Time Domain and Frequency Domain Analysis On Psychological Stress Speech Signals

[1]Bhagyalaxmi Jena,[2]Sudhanshu Sekhar Singh

[1]Department of Electronics and Communication Engineering, Silicon Institute of Technology, Bhubaneswar.
[2]Department of Electronics and Communication Engineering, KIIT University, Bhubaneswar.
Email: bjena@silicon.ac.in

Abstract: This paper is based on finding the difference in pattern of normal speech and the stressed speech. This is accomplished by using the time domain analysis and frequency domain analysis. In the time domain analysis, this paper used the normal energy function, autocorrelation function and the zero crossing rate parameter to study the difference in patterns for normal speech and stressed speech. Likewise in the frequency domain, this paper used the Fast Fourier Transform (FFT), spectrogram and power spectral density analysis (PSD).

Keywords-Energy,Autocorrelation,FFT,Spectrogram

I. INTRODUCTION

This paper is based on finding the difference in pattern of normal speech and the stressed speech. In time domain analysis, this paper used the normal energy function, autocorrelation function and the zero crossing rate parameter to study the difference in patterns for normal speech and stressed speech. Likewise in frequency domain, this paper used the Fast Fourier Transform (FFT) and spectrogram.

A neutral speech can be differentiated from a stressed speech considering different parameters such as amplitude, fundamental frequency, pitch, intensity, spectral energy. Before considering the analysis & synthesis of the system, we define the stress elements of the speech.

II. STRESSED SPEECH

Stress can be defined as any condition that causes a speaker to vary speech production from neutral conditions. If a speaker is in a 'quiet room' with no task obligations, then the speech produced is considered neutral. With this definition, two stress effect areas emerge: perceptual and physiological.

Perceptually induced stress results when a speaker perceives his environment to be different from 'normal' such that his intention to produce speech varies from neutral conditions. Thus, "stressed speech" can be defined as any deviation in speech with respect to the neutral style [1]. This deviation can be in the form of speaking style, selection and usage of words, duration of sentence, etc [2].

1.1 Speech Database

A wide range of speech databases are available which aims for the development of speech synthesis/recognition and for linguistic research[3]. A database of 10 males and 10 females was created and these subjects were evaluated under Exam Stress. Their speech was recorded just before the examination and an hour after the examination. As we know that the pattern of speech changes with the content of utterance of speech, so to make the analysis precise, the phrase, "The weather is too hot today" was taken into account.

2.2 Window Function

A window function is a mathematical tool that limits the input signal. That is, it allows only a defined interval of input signal, while restricting the outer interval of the signal. Thus, we can say that a window function is somewhat a time domain filter which allows only a defined interval of signal to pass while attenuating the signal falling outside the defined interval.

There are many types of window functions, like rectangular, hamming, hanning , blackmann etc[2].

A rectangular window is defined as:

$$w(n) = 1 ; \qquad 0 < n \leq (N\text{-}1) \qquad (1)$$
$$0 ; \qquad \text{otherwise}$$

where,

N is the total number of samples of the signal.

Hamming window is defined as:

$$w(n) = 0.54 + 0.46 \cos\left(\frac{2\pi n}{N}\right) \qquad (2)$$

where,

N is the total number of samples of the input signals.

III. ANALYSIS OF SIGNAL IN TIME DOMAIN

The analysis of mathematical functions, physical signals or time series of environmental data, with respect to time is done by the analysis of signal in time domain . In the time domain, the signal or function's value is known for all real numbers, for the case of continuous time, or at various separate instants in the case of discrete time . Time domain investigation of signals and systems is one of the most essential tool of electrical engineering. When a physical phenomenon is investigated, its time domain behaviour is one of the most important property which should be observed. In info communication often the shape of the received signal

carries the information (e.g., its amplitude, phase, rate of change…). Even if a signal is stored or transmitted in digital form, most essential building blocks of digital signals (bits) are represented by analogue signals in the physical layer. In order to establish a high quality digital communication, the analogue signals must be well-conditioned: high signal-to-noise ratio should be achieved, the state transitions should be sharp enough, oscillation and reflections should be avoided. Simple first- and second-order systems and transmission lines that will be investigated in the measurement are basic building blocks of several complex systems, so it is crucial to be familiar with the time-domain behaviour and measurement technique of these systems. Time domain analysis of speech signal refers to the analysis of mathematical functions and parameters associated with it with respect to time. Thus a time domain graph will represent how the signal changes over a span of time. The time domain mathematical functions and parameters used in this paper are energy, autocorrelation and zero crossing rate.

3.1.1 Energy

In physics, energy is defined as the ability to do work[5]. Hence, as variation in speech occurs, the energy content associated with it also changes [11]. The more is stress is put on a certain word, more is the energy associated with it.

The energy of a signal is calculated as:

$$E = \sum_{-\infty}^{+\infty} x(n)^2 \qquad (3)$$

where,

x(n) is the input signal, and E is the energy. Energy in this context is not, strictly speaking, the same as the conventional notion of energy in physics and the other sciences. When the signal is associated with some physical energy, it means that it gives the energy content in the signal.

3.1.2 Autocorrelation

In simple words, autocorrelation can be defined as the degree of similarity of a signal with its delayed one. Hence, it can be used to find the repeating patterns of a signal [6]. As noise is random in nature, it is highly non periodic and thus highly uncorrelated. Therefore we can say that noise has no attributes while calculating the autocorrelation and we are able to extract out the meaningful parameters from the signal.

The autocorrelation function is given by:

$$R_{ff} = \int_{-\infty}^{+\infty} f(u)\bar{f}(u-\tau) \qquad (4)$$

where,

f(u) is the input function,

$\bar{f}(u-\tau)$ is the delayed complex conjugate of f(u).

In discrete system, the autocorrelation function is given as:

$$R_{yy}(l) = \sum_{n \in Z} y(n)\bar{y}(n-l) \qquad (5)$$

IV. FREQUENCY DOMAIN ANALYSIS

Frequency domain analysis of speech signal is the analysis of mathematical functions and parameters associated with it with respect to time. In other words, a frequency domain graph will represent how the signal varies over a span of frequency band. This paper utilises tools like the Fast Fourier Transform and Spectrogram for the analysis of speech in frequency domain. The analysis of mathematical functions with respect to frequency is known as frequency domain representation. A frequency-domain representation can also include information on the phase shift that must be applied to each sinusoid in order to be able to recombine the frequency components to recover the original time signal. The frequency components of the spectrum is the frequency domain representation of the signal. The conversion from the frequency domain function to a time domain is known as Inverse of the Fourier Transform. A spectrum analyzer is the tool commonly used to visualize real-world signals in the frequency domain.

4.1 Fast Fourier Transform

Fast Fourier Transform (FFT) 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. Thus, for analysing the spectral parts of the speech signal under various stress conditions, this method was used[11].

The DFT is defined as:

$$X_k = \sum_{n=0}^{N-1} x_n \, e^{-i2\pi k \frac{n}{N}} \qquad (7)$$

where,

x(n) is the input signal,

k = 0,1, … N-1.,

N is the total number of samples

The fast Fourier transform (FFT) is an efficient algorithm for computing the DFT of a sequence. Typically the essence of 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 both integers, the exponential term is periodic with period N. This is commonly known as twiddle factor and is represented by,

$$W^N = e^{-j2\pi/N} \qquad \ldots\ldots\ldots\ldots(8)$$

4.2 Spectrogram

A visual representation of different frequency bands present in a signal in the given time intervals or some other variables is known as Spectrogram. In this paper, we have considered the time as the independent variable. Spectrograms were created using the corresponding computed FFT of the given signal. Here, in every time

interval, the spectral components present in that time interval were created and were 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. The frequency and amplitude axes can be either linear or logarithmic, depending on what the graph is being used for. Audio would usually be represented with a logarithmic amplitude axis (probably in decibels, or dB), and frequency would be linear to emphasize harmonic relationships, or logarithmic to emphasize musical, tonal relationships.

$$\lim_{f \to \infty} \frac{1}{2f} \int_{-f}^{+f} |X(f)|^2 df \qquad (8)$$

where,

$X(f)$ is the frequency representation of the input signal

V. TIME DOMAIN ANALYSIS
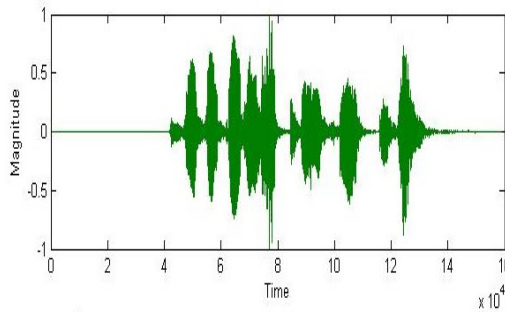
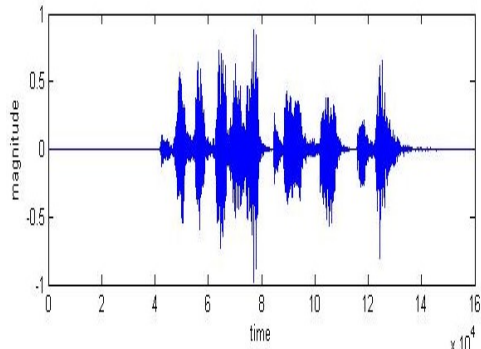Fig. 5.(a) Normal Speech Signal
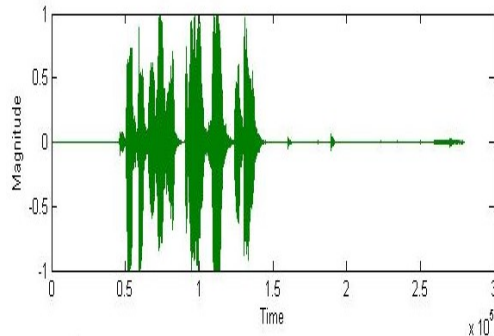
Fig. 5.(b)Stressed Speech Signal
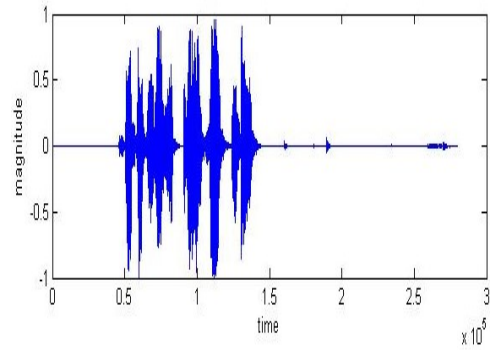

Fig. 5.(c) Windowed Normal Signal
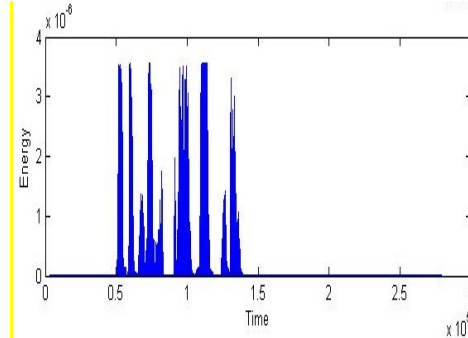

Fig.5.(d)Windowed Stressed Signal
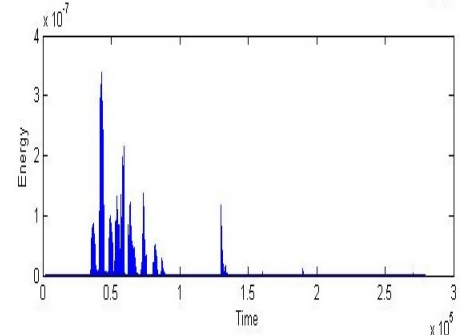

Fig. 5.(e) Energy of Stressed Speech

Fig. 5.(f) Energy of Normal Speech
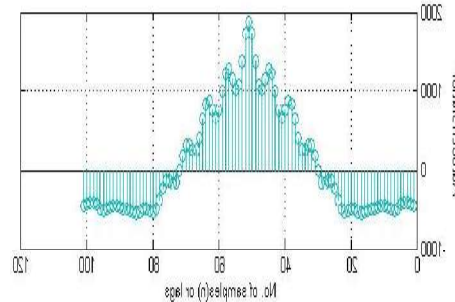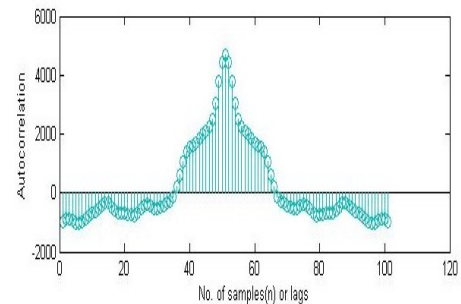

Fig. 5.(g)Autocorrelation of Normal Speech

Fig.5.(h)Autocorrelationof StressedSpeech

VI. FREQUENCY DOMAIN ANALYSIS
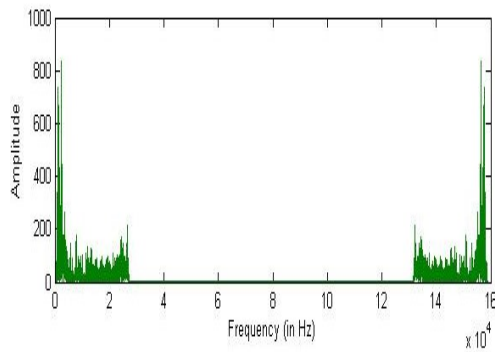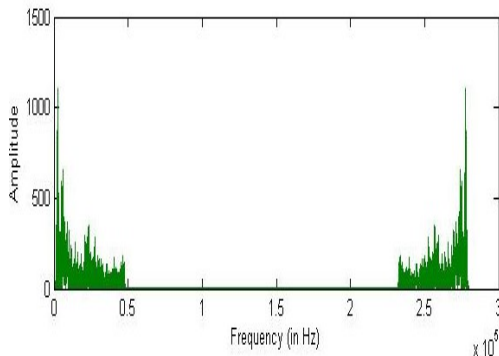

Fig. 6.(a) FFT of Normal Speech

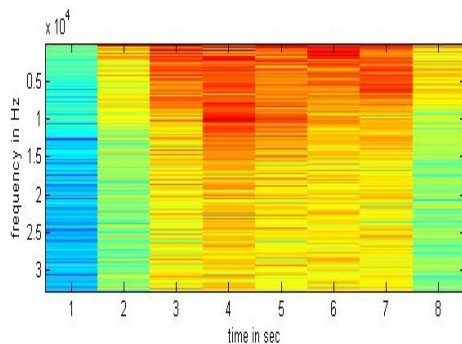Fig. 6.(b) FFT of Stressed Speech

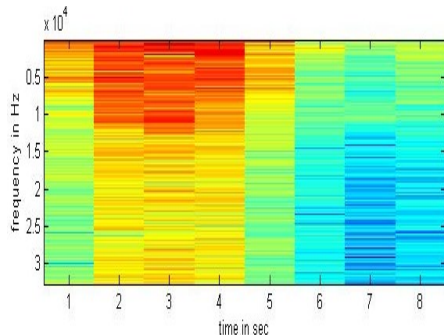Fig. 6. (c) Spectrogram of Normal Speech

Fig. 6.(d) Spectrogram of Stressed Speech

VII. CONCLUSIONS

In this study, we have tried to distinguish between the normal speech and the stressed speech using the parameters of time domain as well as frequency domain. In time domain, we have used parameters like energy, teager energy, autocorrelation and zero crossing rate. In addition to it, we have used fast fourier transform and spectrogram in the frequency domain. The autocorrelation of the normal speech was more appreciable than the stressed speech which states that the normal speech was more predictable than the stressed one. In the frequency domain analysis, we analysed the FFT and spectrogram of the speech signal. The amplitude as well as the frequency content of stressed speech was much greater than the normal speech.

REFERENCES

[1] 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).

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

[3] M. Sigmund.(2006)."Introducing the database ExamStress for speech under stress," Proceedings of 7th IEEE NordicSignal Processing Symposium (NORSIG 2006). Reykjavik,(pp. 290-293).

[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).

[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.

[8] Flanagan, J.L.(1972). Speech Analysis, Synthesis and Perception.second ed.. Springer-Verlag, NY.

[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.

[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 en-

ergy 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).

[15] hou, G., Hansen, J.H.L. &Kaiser, J.F.(2001). Nonlinear featurebased classification of speech under stress. IEEE Trans.Speech Audio Processing 9 (3), (pp.201–216).

[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. Matĕjka.(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] .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

[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)

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

❖ ❖ ❖