresponding frequency terms in normal speech.

References

[1] M. Sigmund. (2006). "Introducing the database ExamStress for speech under stress," *Proceedings of 7th IEEE NordicSignal Processing Symposium (NORSIG 2006)*. Reykjavik, (pp. 290-293). https://doi.org/10.1109/NORSIG.2006.275258.

[2] D. A. Cairns & J. H. L. Hansen. (1994), "Nonlinear analysis and detection of speech understressed conditions," *J. Acoust. Soc. Amer*., vol. 96, (pp.3392–3400). https://doi.org/10.1121/1.410601.

[3] V. Mohan. (2013). "Analysis & Synthesis of Speech Signal Using Matlab", International Journal of Advancements in Research & Technology, Volume 2, Issue 5.

[4] T. Johnstone & K. Scherer. (1999) "The effects of emotions on voice quality," *Proceedings of 14th International Congressof Phonetic Science*. San Francisco, (pp. 2029-2032).

[5] D. Ververidis & C. Kotropoulos. (2006). "Emotional speech recognition: Resources, features, and methods," *SpeechCommunication*, vol. 48, No. 9, (pp. 1162-1181). https://doi.org/10.1016/j.specom.2006.04.003.

[6] L. R. Rabiner & B. H. Juang. (1993) *Fundamentals of Speech Recognition*,Englewood Cliffs, NJ: Prentice-Hall.

[7] Cowie & R.Cornelius, R.R. (2003). Describing the emotional statesthat are expressed in speech. Speech Comm. 40 (1), 5–32.Cowie, R., Douglas-Cowie, E., 1996. Automatic statistical.Rep. 236, Univ. of Hamburg. https://doi.org/10.1016/S0167-6393(02)00071-7.

[8] Flanagan, J.L. (1972). Speech Analysis, Synthesis and Perception.second ed. Springer-Verlag, NY. https://doi.org/10.1007/978-3-662-01562-9.

[9] Heuft, B., Portele & T. Rauth, (1996). Emotions in time domain synthesis. In: Proc. Internat. Conf. on Spoken Language Processing (ICSLP '96), Vol. 3, (pp. 1974–1977).

[10] Markel, J.D., Gray & A.H. (1976). Linear Prediction of Speech. Springer-Verlag, NY. https://doi.org/10.1007/978-3-642-66286-7.

[11] Quatieri, T.F. (2002). Discrete-Time Speech Signal Processing. Prentice-Hall, NJ.

[12] Rahurkar & M.Hansen (2002). Frequency band analysis for stress detection using a Teager energy operator based feature. In: Proc. Internat. Conf. on Spoken Language Processing (ICSLP '02), Vol. 3, (pp. 2021–2024).

[13] Steeneken & Hansen (1999). Speech under stress conditions: overview of the effect of speech production and on system performance. In: Proc. Internat. Conf. on Acoustics, Speech, and Signal Processing (ICASSP '99), Phoenix, Vol. 4, (pp. 2079–2082).

[14] Womack & B.D., Hansen, (1996). Classification of speech under stress using target driven features. Speech Comm. 20, (pp.131–150). https://doi.org/10.1016/S0167-6393(96)00049-0.

[15] Zhou, G., Hansen, J.H.L. & Kaiser, J.F. (2001). Nonlinear feature-based classification of speech under stress. IEEE Trans.Speech Audio Processing 9 (3), (pp.201–216). https://doi.org/10.1109/89.905995.

[16] Deller, J. R., Hansen, J. H. L., Proakis, J. G. (2000). Discete- Time Processing of Speech Signals. N.Y.: Wiley.

[17] M. Sigmund, *Voice Recognition by Computer*. Tectum Verlag, Marburg. (2003).

[18] M. Sigmund & P. Matějka. (2002) "An environment for automatic speech signal labelling," *Proceedings of 28th IASTED International Conference on Applied Informatics.* Innsbruck, (pp. 298-301).

[19] A. Nagoor Kani. (2005). Signals & Systems. Tata McGraw Hill Education.

[20] Sanjit K Mitra. (2009). Digital signal processing, A computer base approach, Tata McGraw Hill.

[21] Lawrence R. Rabiner & Ronald W. Schafer. (2003). Digital Processing of Speech Signals. AT&T.

[22] Alan V. Oppenheim, Alan S. Willsky & S. Hamid Nawab.(2005). Signal & Systems. PHI Learning.

[23] J.H. Hasen & S.E.Ghazale.Getting started with SUSAS. Proceedings of Eurospeech'97. Rhodes, (pp.1743-1746).

[24] M.Kepesi & L.Weruaga. (2006). Adaptive chirp-based time-frequency analysis of speech signals.vol.48, No.5, (pp. 474-492).

[25] B. Gold & N. Morgan. (2000). Speech and AudioSignal Processing. New York. John Wiley and Sons.

[26] Milan Sigmund. (2007). Spectral Analysis of speech under stress. IJCSNS International Journal of Computer Science and Network Security, vol.7.

[27] J.H.L Hansen & B.D.Womack. (1996). Feature analysis and neural network-based classification of speech under stress.(pp. 307-313)

[28] R.J McAulay & T.F. Quatieri. (1986).Speech Analysis based on a Sinusoidal Representation. IEEE Transaction On Audio, Speech, And Language Processing.Vol.14.No.3 https://doi.org/10.1109/TASSP.1986.1164910.

[29] W.Press, S.Teukolsky, W.Vetterling & Flannery. (1992).

[30] Ruhi Sarikya & John N. Gowdy. (1997). Wavelet Based Analysis of Speech under stress.

[31] B.S. Atal. (1976). Automatic Recognition of Speakers from their Voices. Vol.64. no. 4(pp. 460-476) https://doi.org/10.1109/PROC.1976.10155.

[32] D.O' Shauhnessy. (2004). Speech Communication (Human and Machine).

[33] Herman J.M. Steeneken and Johan H.L. Hasen. Speech under Stress Conditions: Overview of the Effect on Speech Production and on System Performance.