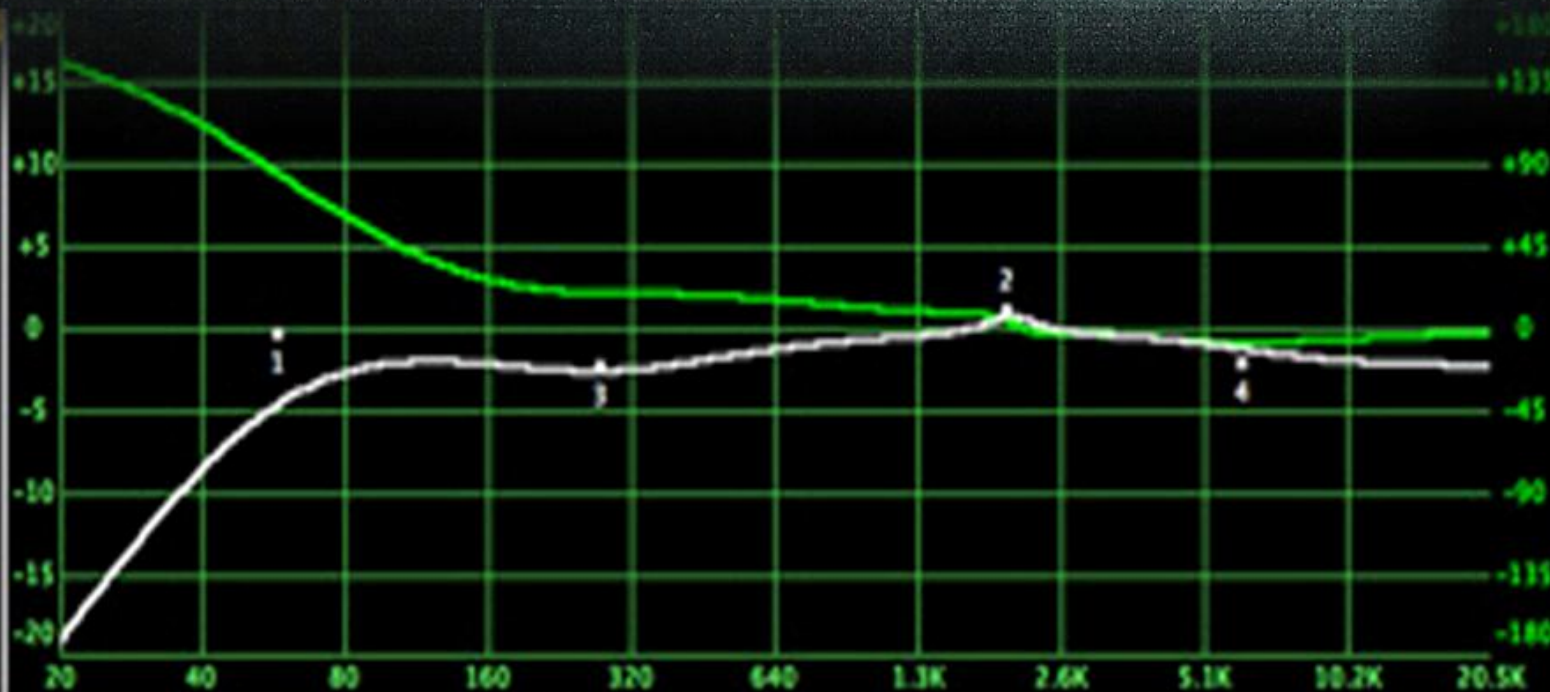


Signal Processing: An International Journal (SPIJ)

ISSN : 1985-2339

VOLUME 3, ISSUE 5

PUBLICATION FREQUENCY: 6 ISSUES PER YEAR



Editor in Chief Dr Saif alZahir

Signal Processing: An International Journal (SPIJ)

Book: 2009 Volume 3, Issue 5

Publishing Date: 31 - 11 - 2009

Proceedings

ISSN (Online): 1985 - 2339

This work is subjected to copyright. All rights are reserved whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication of parts thereof is permitted only under the provision of the copyright law 1965, in its current version, and permission of use must always be obtained from CSC Publishers. Violations are liable to prosecution under the copyright law.

SPIJ Journal is a part of CSC Publishers

<http://www.cscjournals.org>

©SPIJ Journal

Published in Malaysia

Typesetting: Camera-ready by author, data conversion by CSC Publishing Services – CSC Journals, Malaysia

CSC Publishers

Table of Contents

Volume 3, Issue 5, November 2009.

Pages

- 83 – 94 Investigating Multifractality of Solar Irradiance Data through Wavelet Based Multifractal Spectral Analysis.
K. Mofazzal Hossain, Dipendra N. Ghosh, Koushik Ghosh.
- 95 - 109 A Template Matching Approach to Classification of QAM Modulation using Genetic Algorithm.
Negar Ahmadi, Reza Berangi.
- 110 - 119 Moving One Dimensional Cursor Using Extracted Parameter from Brain Signals.
Siti Zuraimi Salleh, Norlaili Mat Safri, Siti Hajar Aminah Ali.
- 132 - 143 Integrated DWDM and MIMO-OFDM System for 4G High Capacity Mobile Communication.
Shikha Nema, Dr Aditya Goel, Dr R P Singh.
- 144 - 152 A New Method for Pitch Tracking and Voicing Decision Based on Spectral Multi-scale Analysis
Mohamed Anouar Ben Messaoud, Aïcha Bouzid, Nouredine Ellouze.
- 153 – 160 Face Recognition using Neural Networks.
P.Latha, Dr.L.Ganesan, Dr.S.Annadurai.

161 – 171 Arabic Phoneme Recognition using Hierarchical Neural Fuzzy
Petri Net and LPC Feature Extraction. Ghassaq S. Mosa,
Abduladhem Abdulkareem Ali.

Investigating Multifractality of Solar Irradiance Data through Wavelet Based Multifractal Spectral Analysis

K. Mofazzal Hossain

email: hossainkm_1976@yahoo.co.in

*Assistant Professor,
Department of Electronics & Instrumentation Engineering
Dr.B.C.Roy Engineering College, Durgapur
Durgapur-713206, West Bengal, India*

Dipendra N. Ghosh

email: ghoshdipen2003@yahoo.co.in

*Associate Professor, Department of Mathematics
Dr.B.C.Roy Engineering College, Durgapur
Durgapur-713206, West Bengal, India*

Koushik Ghosh

email: koushikg123@yahoo.co

*Lecturer, Department of Mathematics
University Institute of Technology, University of Burdwan
Burdwan-713104, West Bengal, India*

Abstract

It has been already revealed that the daily Solar Irradiance Data during the time period from October, 1984 to October, 2003 obtained by Earth Radiation Budget Satellite (ERBS) exhibits an Anti-persistent trend having multi-periodic phenomena. The solar irradiance time series data being a complex non linear signal in this paper we have tried to detect the irregularity and multifractality in the signal using continuous wavelet transform modulus maxima (WTMM) algorithm. Singularity spectrum of the signal has been obtained to measure the degree of multifractality of the Solar Irradiance signal. The qualitative measure of the degree of multifractality of the Solar Irradiance signal will help us to decide the nature of the signal processing tools that can be used to extract the features of the signal in our future work. This may also give an input to the research work of researchers on the solar physics and geophysics.

Keywords: ERBS, Wavelet transform, WTMM, scaling exponent, multifractal dimension, Hölder exponent, singularity spectrum

1. INTRODUCTION

Total solar irradiance describes the electromagnetic radiant energy emitted by the sun over all wavelengths that falls each second on 1 square meter outside the earth's atmosphere. Solar refers to electromagnetic radiation in the spectral range of approximately 1–9ft (0.3–3m), where the shortest wavelengths are in the ultraviolet region of the spectrum, the intermediate wavelengths in the visible region, and the longer wavelengths are in the near infrared. Total solar irradiance means that the solar flux has been integrated over all wavelengths to include the contributions from ultraviolet, visible, and infrared radiation. The solar irradiance had been monitored with absolute radiometers since November 1978, on board six spacecraft (Nimbus-7,

SMM, UARS, ERBS, EURECA, and SOHO), outside the terrestrial atmosphere (Fröhlich and Lean, 1998). Before measuring it from space, this quantity was thought to be constant, because the precision of the ground-based instruments at that time was not high enough to detect such a small variation. It consequently got the name of “solar constant”, which had a value of only $1,353 \text{ W/m}^2$, as a part of the solar radiation is absorbed by the Earth’s atmosphere. But from the data sent by the mentioned spacecraft it reveals that the solar irradiance varies about a small fraction of 0.1% over solar cycle being higher during maximum solar activity conditions. [1]

It is suggested that the solar variability is due to the perturbed nature of the solar core and this variability is provided by the variability of the solar neutrino flux from the solar neutrino detectors i.e., Homestake, Superkamiokande, SAGE and GALLEX-GNO. A major part of the Solar Irradiance variation is explained as a combined effect of the sunspots blocking and the intensification due to bright faculae and plages, with a slight dominance of the bright features effect during the 11-year solar cycle maximum. Solar Irradiance variation within solar cycle is thought to be due to the changing emission of bright magnetic elements, including faculae and the magnetic network. [2]

It has been revealed that the variation of the solar irradiance is anti-persistent and shows multi-periodicity. [3] The periods of the solar irradiance variation detected are 9.08-9.35, 13.53-14.03, 27.50-28.17, 30.26, 35.99-36.37, 51.14-51.52, 68.27-68.60, 101.15, 124.85, 150.63-153.98, 659.90, 729.37, 1259.82, 3464.50 and 4619.33 days.[4]. In this paper we would like to characterize the complex behaviour of the solar irradiance fluctuation by i) tracing the existence of multifractality and ii) scanning the singularities of the time series signal. Here we have computed the signal parameters like scaling exponents $\tau(q)$, multifractal scaling exponents $h(q)$ and generalized multifractal dimensions $D(q)$ which quantifies the multifractality of the signal. For tracking the singularities in the time series signal we have computed the singularity strength or Hölder exponent (α) and obtained the Hausdorff dimension or singularity spectrum $f(\alpha)$. The use of monofractal methods to extract quantitative information from signals is well known. Monofractals are homogeneous objects, in the sense that they have the same scaling properties, characterized by a single singularity exponent. Generally, there exist many observational signals which do not present a simple monofractal scaling behaviour. The need for more than one scaling exponent can derive from the existence of a crossover timescale, which separates regimes with different scaling behaviours. Different scaling exponents could be required for different segments of the same time series, indicating a time variation of the scaling behaviour. Furthermore, different scaling exponents can be revealed for many interwoven fractal subsets of the time series; in this case the process is not a monofractal but multifractal. Thus, multifractals are intrinsically more complex and inhomogeneous than monofractals and characterize systems featured by very irregular dynamics, with sudden and intense bursts of high-frequency fluctuations. The simplest type of multifractal analysis is given by the standard partition function multifractal formalism, developed to characterize multifractality in stationary measures. This method does not correctly estimate the multifractal behaviour of signal affected by trends or non-stationarities. But the solar irradiance time series signal is non stationary in nature. To analyze non-stationary signal wavelet transform based tool are more suitable compared to the traditional Fourier based tools [5]. Hence to characterize the multifractality of non-stationary signals another multifractal method based on the wavelet analysis named as Wavelet Transform Modulus Maxima (WTMM) method is being used in this paper. [6, 7] This method involves tracing the maxima lines in the continuous wavelet transform over all scales. WTMM allows one to detect scaling by means of the maxima lines of the continuous wavelet transform on different scales.

2. THEORY

CONTINUOUS WAVELET TRANSFORM

The continuous wavelet transform (WT) is a mathematical technique introduced in signal analysis in the early 1980s. Since then, it has been the subject of considerable theoretical developments and practical applications in a wide variety of fields. The WT has been early recognized as a mathematical microscope that is well adapted to reveal the hierarchy that governs the spatial distribution of singularities of multifractal measures. The wavelet transform is

a convolution product of the data sequence (a function $f(x)$, where x , referred to as “position”, is usually a time or space variable. In this study x is referred as time (t) and hence the data sequence is time series) with the scaled and translated version of the mother wavelet, $\psi(x)$. The scaling and translation are performed by two parameters; the scale parameter s stretches (or compresses) the mother wavelet to the required resolution, while the translation parameter b shifts the analyzing wavelet to the desired location:

$$Wf(s, b) = \frac{1}{s} \int_{-\infty}^{\infty} f(x) \psi\left(\frac{x-b}{s}\right) dx, \quad (1)$$

where s, b are real, $s > 0$ for the continuous version (CWT). $Wf(s, b)$ are the wavelet transform coefficients. The wavelet transform acts as a microscope: it reveals more and more details while going towards smaller scales, i.e. towards smaller s values [8].

The mother wavelet $\psi(x)$ is generally chosen to be well localized in space (or time) and frequency. Usually, $\psi(x)$ is only required to be of zero mean, but for the particular purpose of multifractal analysis $\psi(x)$ is also required to be orthogonal to some lower order polynomials, up to the degree n :

$$\int x^m \psi(x) dx = 0, \quad \forall m, \quad 0 \leq m < n \quad (2)$$

Thus, while filtering out the trends, the wavelet transform can reveal the local characteristics of a signal, and more precisely its singularities. The Hölder exponent can be understood as a global indicator of the local differentiability of a function.

By preserving both scale and location (time, space) information, the CWT is an excellent tool for mapping the changing properties of non-stationary signals. A class of commonly used real-valued analyzing wavelets, which satisfies the above condition (2), is given by the successive derivatives of the Gaussian function:

$$\psi^{(n)}(x) = \frac{d^n}{dx^n} e^{-x^2/2} \quad (3)$$

Note that the WT of a signal $f(x)$ with $\psi^{(n)}(x)$ in Eq. (3) takes the following simple expression:

$$\begin{aligned} Wf(s, b) &= \frac{1}{s} \int_{-\infty}^{\infty} f(x) \psi^{(n)}\left(\frac{x-b}{s}\right) dx, \\ &= s^n \frac{d^n}{dx^n} Wf(s, b) \end{aligned} \quad (4)$$

Equation (4) shows that the WT computed with $\psi^{(n)}(x)$ at scale s is nothing but the n^{th} derivative of the signal $f(x)$ smoothed by a dilated version $\psi^{(0)}(x/s)$ of the Gaussian function. This property is at the heart of various applications of the WT microscope as a very efficient multi-scale singularity tracking technique. Thus, the higher derivatives, the more vanishing moments, that is, the local polynomial trends of higher order would be eliminated. We choose the third derivative of a Gaussian

$$\psi^{(3)}(x) = \frac{d^3}{dx^3} e^{-x^2/2} \quad (5)$$

which is insensitive to trends up to a quadratic one.

WAVELET TRANSFORM MODULUS MAXIMA (WTMM)

The WTMM method inherits the advantages of the wavelet transform analysis and was developed to deal with strongly non-stationary data. It has an important ability to reveal hierarchical structure of singularities and therefore proves useful in analyzing self-similar structures like fractals. In

small-scale levels s of wavelet transform, sharp hidden transitions (singularities) in Solar Irradiance dynamics would be extracted.

The continuous wavelet transform described in Eq. (1) is an extremely redundant representation, too expensive for most practical applications. To characterize the singular behaviour of functions, it is sufficient to consider the values and position of the Wavelet Transform Modulus Maxima (WTMM). The wavelet modulus maxima is a point (s_0, x_0) on the scale-position (or time) plane, (s, x) , where $|Wf(s_0, x)|$ is locally maximum for x in the neighborhood of x_0 . These maxima are disposed on connected curves in the scale position (s, x) (or scale-time) half-plane, called *maxima lines*. An important feature of these maxima lines, when analyzing singular functions, is that there is at least one maxima line pointing towards each singularity. The WTMM representation has been used for defining the partition function based multifractal formalism.

Let $\{u_n(s)\}$, where n is an integer, be the position (time) of all local maxima at a fixed scale s . By summing up the q 's power of all these WTMM, we obtain the partition function Z : [9]

$$Z(q, s) = \sum_n |Wf(s, u_n)|^q \quad (6)$$

where q can be any real value except zero.

TRACING SINGULARITIES

The rapid changes in a time series $f(x)$ are called singularities and a characterization of their strength is obtained with the Hölder exponents. The strength of the singularity of a function $f(x)$ at point x_0 is given by the Hölder exponent α , i.e., the largest exponent such that $f(x)$ is Lipschitz at x_0 . There exists a polynomial $P_n(x - x_0)$ of order n and a constant C , so that for any point x in a neighborhood of x_0 , one has:

$$|f(x) - P_n(x - x_0)| \leq C |x - x_0|^\alpha \quad (7)$$

where $n \leq \alpha(x_0)$ and $C > 0$.

The Hölder exponent measures the degree of irregularity of $f(x)$ at the point x_0 . When a broad range of exponents is found, signals are considered as multifractal. A narrow range implies monofractality. Let us assume that according to Eq.(7), $f(x)$ has, at the point x_0 , a local scaling (Hölder) exponent $\alpha(x_0)$; then, assuming that the singularity is not oscillating, one can easily prove that the local behaviour of $f(x)$ is mirrored by the WT which locally behaves as per the power law:

$$Wf(s, x_0) \sim s^{\alpha(x_0)}, \quad (8)$$

Taking the log-log plot on both sides of the Eq. (8) Hölder exponent α can be estimated. A very important point (at least for practical purpose) rose by Mallat and Hwang is that the local scaling exponent $\alpha(x_0)$ can be equally estimated by looking at the value of the WT modulus along a maxima line converging towards the point x_0 . Indeed one can prove that Eqs. (8) still holds when following a maxima line from large down to small scales. Depending on the value of $\alpha(x_0)$ at every x_0 we can scan the points of irregularity (opposite of regularity) or singularity.

If $\alpha(x_0)$ is	Regularity of $f(x)$ at x_0	Singularity of $f(x)$ at x_0
Higher	More	Less
Lower	Less	More

MULTIFRACTAL ANALYSIS

A natural way of performing a multifractal analysis of a function lies in generalizing the multifractal formalism using wavelets. From the deep analogy that links the multifractal formalism to thermodynamics [10], one can define the scaling exponent $\tau(q)$ from the power-law behavior of the partition function as given in Eqs (6):

$$Z(q,s) \sim s^{\tau(q)} \quad (9)$$

Here we have varied the value of q from -20 to 20 with an increment of 0.2. Taking the log of the Eq.(9), $\tau(q)$ is being estimated for each value of q . The singularity spectrum $f(\alpha)$ is related to $\tau(q)$ by Legendre Transform as follows: a) from the plot of $\tau(q)$ vs. q the Hölder exponents α as a function of q can be determined from the relationship:

$$\alpha(q) = \frac{d\tau(q)}{dq} \quad (10)$$

b) Singularity spectrum $f(\alpha)$ is calculated from the equation

$$f(\alpha) = q\alpha - \tau(q) \quad (11)$$

From the properties of the Legendre transform, it is easy to see that *homogeneous* mono-fractal functions that involve singularities of unique Hölder exponent $\alpha(q)$ are characterized by a $\tau(q)$ spectrum which is a *linear* function of q . On the contrary, a *nonlinear* $\tau(q)$ curve is the signature of non-homogeneous functions that exhibit *multifractal* properties, in the sense that the Hölder exponent $\alpha(q)$ is a fluctuating quantity. The singularity spectrum $f(\alpha)$ of a multifractal function displays a single humped shape that characterizes intermittent fluctuations corresponding to Hölder exponent values spanning a whole interval $[\alpha_{\min}, \alpha_{\max}]$, where α_{\min} and α_{\max} are the Hölder exponents of the strongest and weakest singularities respectively.

Other than the signal parameters like scaling exponent $\tau(q)$, Hölder exponents $\alpha(q)$ and singularity spectrum $f(\alpha)$ as described above, multifractality can also be detected from the *multifractal scaling exponent or generalized Hurst exponent* $h(q)$ and the *generalized multifractal dimension* $D(q)$. Both $h(q)$ and $D(q)$ can be calculated from the scaling exponent $\tau(q)$ as below:

$$h(q) = \frac{1 + \tau(q)}{q}, \quad q \neq 0 \quad (12)$$

and

$$D(q) = \frac{\tau(q)}{q-1} = \frac{qh(q)-1}{q-1}, \quad q \neq 1 \quad (13)$$

For monofractal time series $h(q)$ is independent of q whereas $D(q)$ depends on q . But for multifractal time series there is significant dependence of $h(q)$ on q . If q is positive, large fluctuations are characterized by a smaller values of $h(q)$ for multifractal time series. And, for negative q values, small fluctuations are usually characterized by larger values of $h(q)$.

From Eq.10, 11 and 12 Hölder exponent $\alpha(q)$ and Singularity spectrum $f(\alpha)$ can also be expressed in terms of the multifractal scaling exponent $h(q)$ as follows:

$$\alpha = h(q) + q \frac{dh(q)}{dq} \quad (14)$$

and

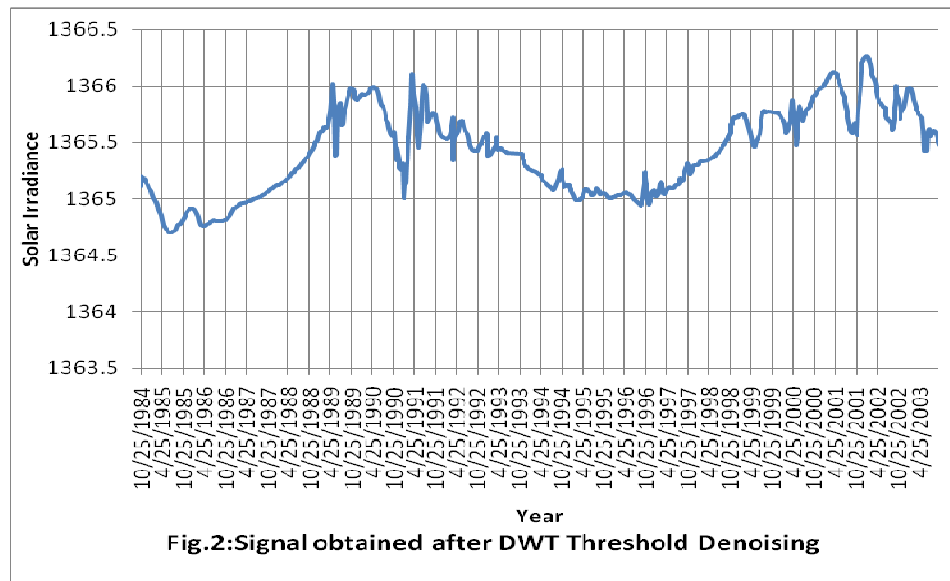
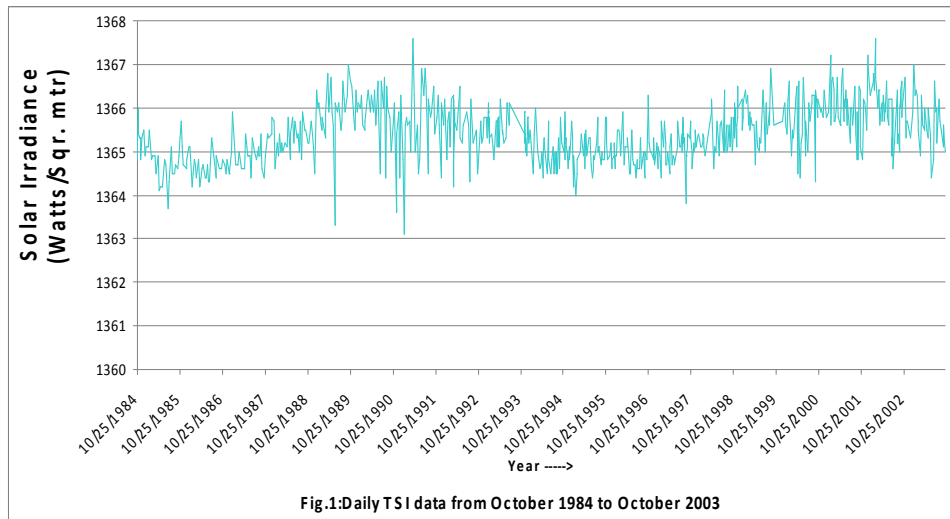
$$f(\alpha) = q[\alpha - h(q)] + 1 \quad (15)$$

Here we like to mention that multifractal scaling exponent or generalized Hurst exponent $h(q)$ is related to Hurst exponent H by the equation

$$H = h(q=2) - 1 \quad (16)$$

3. RESULTS

Fig.1 represents the original signal of the daily Solar Irradiance from October, 1984 to October, 2003 obtained by ERBS after simple exponential smoothing which is being denoised using DWT thresholding and the denoised signal is obtained as in fig.2.[3].



CWT, $Wf(s,b)$ of this data is being taken. The absolute values of the coefficients i.e. $|Wf(s,b)|$ is plotted with color coding, independently at each scale s , using 128 colors from deep brown ($|Wf(s,b)| = 0$) to white ($\max |Wf(s,b)|$) as shown in fig.3. Scale and time are on the vertical and horizontal axis, respectively. The plot was obtained by using the “Wavelet toolbox” of Matlab software.

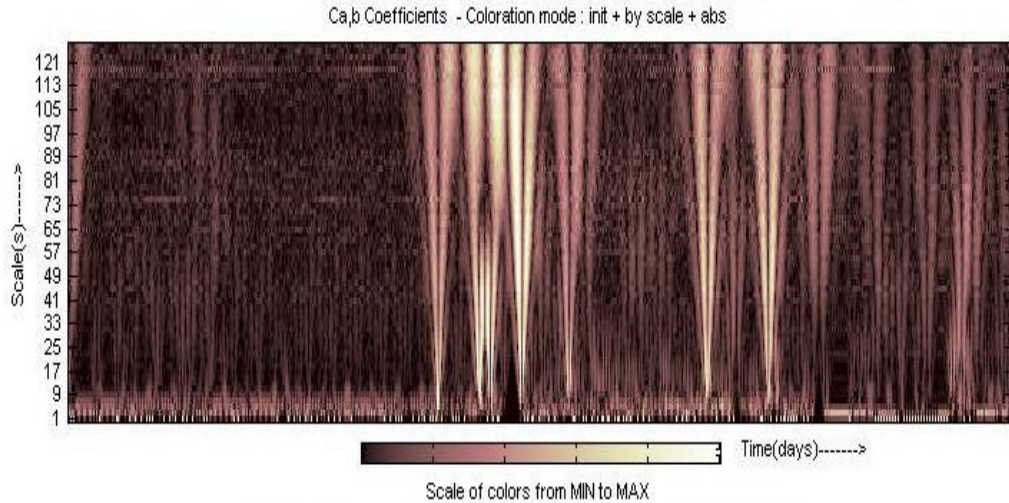


Fig.3: Colour coded CWT coefficient plot (Scale(s) Vs. Time)

Fig.4 represents the WT skeleton defined by the set of all maxima lines.

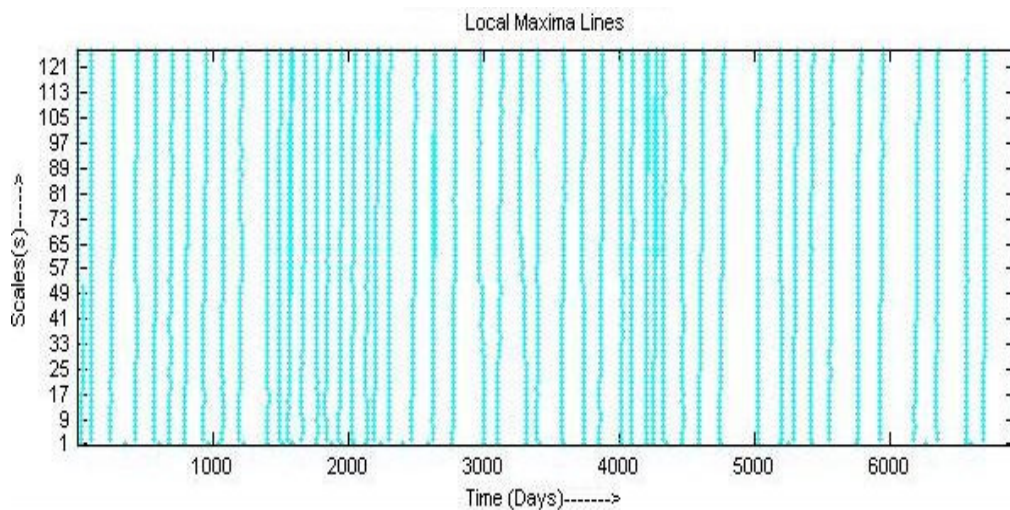
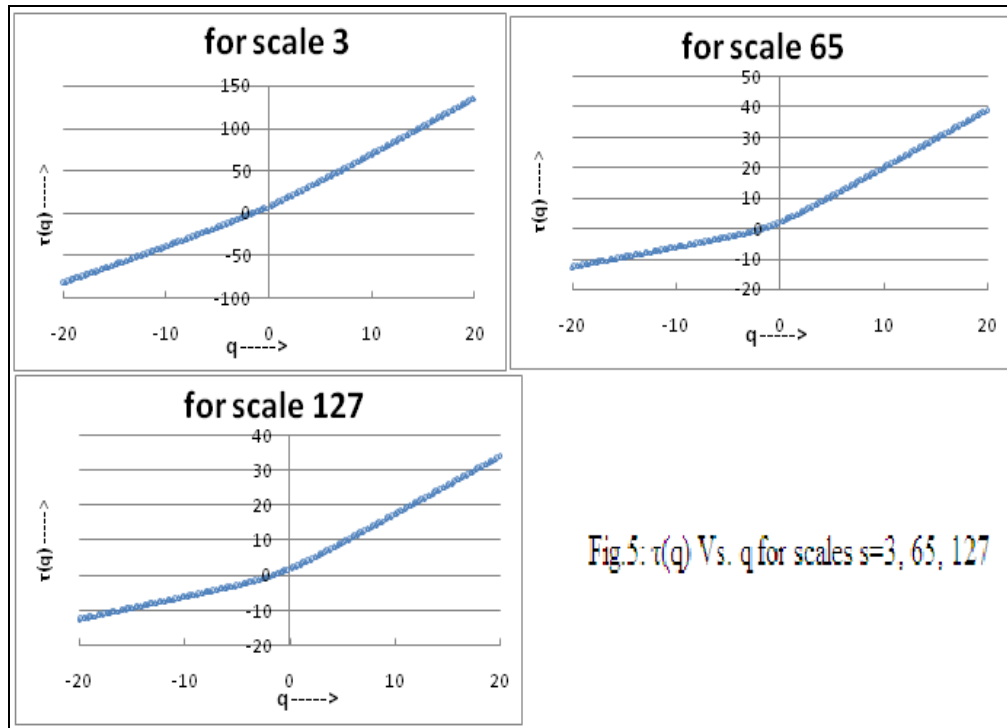


Fig.4: WTMM Skeleton

The plot of $\tau(q)$ vs q for the scale, $s=3, 65, 127$ are being shown in fig.5



The singularity spectrum i.e. $f(\alpha)$ vs. α for the scales $s=3, 65, 127$ is represented in fig.6.

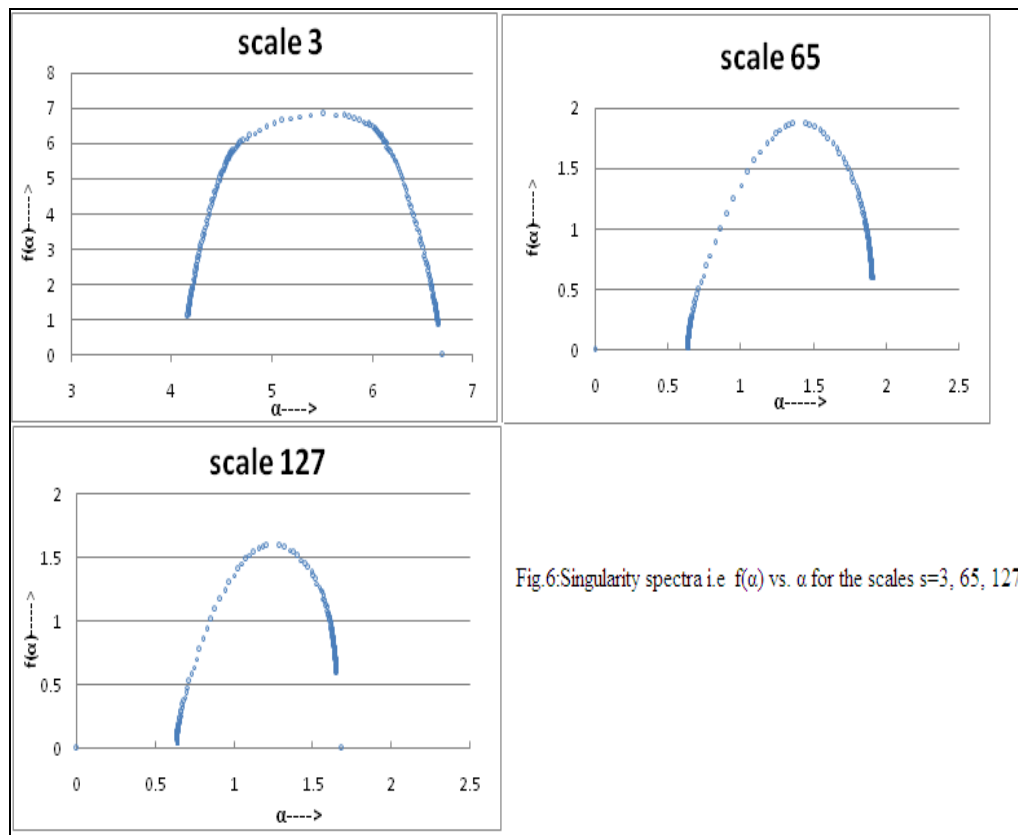


Fig.7 represents the $D(q)$ vs. q curve for the scales $s=3,65$ and 127 as shown below.

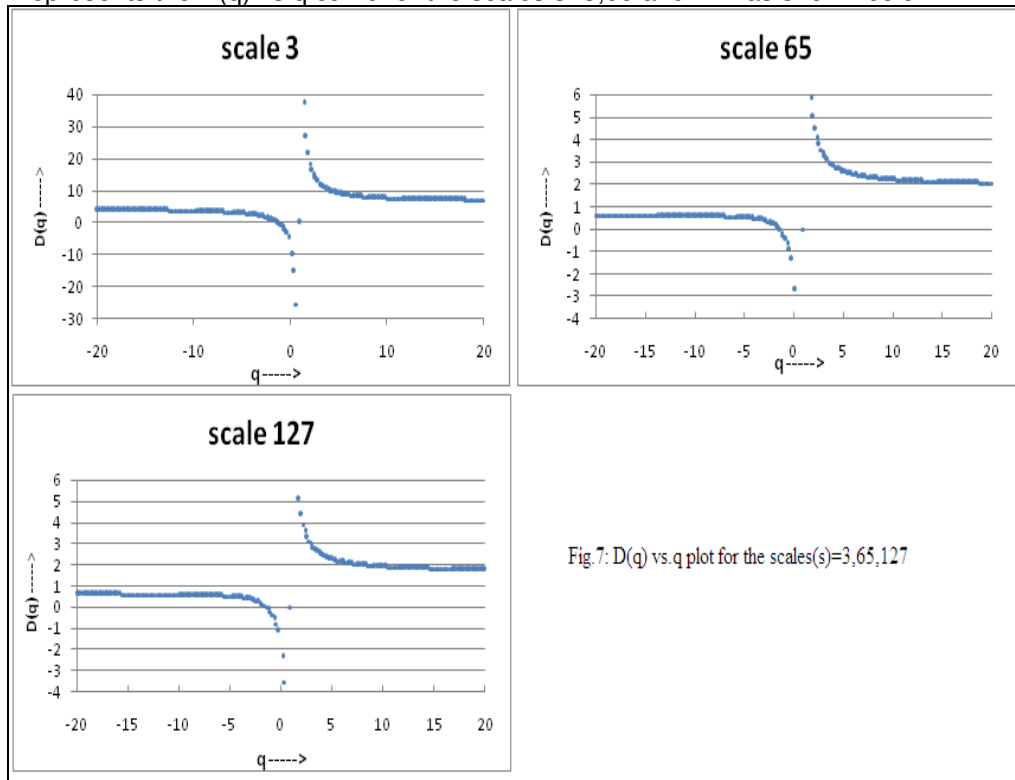


Fig. 7: $D(q)$ vs. q plot for the scales $(s)=3,65,127$

Fig.8 represents the $h(q)$ vs. q curve for the scales $s=3,65$ and 127 as shown below.

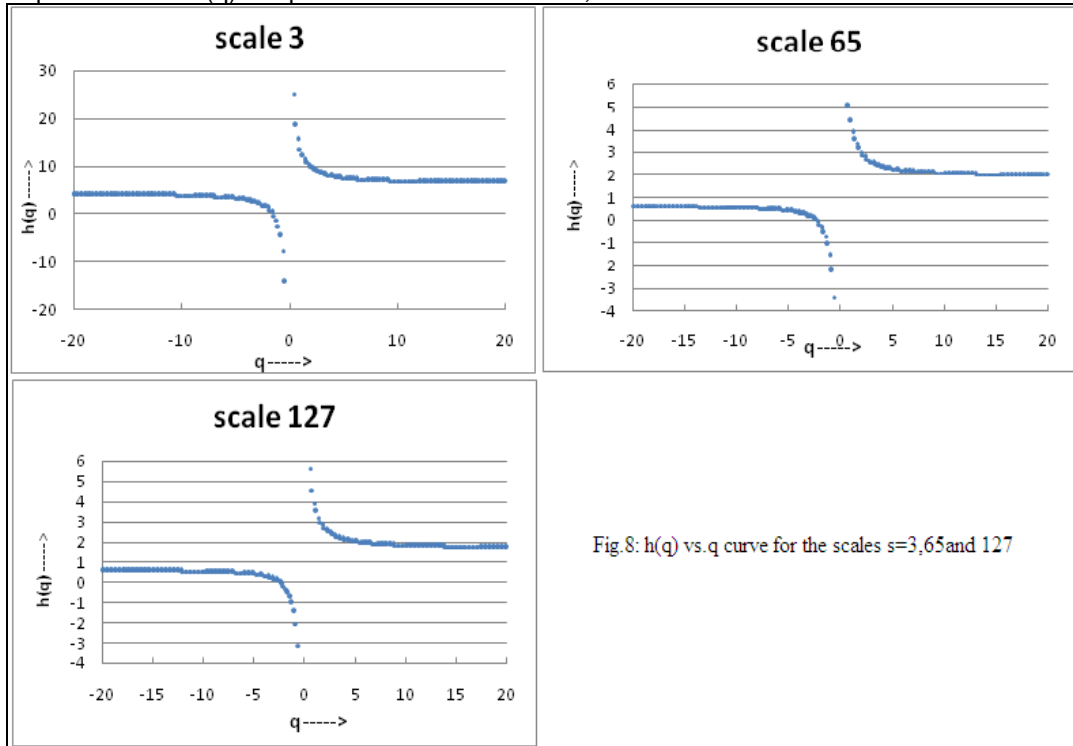


Fig. 8: $h(q)$ vs. q curve for the scales $s=3,65$ and 127

The plot of $D(h)$ vs. h for the scale, $s=3, 65, 127$ are being shown in fig.9.

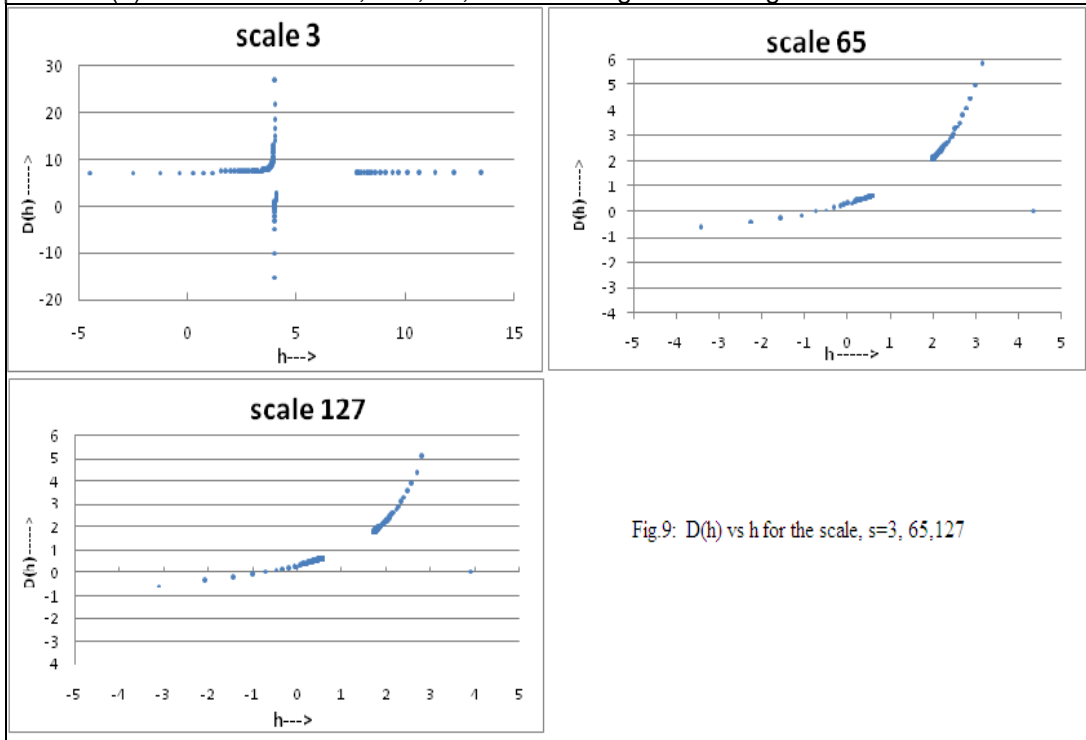
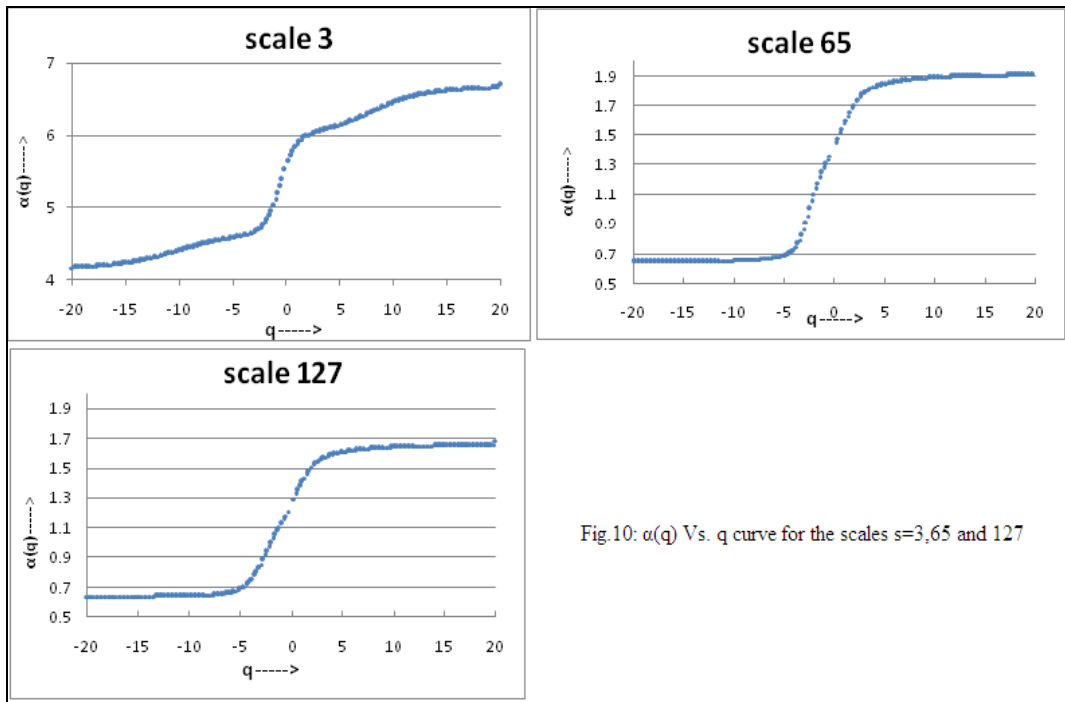


Fig.10: $\alpha(q)$ vs. q curve for the scales $s=3, 65$ and 127 as shown below



4. CONCLUSIONS

WTMM method allows us to determine the multifractal characterization of the nonstationary solar irradiance time series. The concept of WTMM of the solar irradiance time series is used here to have a deeper insight into the process occurring in nonstationary dynamical system such as multi-periodic fluctuation in solar irradiance values. The dependency of the $\tau(q)$ and $h(q)$ on q as observed in fig.5 and fig.8, indicates that the solar irradiance variation has multifractal behavior. This behavior of exhibiting multifractal characteristics can be more established from the singularity spectrum as in fig.6. The multifractal analysis gives information about the relative importance of various fractal exponents present in the series. In particular, the width of the singularity spectrum indicates the range of present exponents. To get the quantitative characterization of multifractal spectra, the singularity spectrum is fitted to a quadratic function around the position of its maximum at α_0 , i.e. $f(\alpha) = A(\alpha - \alpha_0)^2 + B(\alpha - \alpha_0) + C$. The coefficients can be obtained by an ordinary least-squares procedure. [11] In this fitting the additive constant $C = f(\alpha_0)$. With low α_0 , the process becomes correlated; for example if the process had the tendency to move upward in the past, it will move upward with a probability larger than 1/2 in the next time step. Roughly speaking, a small value of α_0 means that the underlying process is more regular in appearance. From the fig.6 we observe that the value of α_0 is very high for lower scales and decreases with increase in the scale. It means that the signal is correlated at higher scales.

To obtain an estimate of the range of possible fractal exponents, we measured the width of the singularity spectrum, extrapolating the fitted curve to zero. The width of the spectrum was then defined as $W = \alpha_{\max} - \alpha_{\min}$ with $f(\alpha_{\max}) = f(\alpha_{\min}) = 0$. The width of the spectrum W is a measure of how wide the range of fractal exponents found in the signal and thus it measures the degree of multifractality of the series. The wider the range of possible fractal exponents, the 'richer' is the process in structure. From the fig.6 we observe that W is decreasing with increase in the scale size i.e. solar irradiance signal is richer in structure at lower scales.

Finally, parameter B serves as an asymmetry parameter, which is zero for symmetric shapes, positive or negative for a left- or right-skewed (centered) shape, respectively. B captures the dominance of low- or high-fractal exponents with respect to the other. A right-skewed spectrum indicates relatively strongly weighted low-fractal exponents, and for left-skewed spectrum indicates relatively strongly weighted high-fractal exponents. From fig.6 we observe that for scale 65 and 127 the singularity spectrum is left skewed whereas for scale 3 the singularity spectrum is more or less symmetrical. Hence we can say that with increasing scales the signal is found to have high fractal exponents. The parameter scale(s) in the wavelet analysis also has a significant role. The high scales correspond to a non-detailed global view (of the signal), whereas the low scales correspond to a detailed view. Similarly, in terms of frequency, low frequencies (high scales) correspond to a global information of a signal (that usually spans the entire signal), whereas high frequencies (low scales) correspond to a detailed information of a hidden pattern in the signal (that usually lasts a relatively short time). So the above discussion regarding the values of α_0 , W , B at various scales give a measure of the detailed or non-detailed global view of the signal.

5. REFERENCES

1. A.Oncica, M. D.Popescu, M. Mierla, G. Maris, "Solar Variability: From Core to Outer Frontiers" In Proceedings of the 10th European Solar Physics Meeting Proc. ESA, SP-506, Prague, 193, 2002.
2. P. Raychaudhuri, "Total Solar Irradiance Variability and the Solar Activity Cycle", <http://arxiv.org/ftp/astro-ph/papers/0601/0601335.pdf>.
3. K. M. Hossain, D. N. Ghosh and K. Ghosh, "Scaling Analysis by FVSM and DWT denoising of the measured values of solar irradiance", accepted for publication in International Journal of Information and Computing Science(IJICS), Kolkata, Dec.2009
4. K. M. Hossain, D. N. Ghosh and K. Ghosh, " Search for Periodicities in Solar Irradiance data from ERBS ", communicated to Solar Physics, US, June 2009

5. Othman O. Khalifa, Sering Habib Harding & Aisha-Hassan A. Hashim, "*Compression Using Wavelet Transform*", Signal Processing: An International Journal(SPIJ), Volume (2) : Issue (5),17-26,Malaysia,September/October,2008
6. J.F. Muzy, E. Bacry, A. Arneodo,"*Wavelets and multifractal formalism for singular signals: Application to turbulence data*", Physical Review. Lett. 67, 3515-3518, 1991.
7. A. Arneodo, E. Bacry, J.F. Muzy," *The thermodynamics of fractals revisited with wavelets*" Physica A 213, 232-275 ,1995
8. Mohamed Nerma, Nidal Kamel, and Varun Jeoti," *An OFDM System Based on Dual Tree Complex Wavelet Transform (DT-CWT)*", Signal Processing: An International Journal (SPIJ),Volume (3) : Issue (2),14-26,Malaysia, March/April,2009
9. Bogdan ENESCU, Kiyoshi ITO, and Zbigniew R. STRUZIK, "*Wavelet-Based Multifractal Analysis of real and simulated time series of earthquakes*", Annuals of Disas. Prev. Res. Inst., Kyoto Univ., No. 47 B,2004
10. H.Eugene Stanley, Paul Meakin,"*Multifractal Phenomena in Physics and Chemistry*", Nature Vol.335, page 405-409,1988
11. Luciano Telesca, Vincenzo Lapenna and Maria Macchiato, "*Multifractal fluctuations in earthquake-related geo-electrical signals*", New J. Phys. 7, 214,2005

A Template Matching Approach to Classification of QAM Modulation using Genetic Algorithm

Negar Ahmadi

*Department of Computer Engineering
Iran University of Science and Technology
Narmak, Tehran, Post Code: 1684613114, Iran*

negar.ahmadi670@gmail.com

Reza Berangi

*Department of Computer Engineering
Iran University of Science and Technology
Narmak, Tehran, Post Code: 1684613114, Iran*

rberangi@iust.ac.ir

Abstract

The automatic recognition of the modulation format of a detected signal, the intermediate step between signal detection and demodulation, is a major task of an intelligent receiver, with various civilian and military applications. Obviously, with no knowledge of the transmitted data and many unknown parameters at the receiver, such as the signal power, carrier frequency and phase offsets, timing information, etc., blind identification of the modulation is a difficult task. This becomes even more challenging in real-world.

In this paper modulation classification for QAM is performed by Genetic Algorithm followed by Template matching, considering the constellation of the received signal. In addition this classification finds the decision boundary of the signal which is critical information for bit detection. I have proposed and implemented a technique that casts modulation recognition into shape recognition. Constellation diagram is a traditional and powerful tool for design and evaluation of digital modulations. The simulation results show the capability of this method for modulation classification with high accuracy and appropriate convergence in the presence of noise.

Keywords: Automatic Modulation Recognition, Genetic Algorithm, Constellation Diagram, Template Matching.

1. INTRODUCTION

Recognition of the modulation type of an unknown signal provides valuable insight into its structure, origin and properties. Automatic modulation classification is used for spectrum surveillance and management, interference identification, military threat evaluation, electronic counter measures, source identification and many others. For example, if the modulation type of an intercepted signal is extracted, jamming can be carried out more efficiently by focusing all resources into vital signal parameters. Other applications may include signal source

identification. This is particularly applicable to wireless communications where different services follow well known modulation standards.

There is another usage for both urban and military applications and recently has attracted many attention that is making possible to build Intelligent receivers which can recognize the modulation type without having any prior information from the transmitting signal. Thus intelligent transmitters-receivers appears that can select the most appropriate modulation type to transmit the information due to the environmental condition and communicative channel, and also the receiver can recognize the changes of the modulation types immediately. Therefore, in the subject of the communication, transparency is developed due to the modulation type [1, 2, and 3].

Modulation is the process of varying a periodic waveform, i.e. a tone, in order to use that signal to convey a message. The most fundamental digital modulation techniques are: Amplitude Shift Keying (ASK), Frequency Shift Keying (FSK), Phase Shift Keying (PSK) and Quadrature - Amplitude Modulation (QAM). The QAM modulation is more useful and efficient than the others and is almost applicable for all the progressive modems.

Modulation recognition is an intermediate step on the path to full message recovery. As such, it lies somewhere between low level energy detection and a full fledged demodulation. Therefore, correct recovery of the message per se is not an objective, or even a requirement [4, 5]. The existing methods for modulation classification span four main approaches. Statistical pattern recognition, decision theoretic (Maximum Likelihood), M-th law non-linearity and filtering and ad hoc [6, 7].

Early on it was recognized that modulation classification is, first and foremost, a classification problem well suited to pattern recognition algorithms. A successful statistical classification requires the right set of features extracted from the unknown signal. There have been many attempts to extract such optimal feature. Histograms derived from functions like amplitude, instantaneous phase, frequency or combinations of the above have been used as feature vectors for classification, Jondral [8], Dominguez et al. [9], Liedtke [10]. Also of interest is the work of Aisbett [11] which considers cases with very poor SNR.

The current state of the art in modulation classification is the decision theoretic approach using appropriate likelihood functional or approximations thereof. Polydoros and Kim [12] derive a quasi-log-likelihood functional for classification between BPSK and QPSK modulations. In a later publication, Huang and Polydoros [13] introduce a more general likelihood functional to classify among arbitrary MPSK signals. They point out that the S-classifier of Liedtke, based on an ad hoc phase-difference histogram, can be realized as a noncoherent, synchronous version of their *qLLR*. Statistical Moment-Based Classifier (SMBC) of Solimon and Hsue [14] are also identified as special coherent version of *qLLR*. Wei and Mendel [15] formulate another likelihood-based approach to modulation classification that is not limited to any particular modulation class. Their approach is the closest to a constellation-based modulation classification advocated here although they have not made it the central thesis of their work. Carrier phase and clock recovery issues are also not addressed. Chugg et al [16] use an approximation of log-ALF to handle more than two modulations and apply it to classification between OQPSK/BPSK/QPSK. Lin and Kuo [17] propose a sequential probability ratio test in the context of hypothesis testing to classify among several QAM signals. Their approach is novel in the sense that new data continuously updates the evidence.

There have been other approaches to modulation classification. A method has been proposed by Ta [18] which uses the energy vectors derived from wavelet packet decomposition as feature vectors to distinguish between ASK, PSK and FSK modulation types.

Past work on modulation recognition has primarily used signal properties in time and/or frequency domain to identify the underlying modulation. One of the typical analysis methods for the modulated signal is the extraction of In-Phase (I) and Quad-Phase (Q) components.

According to these components, we can see the signal as a vector in the $I-Q$ plane which is referred to as the constellation diagram. With the use of modulated signal constellation, modulation classification can be investigated as pattern recognition problem and well known pattern recognition algorithms can be used.

2. SIGNAL TRAJECTORY AND CONSTELLATION

One of the best methods for classification of signal modulation is the use of signal trajectory and its constellation. Since each type of modulations has a unique constellation and signal trajectory recognition of modulation could be performed accurately.

This approach to the analysis of modulated signals is based on the extraction of the *in-phase* (I) and *quadrature* (Q) components of the signal, which are obtained through a suitable demodulator. This allows seeing the modulating signal as a vector in the $I-Q$ plane, whose measured trajectory is presented in a two-dimensional diagram. The two most common diagram types are:

Constellation: presents the values obtained by sampling the (I) and (Q) components at the time instants given by the receiver clock. A constellation diagram thus presents the actual received symbol values (Fig. 1.a);

Vector diagram: presents in the $I-Q$ plane the whole trajectory of the vector associated with the demodulated signal. To obtain a vector diagram the (I) and (Q) components must be sampled at a higher rate than the receiver clock rate (Fig. 1 .b).

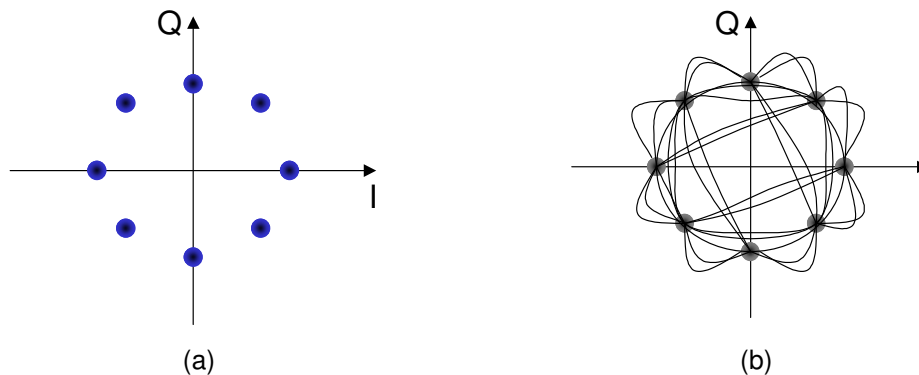


FIGURE 1: $I-Q$ diagrams: a) constellation; b) vector diagram.

From a measurement point of view, the main difference between the two diagrams lies in the different way of sampling the signal. To obtain a constellation diagram, the receiver clock must be available and determines the sampling instants. This may be provided by the system under test as an external clock input, or may be recovered by the measuring instrument from the analysed signal itself. In this way, by obtaining the number of clusters created in $I-Q$ plane, levels and type of modulation could be identified.

To the knowledge of author, one work on constellation diagram is reported by [6, 7], which is worked on fuzzy system. In [6, 7], fuzzy c-means clustering is used for initial processing but for final decision, it used a kind of template matching which uses a maximum likelihood approach.

In This paper proposes a Genetic Algorithm based approach to recognize QAM and PSK modulation schemes from their symbols. It classifies the symbols using the Genetic Algorithm and

hierarchical clustering to find the natural clusters which are equal to the modulation levels. The main advantage of this approach is that the decision boundaries can be accurately identified. The simulation results show that the proposed method has a high success rate at the presence of noise and can be easily applied over the single dimensional modulation schemes such as FSK.

3. CLASSIFICATION OF QAM AND PSK MODULATION USING GENETIC ALGORITHM (GA)

Genetic algorithm is a special kind of evolution algorithms that populations of chromosomes which are nominees to solve the problem are finally led to a better solution. The hypothesis beings with a complete random population and continues in generations. In each generation, all the members are evaluated. Several chromosomes are selected from the current generation based on the worthiness and modified to form the new generation and in the next repeat of algorithm, will be the current generation. In every problem to find the genetic algorithm to obtain an answer, we need two elements: first a method to represent an answer in such a way that genetic algorithm can apply that. Second a method that can calculate the proposed answer by using the fitness functions.

Modifiers selection is the most important part of the genetic algorithm. In fact, genetic algorithm searches the answer to find new answers by the genetic modifiers. This process is repeats for every member and then calls for genetic algorithms modifiers such as selection, cross over, reproduction and mutation forms the next generation. Producing of new generation will be continued to get the best solution. Generally, a genetic algorithm has a main five parts:

1. member presentation in genetic algorithm
2. a method of creation a first population
3. evaluation function to determine the fitness rate of each member
4. genetic modifiers which causes to combine the genetic structure of offspring during reproduction
5. numbers related to the genetic algorithm parameters.

As mentioned, constellation diagram, which consists of In-phase and Quad-phase components, can be used for modulation classification. Since the constellation is symmetric with respect to its axes, in order to reduce complexity, we can map all the received symbols into the first quadrant in the constellation diagram. After obtaining number and location of clusters in first quadrant, centroids of clusters could be extended to the whole constellation, symmetrically. The proposed technique has been designed so that it would be capable of recognizing MPSK and MQAM, so the initial number of clusters has been set to 16 in the first quadrant. Therefore, the initial centroids can be defined as a vector of 16 elements in which each element is a point in the first constellation quadrant. In order to reduce processing, calculation is done in the first quadrant, absolute value of signal's $I-Q$ components are calculated and stored in a 2D matrix and used in future processing.

This method proposes a Genetic Algorithm based approach to recognize QAM and PSK modulation schemes from their symbols. It classifies the symbols using the Genetic Algorithm and hierarchical clustering to find the natural clusters which are equal to the modulation levels. The main advantage of this approach is that the decision boundaries can be accurately identified. The simulation results show that the proposed method has a high success rate at the presence of noise and can be easily applied over the single dimensional modulation schemes such as FSK.

4. REGULATION OF GENETIC ALGORITHM PARAMETERS

In this step, the following parameters are regulated as the essential genetic algorithm parameters:

- Max-Gen parameter that determines the maximum time of reproducing and its number is selected 80 in this algorithm.
- Pop-Size parameter that represent the population size and number of chromosomes in members and is considered 300 here.
- P_c parameter that introduces the probability of two chromosomes crosses over and is considered 0.7 here.
- P_m parameter which shows the probability of mutation in genes of a chromosome and here is 0.15.

In this algorithm is tried to select the related points of a chromosome purposely with all page scattering and some random numbers are increased or decreased to the selected pointed of the first chromosome to select other points in the other chromosomes. By this method, we achieved the speed rise to the algorithm convergence.

To put a problem through GA, we should have a representation for solution of the problem. So, in our case, in order to provide a set of chromosomes which each one presents a solution to the problem, we let each chromosome to be a set of 16 centroids (the centers of clusters). In this way, each chromosome consists of an array of 16 pair values which are I and Q components of the centroids. In order to improve the convergence, the scattering of the initial centroids, as initial chromosomes, was chosen similar to ideal constellation, although relatively, with the PSK and QAM constellations.

In the next step, the fitness of each chromosome is evaluated by comparing with signal symbols and then computing the objective function through HCM (hard C-means) and then ascending sort these values. Fitness function is defined as inverse of the objective function. Therefore, a chromosome with the highest fitness value will have the rank one in the population. In this manner, the rank of each chromosome is defined proportionally. Chromosomes with the high fitness will have a higher chance to be present at the next population. In this step, also the cumulative probabilities are computed to be utilized at the selection step. The pseudo code of this step is shown in figure (2).

```

For  $j = 1$  to Pop-Size do
    Compute the objectives  $Z_j$  for  $U_j$ ;
End for
For  $j = 1$  to Pop-Size do
    Compute the fitness evaluation ( $U_j$ );
End for
For  $j = 1$  to Pop-Size do
    Compute the cumulative probabilities ( $q_j$ );
End for

```

FIGURE 2: The pseudo code of fitness evaluation.

Selection operator uses the fitness of chromosomes in the preceding step and selects some chromosomes within the population, to be employed at the next step. Chromosomes with high fitness values will have a higher chance to be selected. Hence, increasing the probability of selection and propagation leads to that a higher number of duplicates and children, which resulted from chromosomes with high fitness, may be existed at the next generation. Selection of chromosomes to product the next generation is performed by rotating the Roulette Wheel for times equal to the number of members of population, so that in each rotation exactly one chromosome is selected to contribute in the production of the next generation. The pseudo code of selection step is shown in figure (3).

```

For  $j = 1$  to Pop-Size do

    Generate a random real number  $r_s$  in the interval of  $[0, 1]$ 

    If  $(q_j - 1 < r_s < q_j)$  then

        Select  $U_j$  ;

    end if

End for

```

FIGURE 3: The pseudo code of selection step.

Then, the crossover operator operates on the chromosomes of the intermediate population and combines those together. This modifier combines the chromosomes with the hope of reproducing a better chromosome of the offspring than the parent. After the two chromosomes were selected to cross over, they can be combined with different method of crossing over. In the present plan, we use the single point crossing over. For this, we selected two chromosomes among the middle population of chromosomes. Then we select a point of the chromosome randomly and exchange all the gens of these two chromosomes after this point. To improve the algorithm efficiency, we select the random number among 3 and 14 and then with a crisscross way we exchange all the gens after that point and reproduce two offspring chromosome from the two parent chromosomes. After that, offspring chromosomes replace for its parents chromosome. The pseudo code of crossover step is shown in figure (4).

```

For  $j = 1$  to Pop-Size / 2 do

    Generate a random real number  $r_c$  in the interval of  $[0, 1]$  ;

    If  $r_c \leq P_c$ 

        Perform the Crossover on randomly selected the  $l^{th}$  and  $m^{th}$  chromosomes;

    end if

End for

```

FIGURE 4: The pseudo code of crossover step.

Then, it is time to apply the mutation operator that randomly varies one or more elements of a chromosome. When a mutation modifier is implemented on a chromosome, mutation will be created in that chromosome. A current method to implement mutation is to change a one or more gens from a chromosome randomly. In other words, this modifier is defined as a change in one or more bits of a vector of answers. Therefore, in the present problem, mutation modifier is done with changes on the elements of one or more rows of matrixes of answers. Finally, after implementation of mutation modifier, the middle population replace for the initial population so that during the next repeat all the processes of evaluation, crossing over, mutation and

replacement are done on the new generation and lead to the next generation. The pseudo code of mutation step is shown in figure (5). Termination condition of the algorithm is satisfied when the objective function value hasn't considerable variation through the recent iterations. After termination, the chromosome (i.e. set of centroids) with the highest fitness within the final population is presented in the output [19].

```

For  $j = 1$  to Pop-Size do
    For  $k = 1$  to  $n$  do
        Generate a random real number  $r_m$  in the interval of  $[0, 1]$  ;
        If ( $r_m \leq P_m$ ) then
            Generate new elements in the  $k^{th}$  row of the  $j^{th}$  Chromosome ;
        End if
    End for
End for

```

FIGURE 5: The pseudo code of mutation step.

5. MODULATION CLASSIFICATION USING TEMPLATE MATCHING

In template matching method by eliminating of the used post processing step in previous methods, the possibility of error marking was decreased and also this method can recognize 256QAM modulation with high accuracy.

The main idea in this method is assessment and investigating of input signals based on relative similarity that exist between different kind of standard QAM modulations with predefined levels. In fact we suppose that the signal modulation kind is according to 4QAM, 16QAM and 256QAM, so at last our purpose is the recognition of modulation between all different kinds.

In this method the standard models of constellation diagram of QAM modulation signals when they are not on the influence of noises and deviation and etc, are the parameters in decision making and choosing the modulation kind. It is done by analyzing the resulted similarity of constellation of input signals with standards constellations and then the modulation that has the greatest similarity would be chosen. For this purpose, the genetic algorithm had been used that the details of used algorithms are given below.

5.1 Definition of Ideal Cluster Centriods

In the beginning for, every kind of QAM modulation family, the ideal centriods would be defined, so after we can compare them with resulted centriods from signals. We define the template forms of 256QAM, 64QAM, 16QAM and 4QAM modulation constellations in $I-Q$ quarter page. For 4QAM modulation, an ideal cluster centriod in quarter page and for 16QAM modulation, 4 ideal cluster centriods and for 64QAM modulation, 16 cluster centriods and for 256QAM, 64 cluster centriods can be defined. All centriods are defined in $I-Q$ quarter page and in $[0, 1]$ interval. So here because we would have the possibility of comparing of centriods of input signals with ideal centriods, the absolute value of receiving samples from input are calculated and then normalized.

5.2 Recognition of QAM Family

The first kind of modulation that is assessed is 256QAM. First, preliminary centriod for using genetic algorithm is defined and then these preliminary centriods are applies in genetic

program. The centroids that are given by genetic program are compared by 256QAM ideal centroids. This comparison is done by calculating of Euclidean distance of centroids.

The given value is compared with defined corresponding threshold for 256QAM and in a case that the given value would be less than threshold, this kind of modulation is introduced as the detection modulation and the program would be end. But if this case is not full filled, the next kind of modulation that is 64QAM would be assessed in a similar way and in a case the requirement for similarity of modulation is full filled, it will be introduced as the input detection modulation, otherwise the assessment is done for the next modulation, 16QAM. If none of the modulation kinds did not full filled, the requirement for similarity at last the 4QAM is supposed as the input signal and would be introduced in the outputs.

So the assessment of different kinds of modulations is started from 256QAM and would be end with 4QAM. The reason is when the correctness of similarity is full filled for 256QAM, this be full filled for lower levels too. Also if the similarity requirement is full filled for 64QAM, it would be right for 16QAM and 4QAM, but the reverse would not be right. It means that if the requirement for similarity for 16QAM is full filled, it would not be full filled for modulations with higher levels (64QAM and 256QAM), so the assessment is standard from high levels of 256QAM and will be end in 4QAM.

From below threshold value for comparing the similarities are used for recognition of different modulations of QAM families in this program. The 290 threshold was used for 256QAM modulation recognition. Threshold is used for differentiate of 64QAM modulation is set to 320 and at last threshold for recognition of 16QAM modulation is set to 350. Figure (6) shows the flowchart of the proposed method for recognition of modulation.

6. PERFORMANCE EVALUATION AND SIMULATION RESULTS

In order to evaluate the performance of the proposed method, simulation has been performed for various SNR values and various types of QAM modulations. Channel model, applied in this work, has been assumed to be an AWGN channel, and it is also assumed that there is no time and/or frequency synchronization error. Simulation results show that this method has an efficient performance and high accuracy for the recognition of modulation. Figures (7), (8) and (9) show the centroids of the clusters which are obtained from Genetic algorithm and ideal constellation points of related modulation for 16-QAM, 64-QAM and 256-QAM, respectively.

The main program finds a template (ideal constellation points) with matches more to the detected centroids. This judgment is done based on the threshold related to each type of modulation. Finally, the detected centroids in $I-Q$ plane and the recognized modulation type and the fitness value are considered as output. Figure (10), (11), (12) and (13) depict the detected centroids and data symbols in the whole constellation diagram, for types of 4-QAM, 16-QAM, 64-QAM and 256-QAM, respectively.

For recognition of these types of modulation with various SNR, different numbers of samples are used which are presented in table (1). Figure (14) presents the accuracy percentage of the modulation recognition versus SNR, for various types. The accuracy percentages have been obtained by executing algorithm enough times and calculating the ratio between correct recognition and total number of execution.

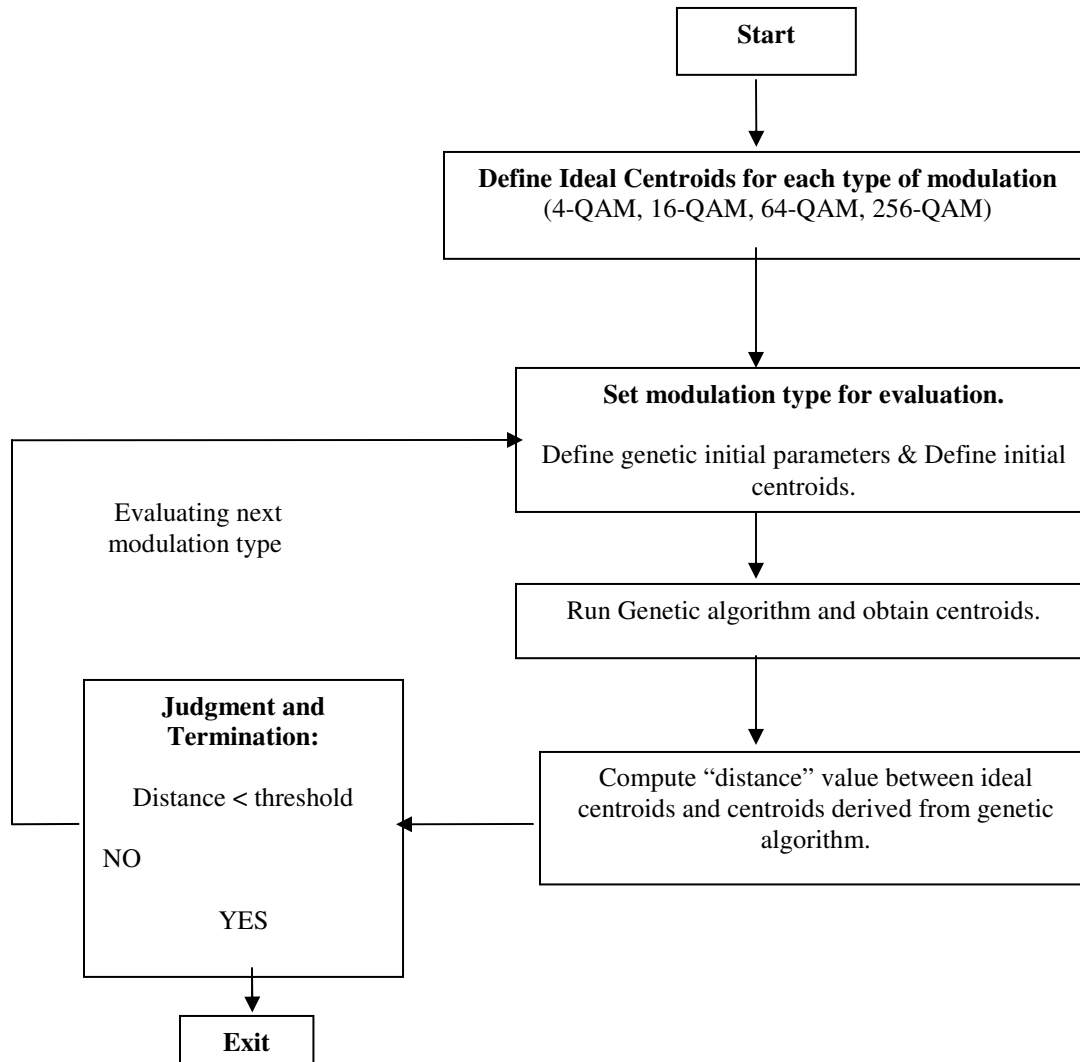


FIGURE 6: The flowchart of propose method.

7. CONCLUSION

In this paper Genetic algorithm and template matching was used to classify different modulation types of QAM, using the constellation diagram of the received signal. The proposed method shows a good performance for recognition even in extremely low SNR condition. Of course, it must be mentioned that the performance could be increased with higher number of data symbols. Another advantage of this method is calculating final centroids of clusters and determining the location of these centroids in constellation diagram.

Using Template matching technique would increase the accuracy of recognition of modulation in low SNR and because of that; we succeed in recognition of 256QAM modulation with SNR equal to 17 with 100% accuracy and acceptable accuracy for lower values. As a result the

capability of proposed method in recognition of 64QAM modulation for SNR equal to 10 and more than that was 100% and for lower values was in acceptable levels. This method is also capable for recognition of 16QAM modulation with SNR equal to 2.5 and values more than that and for lower SNR values was in acceptable accuracy levels. This method can recognize all 4QAM modulations with any signal to noise values with 100% accuracy. At last simultaneously with decreasing of the value of SNR, with increment of the number of input samples, the accuracy of modulation recognition can be increased.

The method that have been used can be expanded and use them for modulation recognition of any PAM signals. These signals have one dimensional constellation while in this research we study the signals with two dimensional constellations which are more complicated. Thus with a little change we can use them for recognition of PAM signals. From these signals the MFSK and MASK modulations can be referred.

With little changes in proposed method it can be used in recognition of modulations which have non standard one dimensional or two dimensional constellations. By rotating the constellation diagram of PAM signals for 45° in $I-Q$ plane, without any change in proposed method, it can recognize modulations.

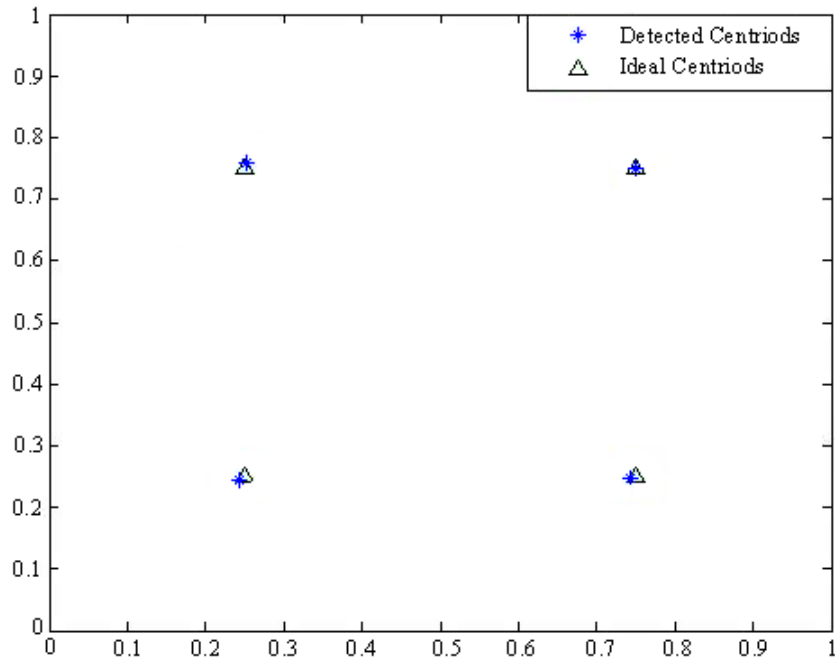


FIGURE 7: Centroids resulted from FCM algorithm in first quadrant for 16-QAM with SNR=5dB, and comparison with the 16-QAM template.

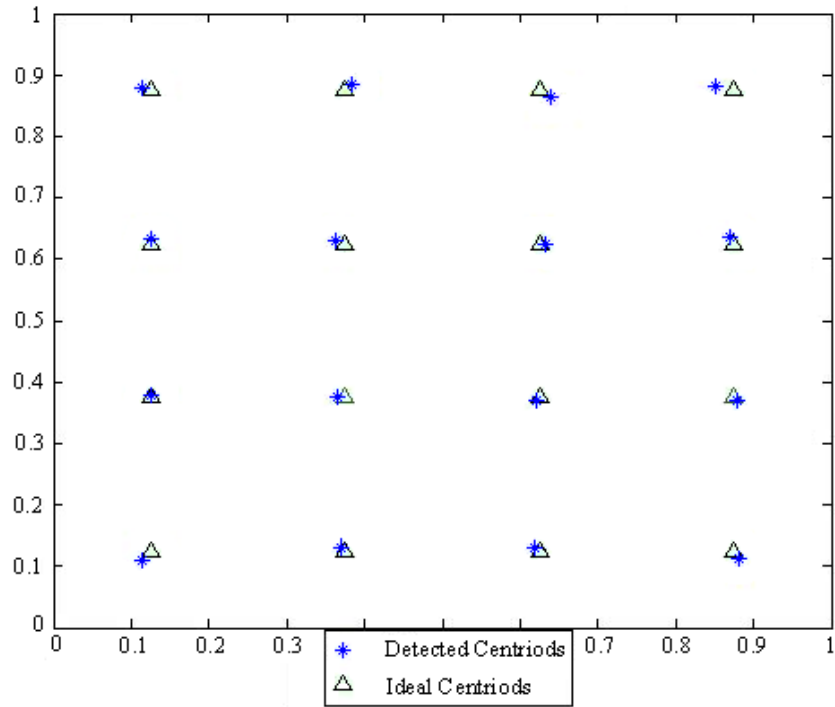


FIGURE 8: Centroids resulted from FCM algorithm in first quadrant for 64-QAM with SNR=12dB, and comparison with 64-QAM template.

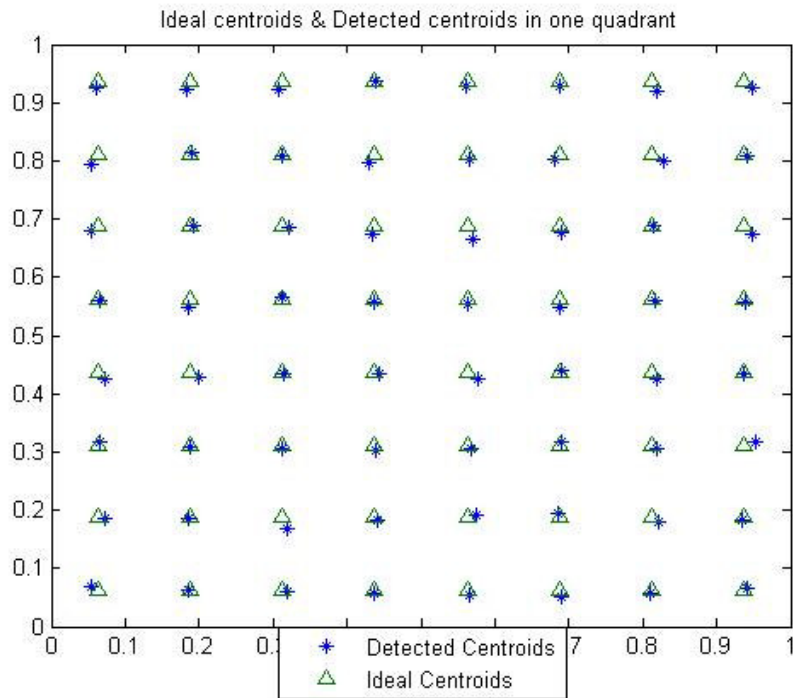


FIGURE 9: Centroids resulted from FCM algorithm in first quadrant for 256-QAM with SNR=23dB, and comparison with 256-QAM template.

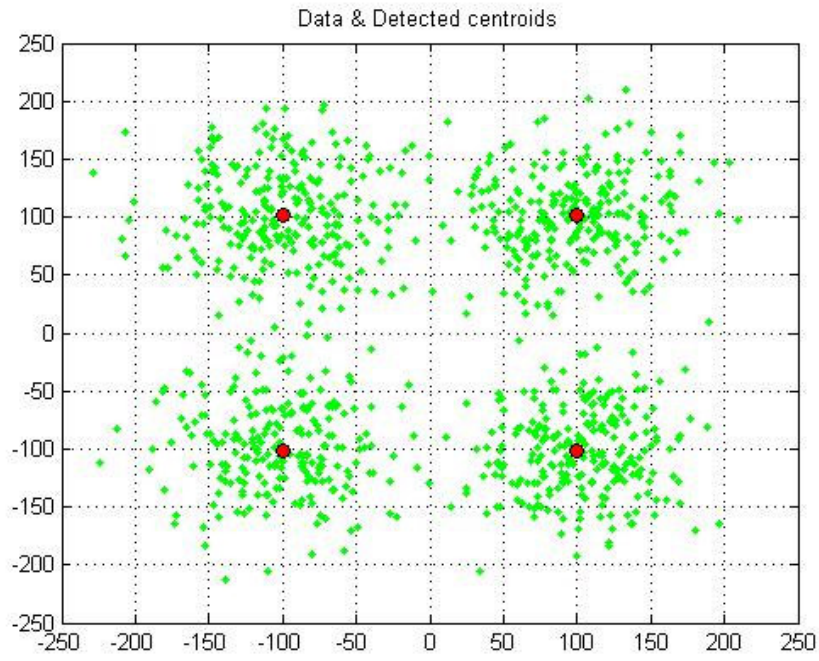


FIGURE 10: Data symbols and resulted centroids after recognition of 4-QAM.

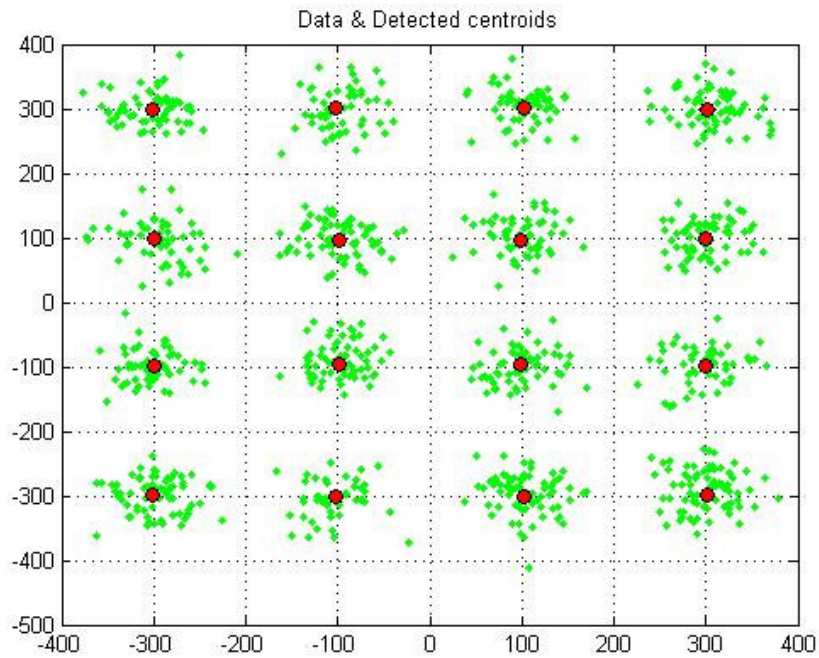


FIGURE 11: Data symbols and resulted centroids after recognition of 16-QAM.

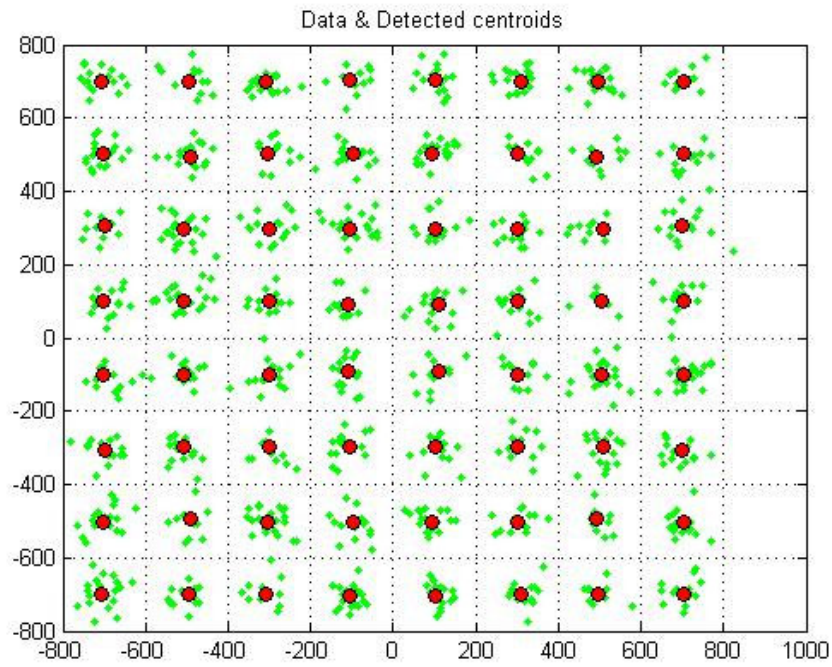


FIGURE 12: Data symbols and resulted centroids after recognition of 64-QAM.

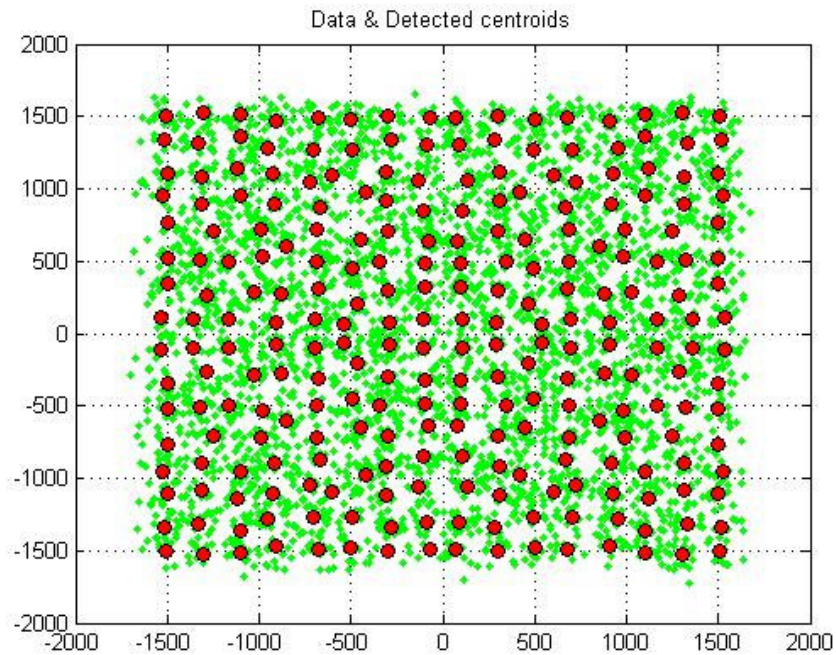


FIGURE 13: Data symbols and resulted centroids after recognition of 256-QAM.

256-QAM	SNR	30dB	25dB	20dB	19dB	18dB	17dB
	Number of Samples	1000	1500	3000	4000	9000	22000
64-QAM	SNR	25dB	20dB	17dB	15dB	12dB	10dB
	Number of Samples	1000	1000	1000	1000	3000	10000
16-QAM	SNR	15dB	10dB	8dB	5dB	3dB	2.5dB
	Number of Samples	1000	1000	1000	1000	3500	10000
4-QAM	SNR	10dB	5dB	3dB	1dB	0dB	-2dB
	Number of Samples	1000	1000	1000	1000	1000	1000

TABLE 1: Number of samples for modulation recognition with various SNR.

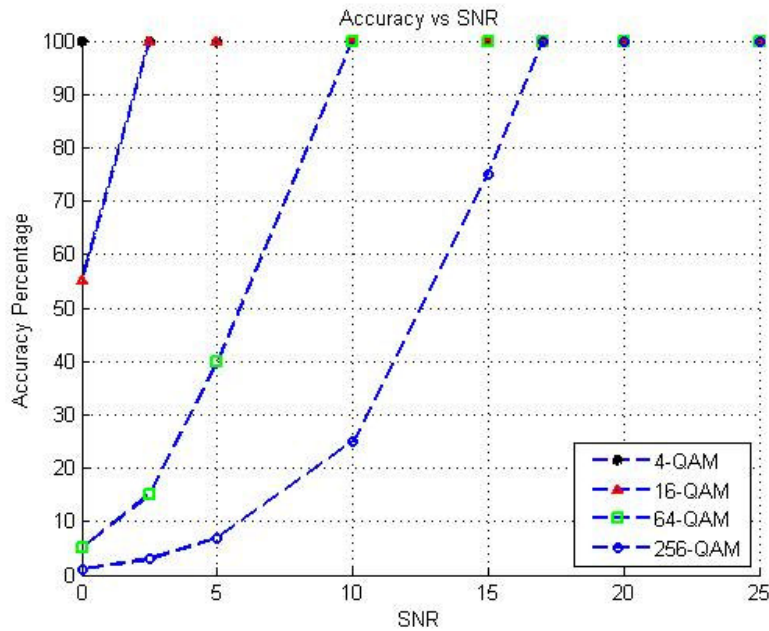


FIGURE 14: Accuracy percentage of recognition versus SNR.

8. REFERENCES

1. J. Lopatka, M. Pedzisz. "Automatic Modulation Classification using Statistical Moments and a Fuzzy Classifier", Signal Processing Proceedings, WCCC- ICSP 2000, 5th international conf. on, 3:1500-1506, 21-25 Aug. 2000
2. Y. O. Al-Jalili. "Identification Algorithm of Upper Sideband and Lower Sideband SSB Signals", Signal Processing, 42:207-213, 1995

3. L. Narduzzi, M. Bertocco. *"Conformance and Performance"*, Department of Electronic and Informatics, Pavova University, 2003
4. J. Reichert. *"Automatic Classification of Communication Signals using Higher Order Statistics"*, ICASSP 92, 221-224,1992
5. R. Schalkoff, *"Pattern Recognition: Statistical, Structural and Neural Approach"*, John Wiley, (1992)
6. Bijan G. Mobaseri. *"Constellation shape as a robust signature for digital modulation recognition"*, Military Communications Conference Proceedings, MILCOM IEEE, 1:442-446, 1999
7. Bijan G. Mobasser. *"Digital Modulation Classification using Constellation Shape"*, Signal Processing, 80(2):251-277,2000
8. F. Jondral. *"Automatic Classification of High Frequency Signals"*, Signal Processing, 9(3):177-190,1985
9. L. Dominguez, J. Borrallo, J. Garcia. *"A General Approach to the Automatic Classification of Radiocommunication Signals"*, Signal Processing, 22(3):239-250,1991
10. F.F. Liedtke. *"Computer Simulation of an Automatic Classification Procedure for Digitally Modulated Communication Signals with Unknown Parameters"*, Signal Processing, 6:311-323,1984
11. J. Aisbett. *"Automatic Modulation Recognition using Time-Domain Parameters"*, Signal Processing, 13(3):323-329,1987
12. A. Polydoros, K. Kim. *"On the Detection and Classification of Quadrature Digital Modulation in Broad-Band Noise"*, IEEE Transactions on Communications, 38(8):1199-1211,1990
13. C. Huang, A. Polydoros. *"Likelihood Method for MPSK Modulation Classification"*, IEEE Transaction on Communications, 43(2/3/4):1493-1503,1995
14. S. Soliman, S. Hsue. *"Signal classification using statistical moments"*, IEEE Transactions on Communications, 40(5):908-915,1992
15. W. Wei, J. Mendel. *"A New Maximum Likelihood for Modulation Classification"* Asilomar-29, 1132-1138, 1999
16. K. Chugg, et al. *"Combined Likelihood Power Estimation and Multiple Hypothesis Modulation Classification"*, Asilomar-29, 1137-1141, 1996
17. Y.Lin, C.C. Kuo. *"Classification of Quadrature Amplitude Modulated (QAM) Signals via Sequential Probability Ratio Test (SPRT)"*, Report of CRASP, University of Southern California, July 15, 1996.
18. Nhi P. Ta. *"A Wavelet Packet Approach to Radio Signal Classification"*, symposium on Time-Frequency and Time Scale Analysis, 508-511, 1994
19. Linhu Zhao, Yasuhiro Tsujihiura, Mitsuo Gien. *"Genetic algorithm for fuzzy clustering"*, Proceedings of IEEE International Conference on, 716-719, 1996

Moving One Dimensional Cursor Using Extracted Parameter from Brain Signals

Siti Zuraimi Salleh

*Department of Electronic Eng.
Faculty of Electrical Engineering
University Technology of Malaysia (UTM)
81310 Skudai, Johor, Malaysia*

missxeetea_z@yahoo.com

Norlaili Mat Safri

*Department of Electronic Eng.
Faculty of Electrical Engineering
University Technology of Malaysia (UTM)
81310 Skudai, Johor, Malaysia*

norlaili@fke.utm.my

Siti Hajar Aminah Ali

*Department of Telecommunication
Faculty of Electrical Engineering
University Technology of Tun Hussein Onn Malaysia (UTHM)
86400 Batu Pahat, Johor, Malaysia*

aminahh@uthm.edu.my

Abstract

This study focuses on developing a method to determine parameters to control cursor movement using noninvasive brain signals, or electroencephalogram (EEG) for brain-computer interface (BCI). Two conditions were applied i.e. Control condition where subjects relax (resting state); and Task condition where subjects imagine a movement. In both conditions, EEG signals were recorded from 19 scalp locations. In Task condition, subjects were asked to imagine a movement to move the cursor on the screen towards target position. Fast Fourier Transform (FFT) was used to analyze the recorded EEG signals. To obtain maximum speed and accuracy, EEG data were divided into various interval and difference in power values between Task and Control conditions were calculated. As conclusion, the present study suggests that difference in delta frequency band between resting and active imagination may be use to control one dimensional cursor movement with parietal region produces the optimum output.

Keywords: Brain-computer interface (BCI), electroencephalogram (EEG), extracted parameter, Fast Fourier Transform (FFT)

1. INTRODUCTION

For normal people, communication is a need to undergo their daily activities. Communication is a process to transmit or transfer information, thought or feeling by or to or between people or groups. It is a connection allowing access between persons by either verbal contact or action.

But some people suffer from “locked-in syndrome”, meaning they are completely unable to control any muscle, preventing them from communicating with their caregivers or environment [1]. For such users, a brain-computer interface or BCI is the only hope for even communicating with loved ones, controlling even simple devices like televisions or lamps or otherwise expressing oneself. BCI is a novel augmentative communication system that translates human intentions into a control signal for an output device such as a computer application [2] or a mobile robot [3], in which users send information using brain activity alone without conventional peripheral nerves and muscles [3].

BCI can be divided into two general categories i.e. invasive and noninvasive [3]. Most noninvasive BCI systems use electroencephalogram (EEG) signals; i.e., the electrical brain activity recorded from electrodes placed on the scalp. The main source of the EEG is the synchronous activity of thousands of cortical neurons. Measuring the EEG is a simple noninvasive way to monitor electrical brain activity, but it does not provide detailed information on the activity of single neurons (a few μ Volts) and noisy environment (especially if recording outside shield rooms) [3]. In invasive BCI systems, the activity of single neurons (their spiking rate) is recorded from microelectrodes implanted in the brain. Such systems are being studied mainly in nonhuman primates [4]. These invasive BCIs face substantial technical difficulties and entail significant clinical risks as they require that recording electrodes be implanted in the cortex and function well for long periods, and they risk infection and other damage to the brain [5]. For human, therefore, noninvasive BCI systems are applied due to the clinical risks and ethics [3].

2. DATA COLLECTION AND ANALYSIS

In BCI studies, the foremost important element is data collection and analysis whereby the recorded and collected data will be used as an input to the system. The input referred here is EEG signal; it was digitized, analyzed and processed for extracting important and useful information [6].

2.1 Participants

Six normal healthy subjects aged 20-26 years old gave informed consent to participate in these experiments.

2.2 Condition

Subjects sat on a chair facing a monitor screen that was placed one meter in front of the subjects. Two conditions were applied, i.e. Control condition and Task condition. In Control condition, subjects were asked to relax (resting) while in Task condition, subjects were asked to imagine voluntary movement, e.g. imagine moving a cursor to a target location on a computer screen. In Control condition, subjects were instructed to fix their eyes on the centre of the screen. No image was displayed on the screen during the entire experiment. In Task condition, subjects were asked to imagine a movement to move the cursor on the screen towards the target. Cursor moving towards the target was displayed on the screen during the entire experiment to imitate real application.

2.3 Data Acquisition

EEG signals were obtained from 19 scalp electrodes, placed on the scalp based on 10-20 electrode placement system (Figure 1). Signals were recorded with passbands of 0.5 -120 Hz and stored in a personal computer with a sampling frequency of 1 kHz. A single trial lasted for 10 seconds and four trials, as illustrated in Figure 2, were conducted for each condition with intervening rest period to avoid fatigue.

The 10-20 electrode placement system is a method used to describe the location of scalp electrodes. These scalp electrodes are used to record the EEG using a machine called an electroencephalograph. The system is based on the relationship between the location of an electrode and the underlying area of cerebral cortex. Each electrode site has a combination of a letter and a number (or another letter) to identify the lobe and the hemisphere location, respectively. The letters F, T, C, P and O stand for Frontal, Temporal, Central, Parietal and Occipital. Even numbers 2, 4, 6 and 8 indicate the electrodes at right hemisphere while electrodes at the left hemisphere are indicated by odd numbers 1, 3, 5 and 7. Other than that, small letter 'z' refers to an electrode placed in the midline. The '10' and '20' are referred to the 10% or 20% inter-electrode distance [7].

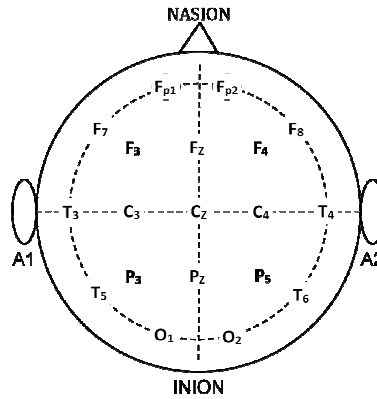


FIGURE 1: Scalp Locations Based On 10-20 Electrode Placement System.

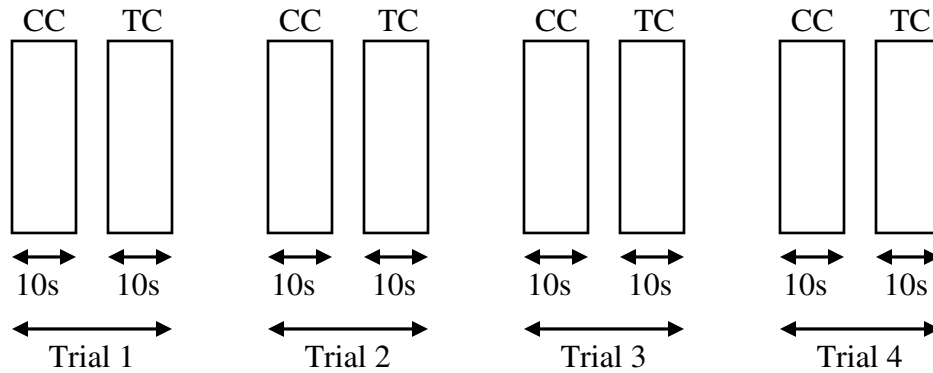


FIGURE 2: The Sequence of Condition. CC and TC represent Control condition and Task condition, respectively.

2.4 Data Analysis

10 seconds of EEG data of Control and Task conditions were divided into ten frames so that each frame consists of one second data. The EEG data were divided into various time intervals, i.e. 1024 ms and 512 ms, as depicted in Figure 3, to investigate time interval that provide optimum speed and accuracy. Data in frequency domain were obtained by applying Fast Fourier Transform (FFT) with butterfly operation for each time interval in a frame [8]. Frequency was divided into six groups, i.e. delta band (0 to 4 Hz), theta band (4 to 7 Hz), alpha band (8 to 12 Hz), beta band (13 to 30 Hz), gamma band (31 to 50 Hz) and high gamma band (>51 Hz). Each frequency band in Task condition was compared to the Control Condition. Differences between

these two condition were obtained and observed, i.e. difference in power (f) (DP) = power in Task (f) – power in Control (f).

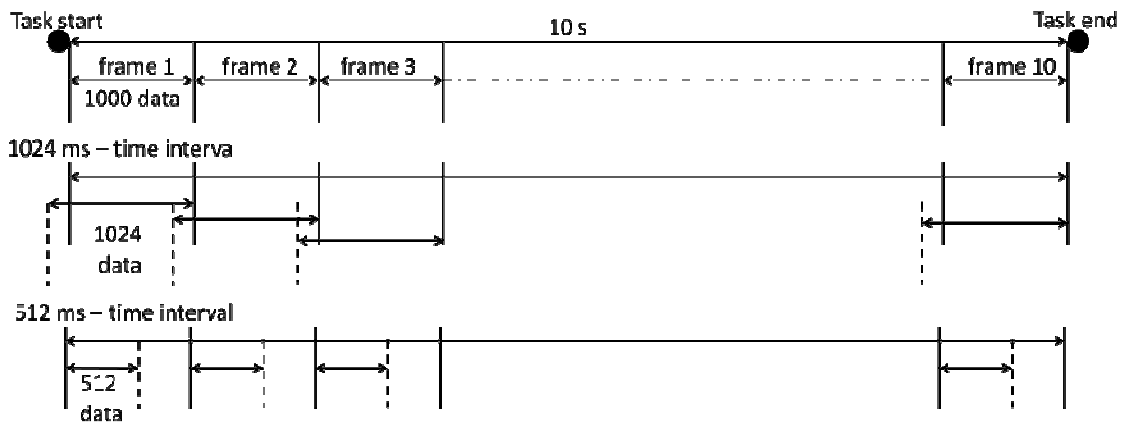
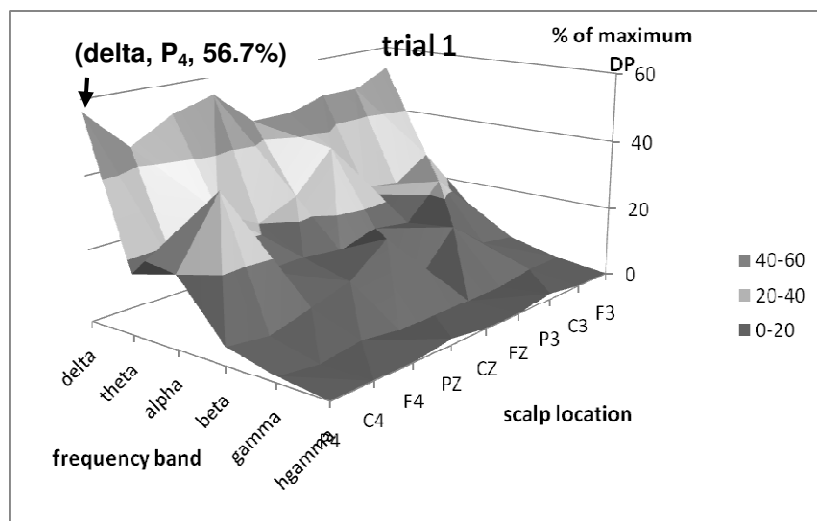


FIGURE 3: Division of Data at 1024 ms and 512 ms Time Interval.

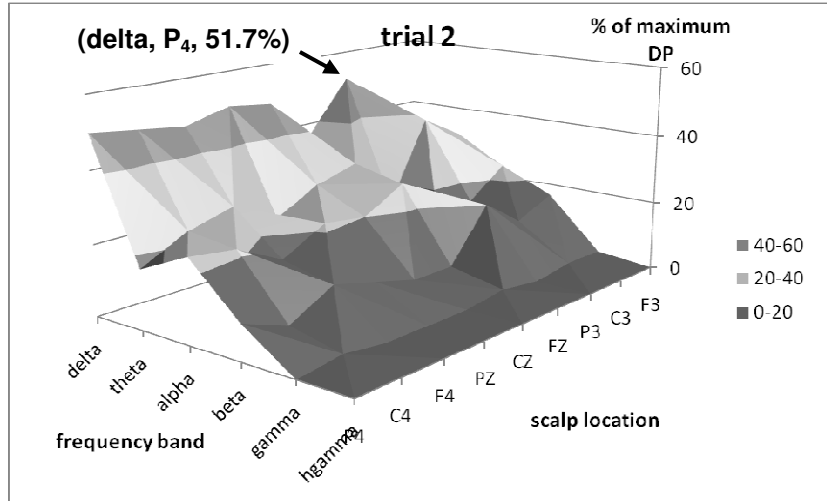
3. PARAMETER EXTRACTION RESULT

3.1 Time Interval: 1024 ms

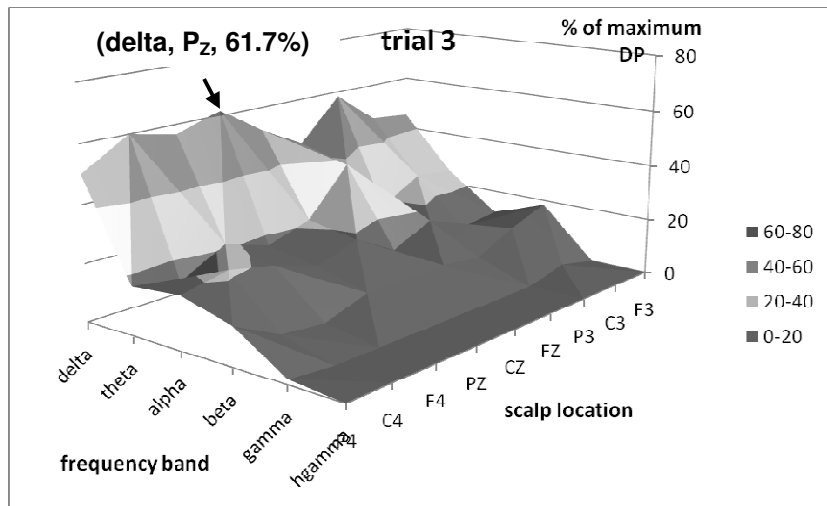
Figure 4 shows the result of percentage of maximum DP using 1024 ms time interval for all trials (6 subjects x 10 frames). For trial 1 (Figure 4(a)), maximum DP was found at P_4 site with 56.7% occurrence. For trial 2 (Figure 4(b)), maximum DP was observed at another site i.e. P_3 with 51.7% occurrence. For trials 3 and 4 (Figure 4 (c) and (d)), both maximum DP was found at P_Z site with 61.7% and 60.0% occurrences, respectively. Generally, the maximum DP occurred at delta band, which ranged 0-4 Hz, for all trials.



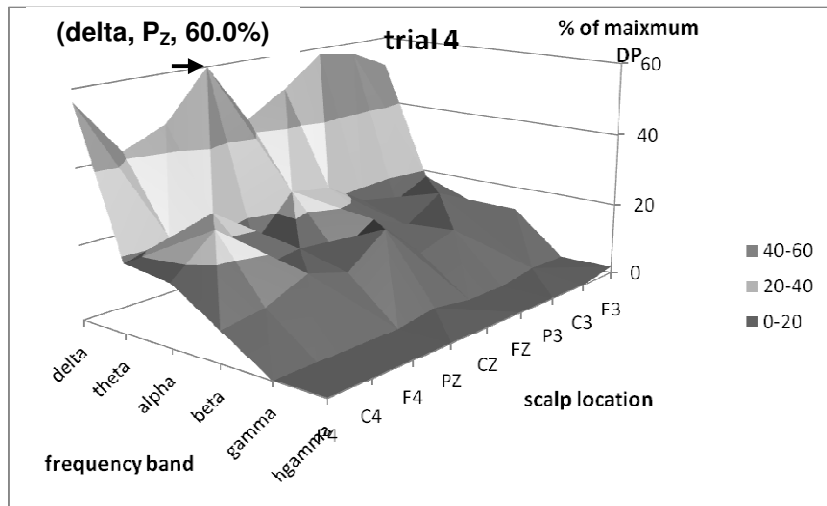
(a)



(b)



(c)

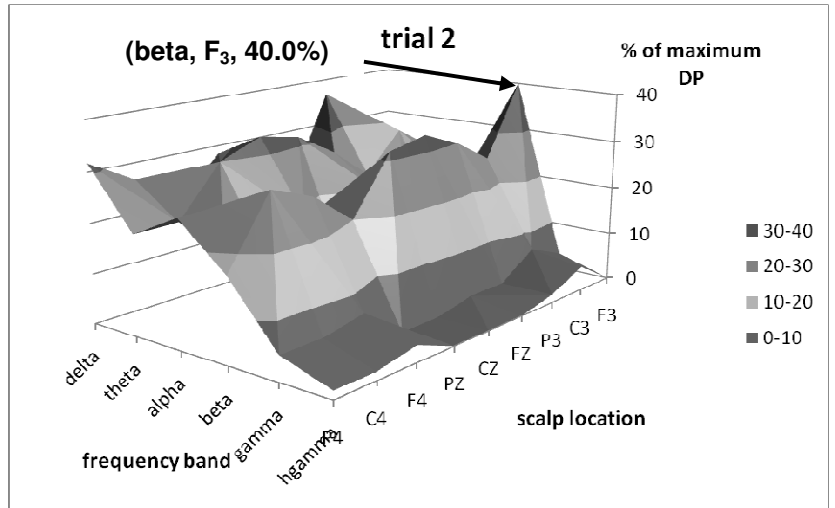
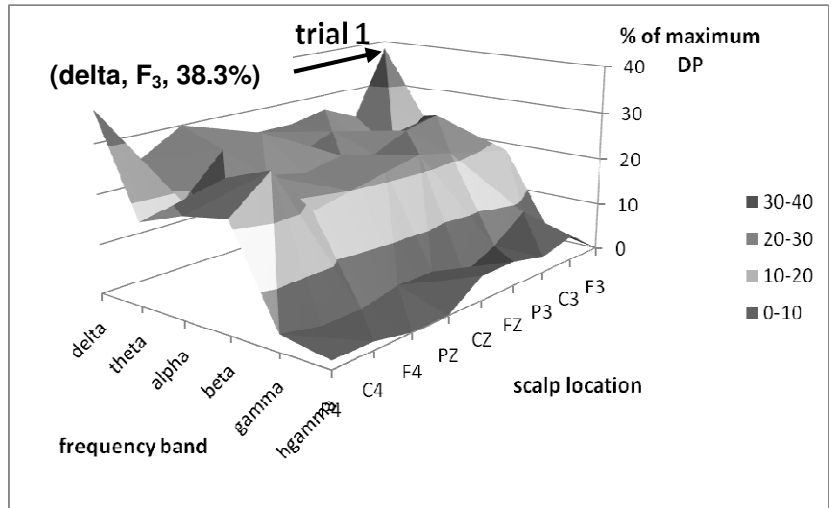


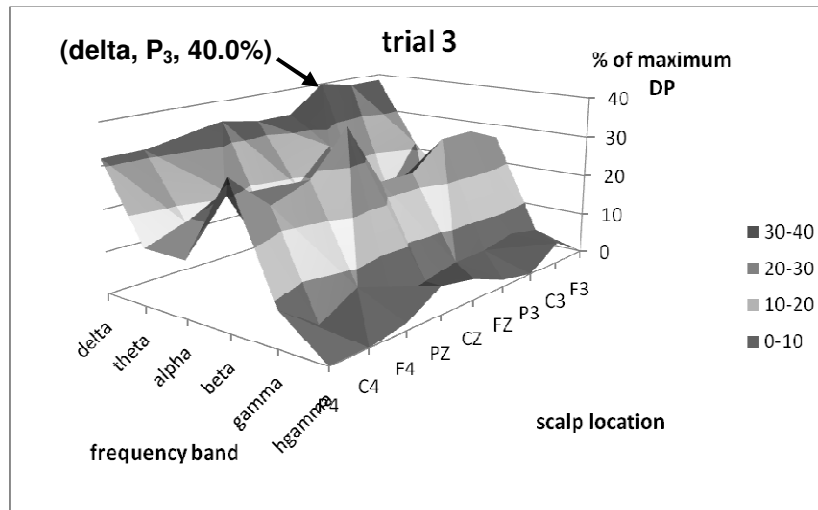
(d)

FIGURE 4: Result for Trial 1, 2, 3 and 4 Using 1024 ms Sampling Interval.

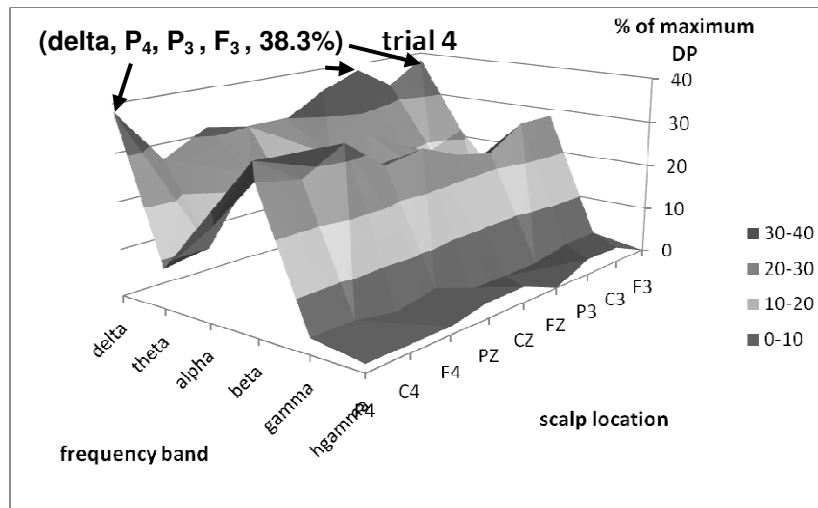
3.2 Time Interval: 512 ms

Maximum DP for all trials (6 subjects x 10 frames) using 512 ms time interval are shown in Figure 5. In section 3.1, the frequency range for maximum DP was constant for all trials i.e. in delta band. However, there were two different frequency band observed in time interval 512 ms. For trials 1 and 2 (Figures 5(a) and (b)), both maximum DP occurred at F₃ site with 38.3% and 40% occurrences, respectively. However, the frequency band in which the maximum DP occurrence appeared was different. For trial 1, maximum DP occurrence was obtained in delta band whereas for trial 2, it was obtained in beta band. For trials 3 and 4 (Figures 5(c) and (d)), the frequency range in which maximum DP was found was seen constant but not the scalp location. The maximum DP occurred in delta band for both trials 3 and 4. In trial 3, the maximum DP was found at P₃ site with 40% occurrence. Meanwhile, in trial 4, there were three different sites assembled the maximum DP awith similar occurrence percentage, 38.3%, i.e. F₃, P₃ and P₄.





(c)



(d)

FIGURE 5: Result for Trial 1, 2, 3 and 4 Using 512 ms Time Interval.

3.3 Averaging for All Trials

Initially, it is expected that every trial will produce similar result to determine the parameters. However, it was not the case here. The maximum DP was found at various scalp locations for all the trials. To overcome these distinctions, averaging process based on location was done for all the trials. Results are shown in Figures 6 and 7 for time interval 1024 ms and 512 ms, respectively.

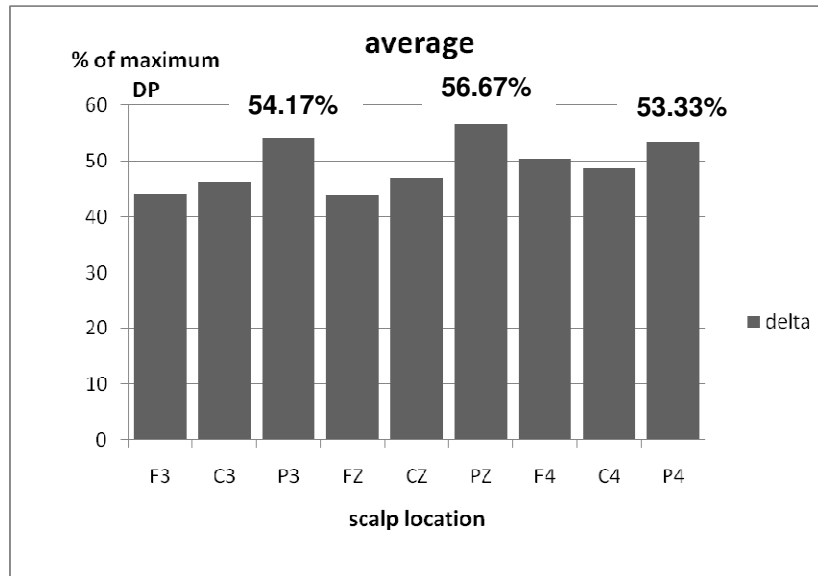


FIGURE 6: Percentage of Averaged Maximum DP in Delta Band for Time Interval 1024 ms.

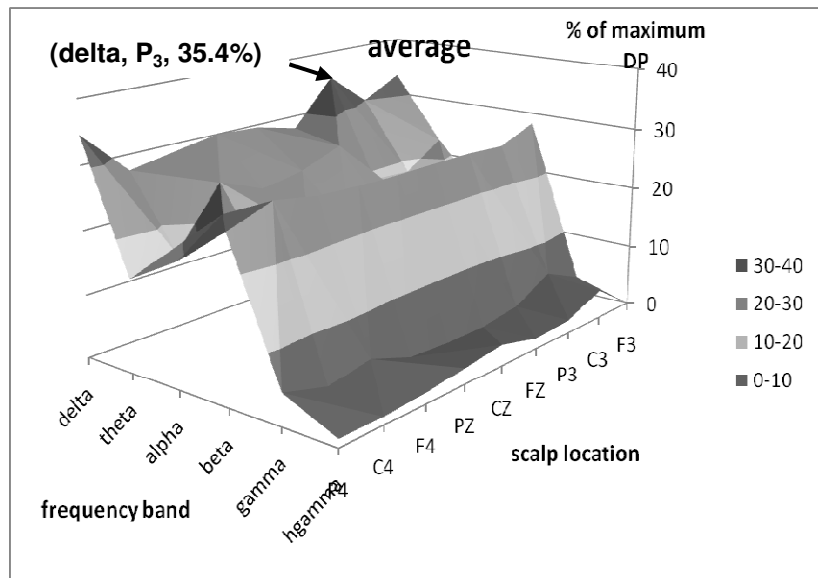


FIGURE 7: Percentage of Averaged Maximum DP for Time Interval 512 ms.

Figure 6 shows that the scalp location with highest percentage of averaged maximum DP using time interval 1024 ms was P_z (56.67%), followed by P_3 (54.17%) and P_4 (53.33%). For time interval 512 ms, the highest percentage of averaged maximum DP occurred at P_3 site within delta frequency range with 35.4% occurrence. The other two locations were P_4 and F_3 , with 34.6% and 33.3% occurrences, respectively, also in the delta frequency band (Figure 7).

4. DISCUSSION AND CONCLUSION

In this study, we were interested to observe the maximum difference in power between resting and active imagination, i.e. at which scalp location and frequency band it occurred. The two features (scalp location and frequency band) are similarly identified by Leuthardt et al. (2004) in their BCI study using electrocorticographic (ECoG) [9]. However, they focus on r^2 value instead of maximum difference in power.

From our findings, the maximum different in power (DP) between the two conditions occurs in delta band at posterior area, i.e. P_z for 1024 ms time interval and P_3 for 512 ms time interval. Comparing the two time intervals, delta band at central posterior area (P_z) provided higher percentage of difference in power, hence, can be use to control one dimensional cursor movement [10] for future study of online BCI. By selecting the power difference of delta frequency band (< 4 Hz) in central posterior area, it is expected that the cursor can move further and faster in one dimensional direction towards targeted location. It is also expected that no training is required to obtain optimum results. Many researches have shown that BCI training must be conducted many times to achieve best performance for each subject. For example, Wolpaw et al. (2004) had conducted over twenty sessions per subject, at a rate of two to four per week [5]. In this study, the result was based on a single session of an experiment; hence, it is believed that the extracted feature can be used to control a one dimensional cursor movement without prior training. However, further study is needed to delineate this speculation.

Although it is always been reported that BCI researchers use mu and beta rhythm which is associated with actual movement or imagination of movement, in this research, the values in the delta band are used instead to convert EEG signals to cursor movement. Wolpaw et al. (2002) reported that movement or preparation for movement is typically accompanied by a decrease in mu (8-12 Hz) and beta (13-30 Hz) rhythm amplitudes [1], therefore, minimizing their different amplitude values in power. In this study, we found maximum difference in power occurs in delta frequency band even though the slow rhythm is always being associated with sleep wave in adult.

In the study, the maximum power difference in the delta frequency band was found at posterior area that contains primary and association cortices for somatosensation [11]. The region can be divided into two functional regions, one that involves sensation and perception and the other that concern with integrating sensory input, primarily with the visual system [12]. Since the subject was provided with the vision of targeted location and the cursor location, the posterior area probably constructs a spatial coordinate system to represent the two locations. Further study is needed in this aspect.

In conclusion, the present study suggests that difference in delta power at posterior area between resting and active imagination may be use to control a one dimensional cursor movement.

5. REFERENCES

1. J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan. "Brain-computer interfaces for communication and control". *Clinical Neurophysiology*, 113: 767-791, 2002
2. D.J. McFarland and J.R. Wolpaw. "Sensorimotor rhythm-based brain-computer interface (BCI): feature selection by regression improves performance". *IEEE Transaction on neural Systems and Rehab.*, 13(3): 372-379, 2005
3. J.R. Millan, F. Renkens, J. Mourino and W. Gerstner. "Brain-actuated interaction". *Artificial Intelligence*, 159: 241-259, 2004

4. J.M. Carmena, M.A. Lebedev, R.E. Crist, J.E.O'Doherty, D.M. Santucci, D.F. Dimitrov, P.G. Patil, C. S. Henriquez and M.A.L. Nicolelis. "*Learning to control a brain-machine interface for reaching and grasping by primates*". PLOS Biology, 1(2): 193-208, 2003
5. J.R. Wolpaw, D.J. McFarland, T.M. Vaughan and G. Schalk. "*Control of a two dimensional movement signal by a noninvasive brain-computer interface in humans*". PNAS, 101(51):17849-17854, 2004
6. M. M. Ahmed and D. Mohammad. "*Segmentation of brain MR images for tumor extraction by combining kmeans clustering and Perona-Malik anistropic diffusion model*". International Journal of Image Processing, 2(1), 27-34, 2008
7. "*Biomedical Signals Amplifier*", ElettronicaVeneta, pp. 27 (2006)
8. R. S. Manzoor, R. Gani, V. Jeoti, N. Kamel and M. Asif. "*Dwpt based FFT and its application to SNR estimation in OFDM Systems*". Signal Processing: An International Journal, 3(2), 22-33, 2009
9. E.C. Leuthardt, G. Schalk and J.R. Wolpaw. "*A brain-computer interface using electrocorticographic signals in human*". J. Neural Eng., 1: 63-71, 2004
10. J.R. Wolpaw, D.J. McFarland, T.M. Vaughan. "*Brain-computer interface research at the Wadsworth Center*". IEEE Transaction on Neural Systems and Rehab.,8(2): 222-226, 2003
11. G. N. Martin. "*Human Neuropsychology*", Prentice Hall, pp. 90, (1998)
12. J. Kandel, J. Schwartz and T. Jessel. "*Principles of Neural Science*", Elsevier, (1991)

Noise Cancellation in ECG Signals using Computationally Simplified Adaptive Filtering Techniques: Application to Biotelemetry

Md. Zia Ur Rahman

*Department of Electronics and Communication Engg.
Narasaraopeta Engg. College
Narasaraopet, 522601, India*

mdzr_5@yahoo.com

Rafi Ahamad Shaik

*Department of Electronics and Communication Engg.
Indian Institute of Technology
Guwahati, 781039, India*

rafiahamed@iitg.ernet.in

D V Rama Koti Reddy

*Department of Instrumentation Engineering
College of Engineering, Andhra University
Visakhapatnam, 530003, India*

rkreddy_67@yahoo.co.in

Abstract

Several signed LMS based adaptive filters, which are computationally superior having multiplier free weight update loops are proposed for noise cancellation in the ECG signal. The adaptive filters essentially minimize the mean-squared error between a primary input, which is the noisy ECG, and a reference input, which is either noise that is correlated in some way with the noise in the primary input or a signal that is correlated only with ECG in the primary input. Different filter structures are presented to eliminate the diverse forms of noise: 60Hz power line interference, baseline wander, muscle noise and the motion artifact. Finally, we have applied these algorithms on real ECG signals obtained from the MIT-BIH data base and compared their performance with the conventional LMS algorithm. The results show that the performance of the signed regressor LMS algorithm is superior than conventional LMS algorithm, the performance of signed LMS and sign-sign LMS based realizations are comparable to that of the LMS based filtering techniques in terms of signal to noise ratio and computational complexity.

Keywords: Adaptive filtering, Artifact, ECG, LMS algorithm, Noise cancellation.

1. INTRODUCTION

The extraction of high-resolution ECG signals from recordings contaminated with background noise is an important issue to investigate. The goal for ECG signal enhancement is to separate the valid signal components from the undesired artifacts, so as to present an ECG that facilitates

easy and accurate interpretation. Many approaches have been reported in the literature to address ECG enhancement [2]-[5]. In recent years, adaptive filtering has become one of the effective and popular approaches for the processing and analysis of the ECG and other biomedical signals. Adaptive filters permit to detect time varying potentials and to track the dynamic variations of the signals. Besides, they modify their behavior according to the input signal. Therefore, they can detect shape variations in the ensemble and thus they can obtain a better signal estimation.

Several papers have been presented in the area of biomedical signal processing where an adaptive solution based on the LMS algorithm is suggested [5]-[8]. The fundamental principles of adaptive filtering for noise cancellation were described by Widrow et al. [1]. Thakor and Zhu [5] proposed an adaptive recurrent filter to acquire the impulse response of normal QRS complexes, and then applied it for arrhythmia detection in ambulatory ECG recordings. The reference inputs to the LMS algorithm are deterministic functions and are defined by a periodically extended, truncated set of orthonormal basis functions. In these papers, the LMS algorithm operates on an "instantaneous" basis such that the estimate. In a recent study, however, a steady state convergence analysis for the LMS algorithm with deterministic reference inputs showed that the steady-state weight vector is biased, and thus, the adaptive estimate does not approach the Wiener solution. To handle this drawback another strategy was considered for estimating the coefficients of the linear expansion, namely, the block LMS (BLMS) algorithm [7], in which the coefficient vector is updated only once every occurrence based on a block gradient estimation. A major advantage of the block, or the transform domain LMS algorithm is that the input signals are approximately uncorrelated.

Complexity reduction of the noise cancellation system, particularly in applications such as wireless biotelemetry system has remained a topic of intense research. This is because of the fact that with increase in the ECG data transmission rate, the channel impulse response length increases and thus the order of the filter increases. Thus far, to the best of our knowledge, no effort has been made to reduce the computational complexity of the adaptive algorithm without affecting the signal quality. In order to achieve this, we considered the sign based adaptive algorithms. These algorithms enjoy less computational complexity because of the sign present in the algorithm. In the literature, there exist three versions of the signed LMS algorithm, namely, the signed regressor algorithm, the sign algorithm and the sign-sign algorithm. All these three require only half as many multiplications as in the LMS algorithm, thus making them attractive from practical implementation point of view [9]-[11]. In this paper, we considered the problem of noise cancellation and arrhythmia detection in ECG by effectively modifying and extending the framework of [5]. For that, we carried out simulations on MIT-BIH database. The simulation results shows that the performances of the sign based algorithms are comparable with LMS counterpart to eliminate the noise from ECG signals.

2. PROPOSED IMPLEMENTATION

When the doctors are examining the patient on-line and want to review the ECG of the patient in real-time, there is a good chance that the ECG signal has been contaminated by noise. The predominant artifacts present in the ECG includes: Power-line Interference (PLI), Baseline wander (BW), Muscle artifacts (MA) and Motion artifacts (EM), mainly caused by patient breathing, movement, power line interference, bad electrodes and improper electrode site preparation. The low frequency ST segments of ECG signals are strongly affected by these contaminations, which lead to false diagnosis. To allow doctors to view the best signal that can be obtained, we need to develop an adaptive filter to remove the noise in order to better obtain and interpret the ECG data.

2.1 Basic Adaptive Filtering Structure

Figure 1 shows an adaptive filter with a primary input that is an ECG signal s_1 with additive noise n_1 . While the reference input is noise n_2 , possibly recorded from another generator of noise n_2 that is correlated in some way with n_1 . If the filter output is y and the filter error $e = (s_1 + n_1) - y$, then

$$e^2 = (s_1 + n_1)^2 - 2y(s_1 + n_1) + y^2$$

$$= (n_1 - y)^2 + s_1^2 + 2s_1 n_1 - 2y s_1. \quad (1)$$

Since the signal and noise are uncorrelated, the mean-squared error (MSE) is

$$E[e^2] = E[(n_1 - y)^2] + E[s_1^2] \quad (2)$$

Minimizing the MSE results in a filter error output that is the best least-squares estimate of the signal s_1 . The adaptive filter extracts the signal, or eliminates the noise, by iteratively minimizing the MSE between the primary and the reference inputs.

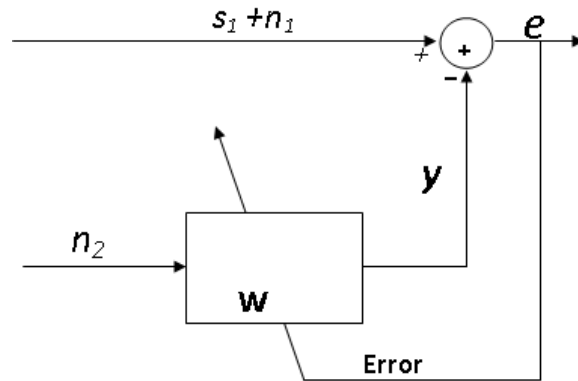


FIGURE 1: Adaptive Filter Structure.

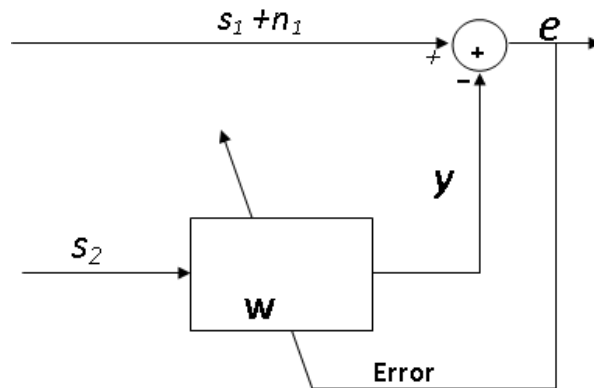


FIGURE 2: Alternate Adaptive Filter Structure.

Figure 2 illustrates another situation where the ECG is recorded from several electrode leads. The primary input $s_1 + n_1$ is a signal from one the leads. A reference signal s_2 is obtained from a second lead that is noise free. The signal s_1 can be extracted by minimizing the MSE between the primary and the reference inputs. Generally in biomedical signal processing the filter structure

shown in figure 1 is used, since it is difficult to obtain a noise free signal. Using the same procedure similar to (1) we can show that

$$E[e^2] = E[(s_1 - y)^2] + E[n_1^2] \quad (3)$$

Minimizing the MSE results in a filter error output y that is the best least-squares estimate of the signal s_1 .

2.2 Simplified Adaptive Algorithms

The LMS algorithm is a method to estimate gradient vector with instantaneous value. It changes the filter tap weights so that $e(n)$ is minimized in the mean-square sense. The conventional LMS algorithm is a stochastic implementation of the steepest descent algorithm. It simply replaces the cost function $\xi(n) = E[e^2(n)]$ by its instantaneous coarse estimate.

The error estimation $e(n)$ is

$$e(n) = \mathbf{d}(n) - \mathbf{w}(n) \Phi(n) \quad (4)$$

Coefficient updating equation is

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \Phi(n) e(n), \quad (5)$$

Where μ is an appropriate step size to be chosen as $0 < \mu < (2 / \text{tr } R)$ for the convergence of the algorithm.

The most important members of simplified LMS algorithms are:

The Signed-Regressor Algorithm (SRLMS): The signed regressor algorithm is obtained from the conventional LMS recursion by replacing the tap-input vector $x(n)$ with the vector $\text{sgn}\{x(n)\}$. Consider a signed regressor LMS based adaptive filter that processes an input signal $x(n)$ and generates the output $y(n)$ as per the following:

$$y(n) = \mathbf{w}^t(n)x(n), \quad (6)$$

where, $\mathbf{w}(n) = [w_0(n), w_1(n), \dots, w_{L-1}(n)]^t$ is a L-th order adaptive filter. The adaptive filter coefficients are updated by the Signed-regressor LMS algorithm as,

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \text{sgn}\{\Phi(n)\}e(n), \quad (7)$$

Because of the replacement of $\Phi(n)$ by its sign, implementation of this recursion may be cheaper than the conventional LMS recursion, especially in high speed applications such as biotelemetry these types of recursions may be necessary.

The Sign Algorithm (SLMS): This algorithm is obtained from conventional LMS recursion by replacing $e(n)$ by its sign. This leads to the following recursion:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \Phi(n) \text{sgn}\{e(n)\}, \quad (8)$$

The Sign – Sign Algorithm (SSLMS): This can be obtained by combining signed-regressor and sign recursions, resulting in the following recursion:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \text{sgn}\{\Phi(n)\} \text{sgn}\{e(n)\}, \quad (9)$$

where $\text{sgn}\{.\}$ is well known signum function,

$e(n) = d(n) - y(n)$ is the error signal.

The sequence $d(n)$ is the so-called desired response available during initial training period. The performance of these algorithms compared from the convergence characteristics shown in figure 3. From the convergence curves it is clear that the performance of the signed-regressor algorithm is only slightly worse than the conventional LMS algorithm. However the sign and sign – sign algorithms are both slower than the LMS algorithm. Their convergence behavior is also rather peculiar. They converge very slowly at the beginning, but speed up as the MSE level drops.

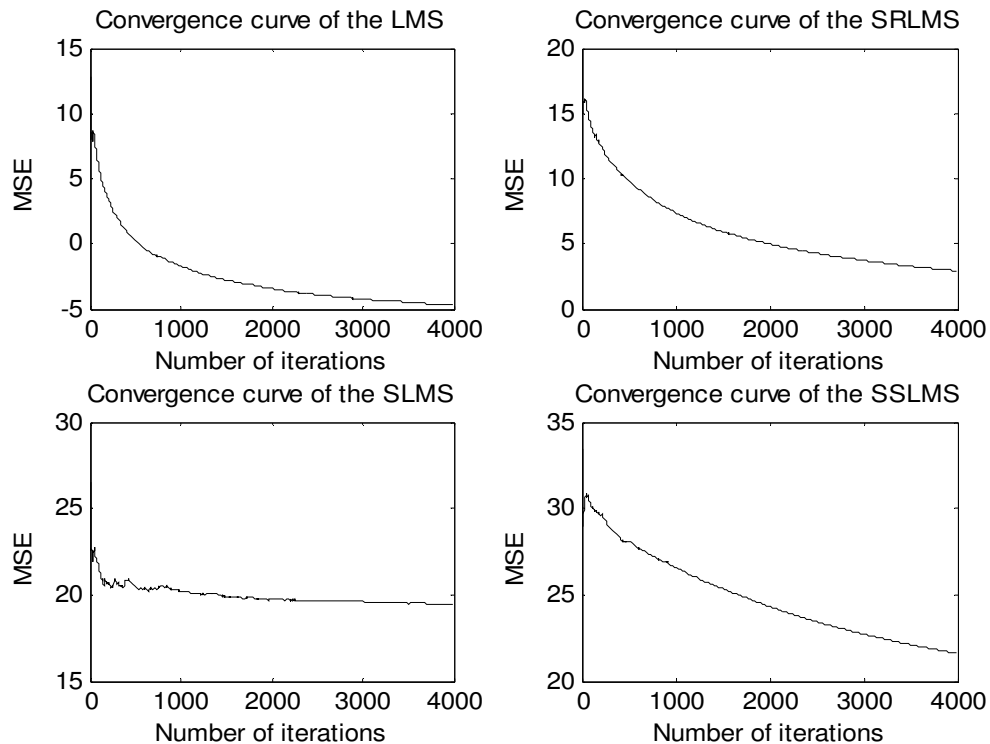


FIGURE 3: Convergence Characteristics of various algorithms

2.3 Noise Generator

The reference signal n_2 shown in figure 1 is taken from noise generator. A synthetic PLI with 1mv amplitude is simulated for PLI cancellation. No harmonics are synthesized. In order to test the filtering capability in non-stationary environment we have considered real BW, MA and EM noises. These are taken from MIT-BIH Normal Sinus Rhythm Database (NSTDB). This database was recorded at a sampling rate of 128Hz from 18 subjects with no significant arrhythmias. A random noise with variance of 0.001 is added to the ECG signals to evaluate the performance of the algorithm. The input SNR for the above non-stationary noise is taken as 1.25dB. In these three simplified algorithms because of the sign present in the recursion some tiny noise remains along the ST segment of the ECG signal. In order to extract the residual noise a tiny PLI is added to the noise reference signal. This improves the performance of the filter.

2.4 Computational Complexity Issues

The computational complexity figures required to compute all the three versions of sign LMS, as proposed above are summarized in Table 1, offers significant reduction in the number of operations required for LMS algorithm. Further, as these sign based algorithms are largely free from multiplication operation, these algorithms provides elegant means for removing the noise from the ECG signals. For LMS algorithm $L+1$ multiplications and $L+1$ additions are required to compute the weight update equation (5). In case of signed regressor algorithm only one multiplication is required to compute $\mu e(n)$. Where as other two signed LMS algorithms does not require multiplication if we choose μ value a power of 2. In these cases multiplication becomes shift operation which is less complex in practical realizations.

Algorithm	Multiplications	Additions	Shifts
LMS	$L+1$	$L+1$	Nil
SRLMS	1	$L+1$	Nil
SLMS	Nil	$L+1$	L
SSLMS	Nil	$L+1$	Nil

TABLE 1: A Computational Complexity Comparison Table.

3. SIMULATION RESULTS

To show that signed LMS algorithms are appropriate for ECG denoising we have used real ECG signals. We used the benchmark MIT-BIH arrhythmia database ECG recordings as the reference for our work. The data base consists of 48 half hour excerpts of two channel ambulatory ECG recordings, which were obtained from 47 subjects, including 25 men aged 32-89 years, and women aged 23-89 years. The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. In our simulation, first we collected 4000 samples of ECG signal. In this simulation μ for all the filters is chosen as 0.001 and the filter length as 5. For all the figures in this section *number of samples* is taken on x-axis and *amplitude* on y-axis, unless stated. Figure 4 shows the clean ECG signal (data105) and its frequency spectrum. In our experiments we have considered a dataset of five ECG records: data100, data105, data108, data203 and data228 to ensure the consistency of the results

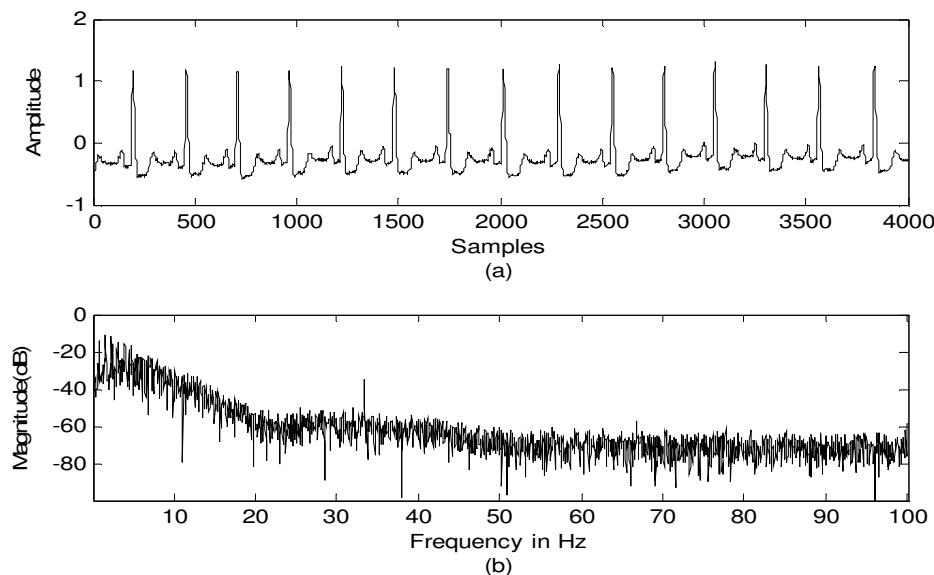


FIGURE 4: Clean ECG signal (data105) and its Spectrum.

3.1 Adaptive Power-line Interference (PLI) Cancellation

In this experiment, first we collected 4000 samples of ECG signal and corrupted with synthetic PLI with frequency 60Hz, sampled at 200Hz. This signal is applied as primary input to the adaptive filter shown in figure 1. The experiment is performed over the dataset average SNR improvement is considered to compare the performance of the algorithms. The reference signal is a synthesized PLI, the output of the filter is recovered signal. These results for data105 are shown in figure 5. Table 2 shows the SNR improvement for the dataset. In SNR measurements it is found that signed-regressor LMS algorithm gets average SNR improvement 29.5441dB, sign LMS gets 22.5405dB, sign-sign LMS improves 20.5345dB and conventional LMS algorithm

improves 31.0146dB. Figure 4 shows the power spectrum of the noisy signal before and after filtering with sign regressor LMS algorithm. The spectrum clears that the sign regressor LMS algorithm filters the PLI efficiently comparable to LMS filter with reduced number of computations.

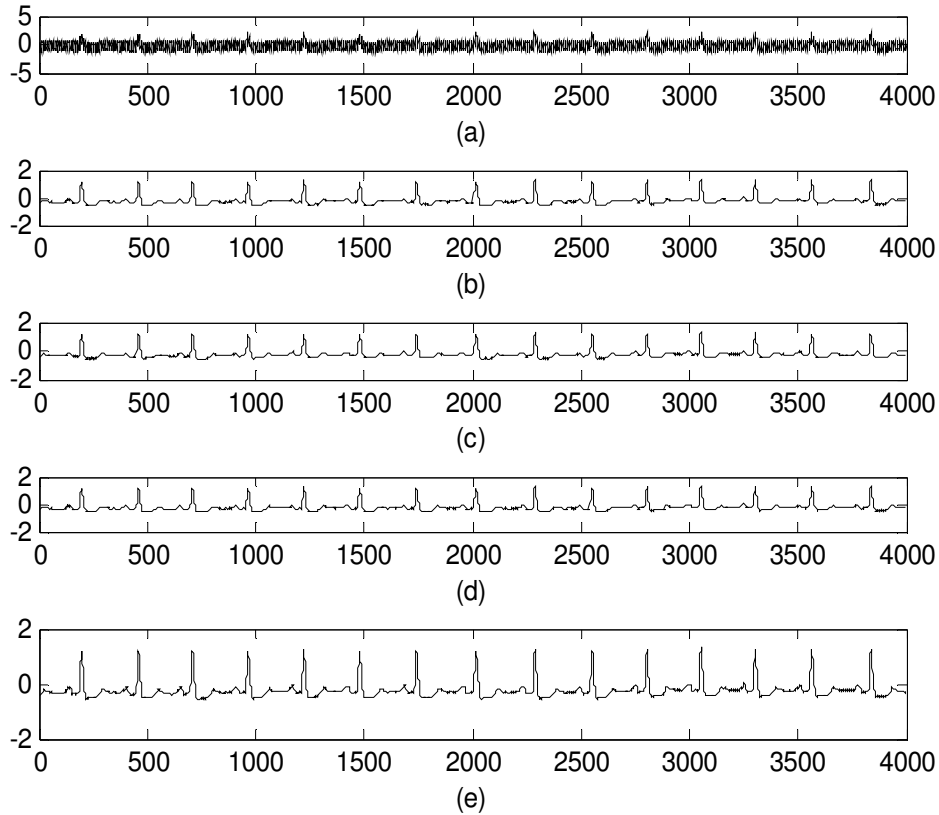


FIGURE 5: Typical filtering results of PLI Cancellation (a) MIT-BIH record 105 with 60Hz noise, (b) recovered signal using LMS algorithm, (c) recovered signal using signed regressor LMS algorithm, (d) recovered signal using sign LMS algorithm (e) recovered signal using sign sign LMS algorithm.

Rec. No	SNR Before Filtering	LMS		SRLMS		SLMS		SSLMS	
		SNR After Filtering	SNR Imp	SNR After Filtering	SNR Imp	SNR After Filtering	SNR Imp	SNR After Filtering	SNR Imp
100	-2.9191	28.7206	31.6397	26.6853	29.6044	17.8050	20.7241	14.1486	18.6195
105	-2.6949	28.5262	31.2211	26.9251	29.6200	20.3215	23.0164	18.0484	20.7433
108	-3.0647	28.4051	31.4698	26.4778	29.5425	22.4489	25.5136	19.3579	22.4226
203	-1.4531	27.3762	28.8293	26.8677	28.3208	18.5911	20.0442	17.1029	18.5560
228	-3.5242	28.3893	31.9135	27.1089	30.6331	19.8804	23.4046	18.8069	22.3311
Avg. (dBs)	-2.7312	28.2834	31.0146	26.8129	29.5441	19.8093	22.5405	17.4929	20.5345

TABLE 2: SNR Improvement of various algorithms for PLI Cancellation

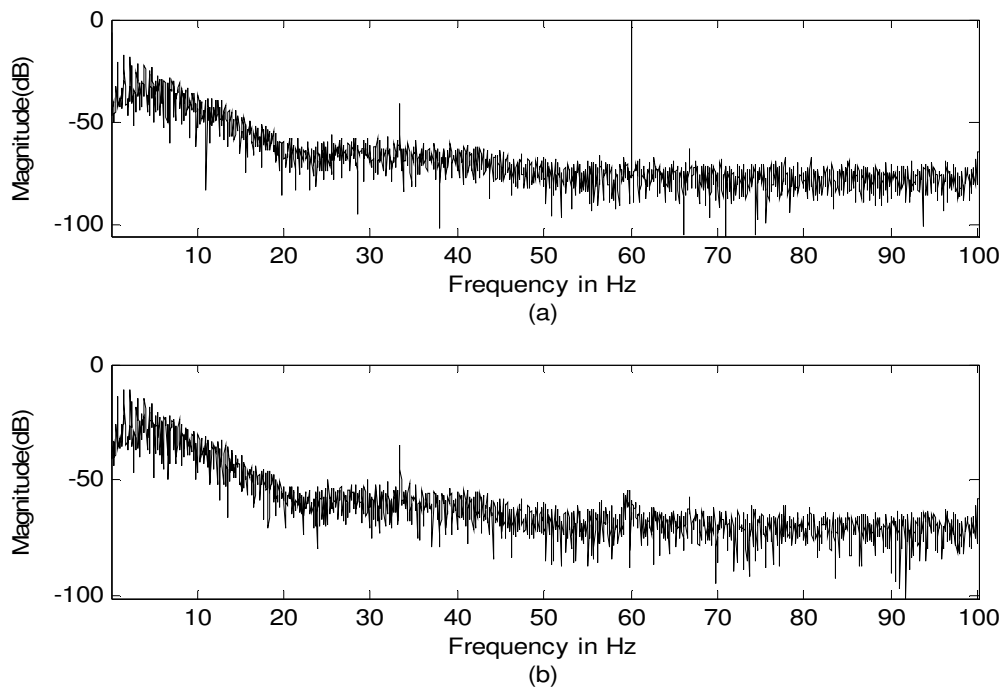


FIGURE 6: (a) Frequency spectrum of ECG with PLI, (b) Frequency spectrum after filtering with Sign regressor LMS algorithm.

3.2 Baseline Wander (BW) Reduction

In this experiment, first we collected 4000 samples of ECG signal (data105) and corrupted with real baseline wander (BW of MIT-BIH NSTDB), it is used as primary input to the adaptive filter of figure 1. The algorithms are applied on entire dataset. Simulation results for data105 are shown in figure 7. For the evaluating the performance of the proposed adaptive filter structures we have measured the average SNR improvement and compared with LMS algorithm. The sign-regressor LMS algorithm gets SNR improvement 10.1255dB, sign LMS gets 6.0443dB, sign-sign LMS improves 4.9937dB and conventional LMS algorithm improves 9.7282dB. Table 3 shows the SNR improvement for the dataset.

Rec. No	SNR Before Filtering	LMS		SRLMS		SLMS		SSLMS	
		SNR After Filtering	SNR Imp	SNR After Filtering	SNR Imp	SNR After Filtering	SNR Imp	SNR After Filtering	SNR Imp
100	1.2500	11.1571	9.9071	11.6220	10.3720	6.7036	5.4536	6.4829	5.2329
105	1.2500	12.3824	11.1324	13.1645	11.9561	8.0460	6.7960	6.4677	5.4177
108	1.2500	11.6224	10.3724	12.1420	10.8920	7.1091	5.8591	5.8679	4.6179
203	1.2500	6.8122	5.5622	6.6976	5.7260	6.4628	5.2128	5.0930	3.8430
228	1.2500	12.9172	11.6672	12.9314	11.6814	8.1500	6.9000	7.1053	5.8553
Avg. (dBs)	1.2500	10.9782	9.7282	11.3115	10.1255	7.2943	6.0443	6.2033	4.9937

TABLE 3: SNR Improvement of various algorithms for Baseline wander removal

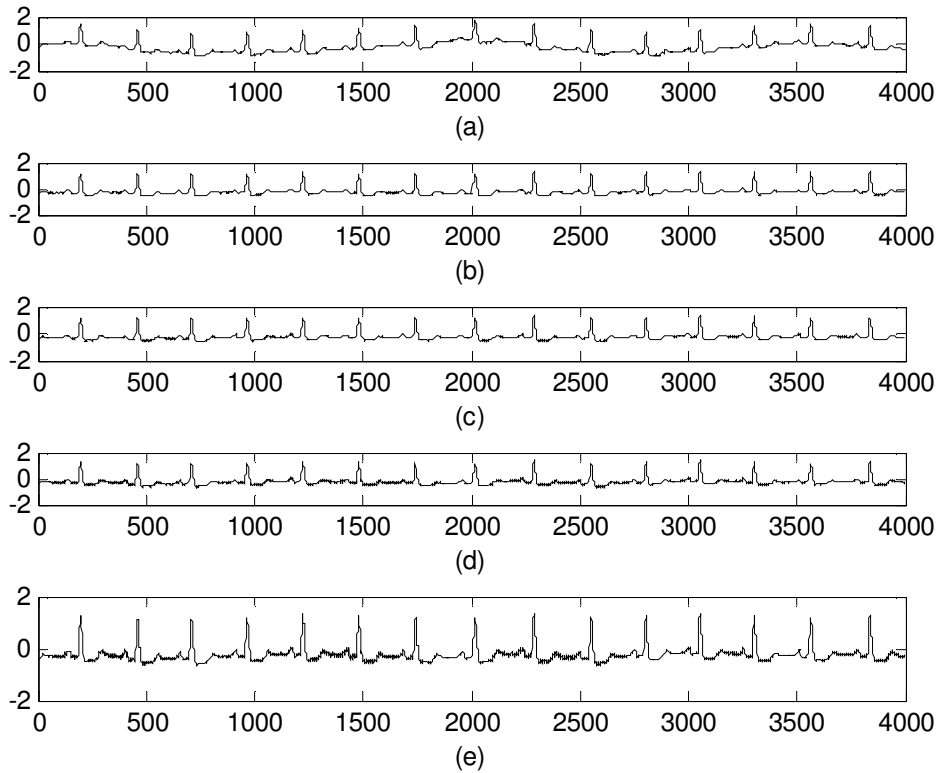


FIGURE 7: Typical filtering results of baseline wander reduction (a) MIT-BIH record 105 with real baseline wander, (b) recovered signal using LMS algorithm, (c) recovered signal using signed regressor LMS algorithm, (d) recovered signal using sign LMS algorithm, (e) recovered signal using sign sign LMS algorithm.

3.3 Muscle Artifacts (MA) Removal

The MA originally had a sampling frequency of 360Hz. The original ECG signal with MA is given as input to the adaptive filter. The results of data105 are shown in figure 8. The average SNR improvement of sign-regressor LMS algorithm is 12.2192dB, sign LMS gets 7.6995 dB, sign-sign LMS improves 6.9517dB and conventional LMS algorithm improves 11.4306dB. Table 4 shows the SNR improvement for the dataset.

Rec. No	SNR Before Filtering	LMS		SRLMS		SLMS		SSLMS	
		SNR After Filtering	SNR Imp	SNR After Filtering	SNR Imp	SNR After Filtering	SNR Imp	SNR After Filtering	SNR Imp
100	1.2500	11.4058	10.1558	12.3791	11.1291	7.8347	6.5847	7.0363	5.7863
105	1.2500	12.4265	11.1765	12.9827	11.7327	8.5680	7.3180	8.2148	6.9648
108	1.2500	12.3752	11.1252	13.4397	12.1897	8.0919	6.8414	7.4295	6.1795
203	1.2500	13.8786	12.6286	15.1749	13.9249	10.0800	8.8300	9.2585	8.0085
228	1.2500	13.3169	12.0669	13.3698	12.1198	10.1735	8.9235	9.0695	7.8195
Avg. (dBs)	1.2500	12.6806	11.4306	13.4692	12.2192	8.9496	7.6995	8.2017	6.9517

TABLE 4: SNR Improvement of various algorithms for adaptive cancellation of muscle artifacts

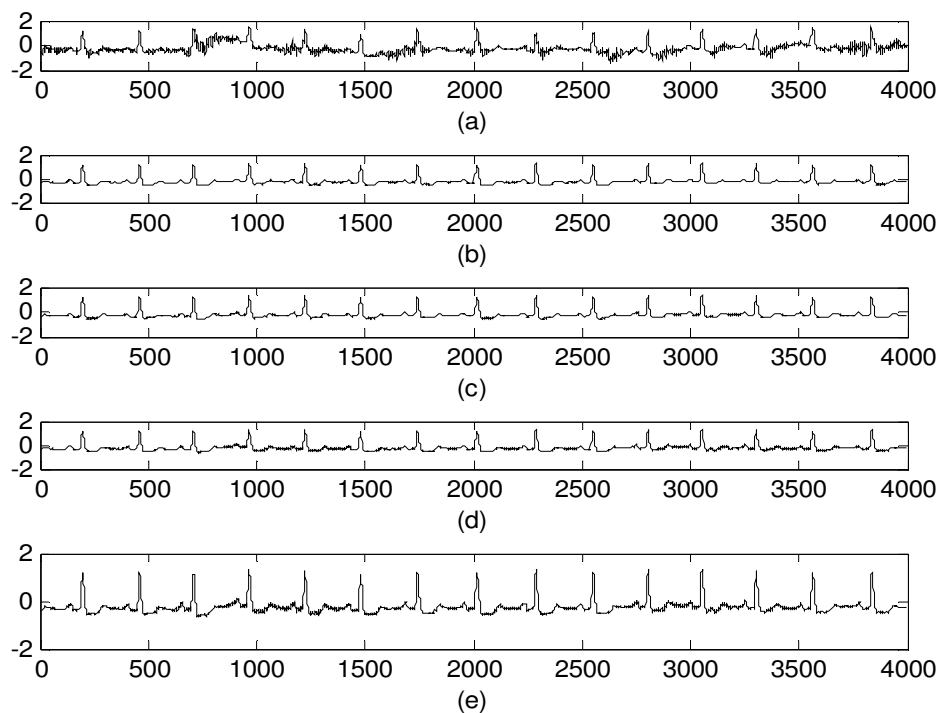


FIGURE 8: Typical filtering results of muscle artifacts removal (a) MIT-BIH record 105 with real muscle artifacts (b) recovered signal using LMS algorithm, (c) recovered signal using signed regressor LMS algorithm, (d) recovered signal using sign LMS algorithm, (e) recovered signal using sign sign LMS algorithm.

3.4 Motion Artifacts (EM) Removal

To demonstrate this we used MIT-BIH record number 105 ECG data with real electrode motion artifact (EM) added. The ECG signal corresponds to record 105 is corrupted with EM is given as input to the adaptive filter. The reference signal is taken from noise generator. The algorithms are tested for the dataset. Figure 9 shows the results correspond to data105. The average SNR improvements for various algorithms are 11.8950dB, 7.2525dB, 5.7464dB and 10.3374dB for signed regressor, sign, sign-sign and LMS algorithms respectively. Table 5 shows the SNR improvement for the dataset.

Rec. No	SNR Before Filtering	LMS		SRLMS		SLMS		SSLMS	
		SNR After Filtering	SNR Imp	SNR After Filtering	SNR Imp	SNR After Filtering	SNR Imp	SNR After Filtering	SNR Imp
100	1.2500	11.5749	10.3249	13.3180	12.0680	7.6309	6.3809	6.4164	5.1664
105	1.2500	12.5709	11.3209	14.4069	13.1569	8.2145	6.9645	6.7265	5.4765
108	1.2500	12.4709	11.1809	14.9770	13.7270	9.0952	7.8455	7.0101	5.7601
203	1.2500	8.9543	7.7043	10.4778	9.2278	8.6879	7.4379	7.0210	5.7710
228	1.2500	12.4062	11.1562	12.5457	11.2957	8.8840	7.6340	7.8080	6.5580
Avg. (dBs)	1.2500	11.5954	10.3374	13.1450	11.8950	8.5025	7.2525	6.9964	5.7464

TABLE 5: SNR Improvement of various algorithms for motion artifacts Cancellation.

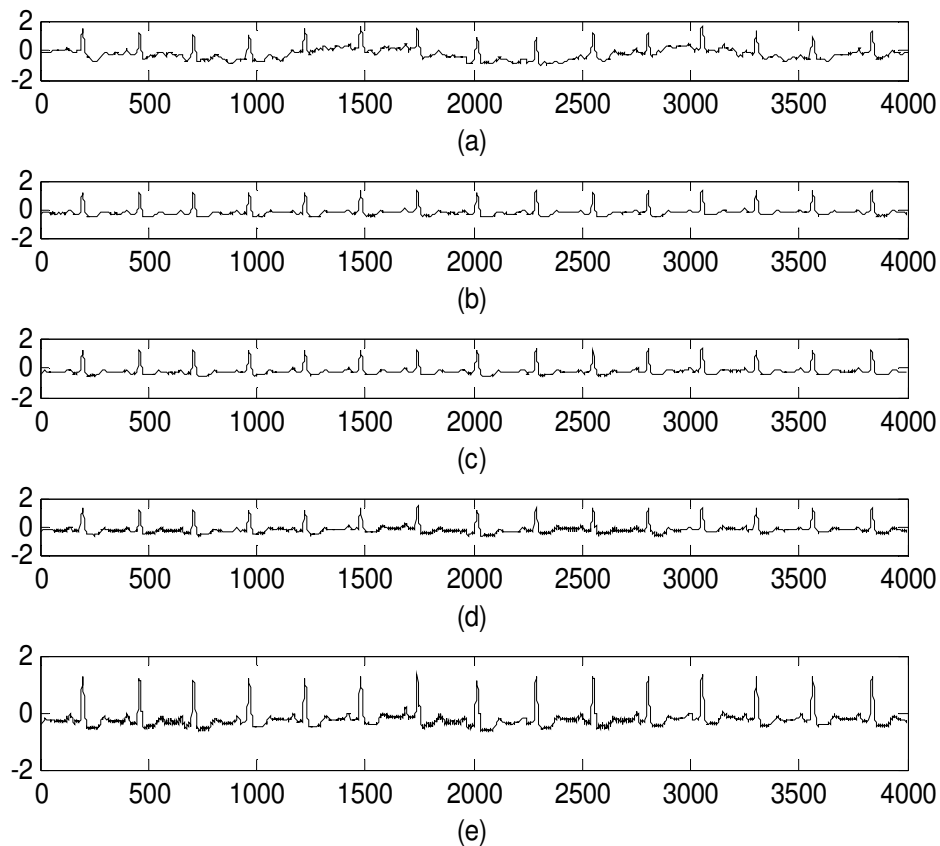


FIGURE 9: Typical filtering results of motion artifacts removal (a) MIT-BIH record 105 with real motion artifacts, (b) recovered signal using LMS algorithm, (c) recovered signal using signed regressor LMS algorithm, (d) recovered signal using sign LMS algorithm, (e) recovered signal using sign sign LMS algorithm.

4. CONCLUSION

In this paper the problem of noise removal from ECG using Signed LMS based adaptive filtering is presented. For this, the same formats for representing the data as well as the filter coefficients as used for the LMS algorithm were chosen. As a result, the steps related to the filtering remain unchanged. The proposed treatment, however exploits the modifications in the weight update formula for all categories to its advantage and thus pushes up the speed over the respective LMS-based realizations. Our simulations, however, confirm that the corresponding show-down effect with regard to the algorithm convergence is quit minor and is acceptable for all practical purposes. From the simulation results it is clear that the signed regressor LMS algorithm performs better than LMS in both SNR improvement and computational complexity, hence it is more suitable for wireless biotelemetry ECG systems.

5. REFERENCES

- [1] B. Widrow, J. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, and R. Goodlin, "Adaptive noise cancelling: Principles and applications", Proc. IEEE, vol. 63, pp. 1692-1716, Dec. 1975.
- [2] A. K. Barros and N. Ohnishi, "MSE behavior of biomedical event-related filters," IEEE Trans. Biomed. Eng., vol. 44, pp. 848-855, Sept. 1997.
- [3] O. Sayadi and M. B. Shamsollahi, "Model-based fiducial points extraction for baseline wander electrocardiograms," IEEE Trans. Biomed. Eng., vol. 55, pp. 347-351, Jan. 2008.
- [4] Y. Der Lin and Y. Hen Hu, "Power-line interference detection and suppression in ECG signal processing," IEEE Trans. Biomed. Eng., vol. 55, pp. 354-357, Jan. 2008.
- [5] N. V. Thakor and Y.-S. Zhu, "Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection," IEEE Transactions on Biomedical Engineering, vol. 38, no. 8, pp. 785-794, 1991.
- [6] Ziarani. A. K, Konrad. A, "A nonlinear adaptive method of elimination of power line interference in ECG signals", IEEE Transactions on Biomedical Engineering, Vol49, No.6, pp.540-547, 2002.
- [7] S. Olmos, L. Sornmo and P. Laguna, "Block adaptive filter with deterministic reference inputs for event-related signals: BLMS and BRLS," IEEE Trans. Signal Processing, vol. 50, pp. 1102-1112, May. 2002.
- [8] P. Laguna, R. Jane, S. Olmos, N. V. Thakor, H. Rix, and P. Caminal, "Adaptive estimation of QRS complex by the Hermite model for classification and ectopic beat detection," Med. Bio. Eng. Comput., vol. 34, pp. 58-68, Jan. 1996.
- [9] Farhang-Boroujeny, B., "Adaptive Filters- Theory and applications", John Wiley and Sons, Chichester, UK, 1998.
- [10] E. Eweda, "Analysis and design of a signed regressor LMS algorithm for stationary and nonstationary adaptive filtering with correlated Gaussian data," IEEE Transactions on Circuits and Systems, Vol. 37, No. 11, pp. 1367-1374, 1990.
- [11] S. Koike, "Analysis of Adaptive Filters using Normalized Signed Regressor LMS algorithm", IEEE Transactions on Signal Processing, Vol. 47, No. 1, pp. 2710-2733, 1999.

Integrated DWDM and MIMO-OFDM System for 4G High Capacity Mobile Communication

Shikha Nema

seeshikhanema@yahoo.co.in

*Department of Electronics and Communication Engineering,
MANIT(Deemed University)
Bhopal -462051,India*

Dr Aditya Goel

adityagoel2@rediffmail.com

*Department of Electronics and Communication Engineering,
MANIT(Deemed University)
Bhopal -462051,India*

Dr R P Singh

prof_rpsingh@rediffmail.com

*Department of Electronics and Communication Engineering,
MANIT(Deemed University)
Bhopal -462051,India*

Abstract

Dense wavelength-division multiplexing (DWDM) technique is a very promising data transmission technology for utilizing the capacity of the fiber. By DWDM, multiple signals (video, audio, data etc) staggered in wavelength domain can be multiplexed and transmitted down the same fiber. The Multiple-input multiple-output (MIMO) wireless technology in combination with orthogonal frequency division multiplexing (MIMO-OFDM) is an attractive air-interface solution for next-generation wireless local area networks (WLANs) and fourth generation mobile communication system. This paper provides an overview of the modified integrated DWDM MIMO-OFDM technology and focuses on DWDM transmitter design with adequate dispersion compensation for high data rate of 10Gbps, MIMO-OFDM system design and receiver design. The performance analysis in terms of bit error rate for Integrated system has also been carried out. Here a 64 channel DWDM system is simulated for transmission of baseband NRZ signal over fiber. Each of the transmission is at the bit rate of 10 Gbps leading to high data rate transmission of 640 Gbps. The resultant Bit Error Rate (BER) is in the range 10^{-12} for DWDM system which is given as input to MIMO-OFDM system. The system performance is analyzed in terms of BER with Signal to Noise Ratio(SNR) for Rayleigh and AWGN channels and desirable BER of 10^{-4} [3] is achieved at SNR of 10dB .

Keywords: DWDM system , 0.5nm channel spacing ,MIMO-OFDM system, Space Time Coding

1 Introduction

Tremendous consumer interest in multimedia applications requires high data rates in mobile communication system. With the advent of 4G mobile communication systems, many broadband wireless applications can be supported like Video Conferencing, Wireless Scada [1] and HDTV. High capacity and variable bit rate information transmission with high bandwidth efficiency are the key requirements that the modern transceivers have to meet in order to provide a variety of new high quality services to be delivered to the customers.

For achieving high capacity transmission, Optical fiber network plays an important role. **BROAD-BAND** millimeter-wave fiber-radio access system will meet demands for “*wireless first/last hop*” to the customers, which can support broad-band and portable services [2]. It will also resolve the scarcity of available microwave-band. For millimeter-wave fiber-radio systems, the only feasible option to connect between the central control office (CO) and the micro- or pico-cellular antenna base stations (BSs) would be an optical generation and transport technique of millimeter-wave RF signals over optical fiber links. In the micro- or pico-cellular fiber-radio access system, more than 1000 BSs are likely to be located under the coverage of a single CO; therefore, it would be desirable to accommodate a large number of BSs [3], and the promise for support will be wavelength division multiplexed (WDM) technology. Recently, there has been rapid progress in WDM transmission technologies. Dense WDM (DWDM) shows promise to increase the transmission capacity of trunk lines within the spectral regions limited by the gain bandwidths of optical fiber amplifiers.

The key challenge faced by future mobile communication system[16] is to provide high-data-rate wireless access at high quality of service (QoS). Combined with the facts that spectrum is a scarce resource and propagation conditions are hostile due to fading (caused by destructive addition of multipath components) and interference from other users, this requirement calls for means to radically increase spectral efficiency and to improve link reliability. Multiple-input multiple-output (MIMO) wireless technology [4] seems to meet these demands by offering increased spectral efficiency through spatial multiplexing gain, and improved link reliability due to antenna diversity gain. Even though there are still a large number of open research problems in the area of MIMO wireless, both from a theoretical perspective and a hardware implementation perspective, the technology has reached a stage where it can be considered ready for use in practical systems.

In this paper simulation is performed for 64 channel DWDM system integrated with MIMO-OFDM technology. The simulation is carried out using powerful software tools **Optisystem** and MATLAB. Section 2 describes the DWDM transmitter module which describes the channel properties of optical fiber with dispersion compensated fiber. Section 3 deals with MIMO-OFDM system design. Numerical results and analysis are provided in Section 4. Finally, Section 5 concludes the paper.

2 Design of 64 channel DWDM System

2.1 Transmitter Module

The transmitter module shown in Fig1 is divided into three parts. First part consists of sixty four NRZ Transmitters. For the generation of 10Gbps NRZ signal, a Pseudorandom bit sequence generator is used whose output in turn is given to a pulse generator to generate NRZ pulses. These Pulses are used to directly modulate externally Modulated LASER which operates at 1566nm wavelength and all subsequent sources are located at the wavelength difference of 0.5 nm. The Mech-Zehnder Modulator which is an intensity modulator based on an interferometric principle is used [8]. It consists of two 3 dB couplers which are connected by two waveguides of equal length. By means of an electro-optic effect, an externally modulated applied voltage can be used to vary the refractive indices in the waveguide branches. The different paths can lead to constructive and destructive interference at the output depending on the applied voltage. Then the output intensity can be modulated according to the voltage. The model implements a continuous wave (CW)

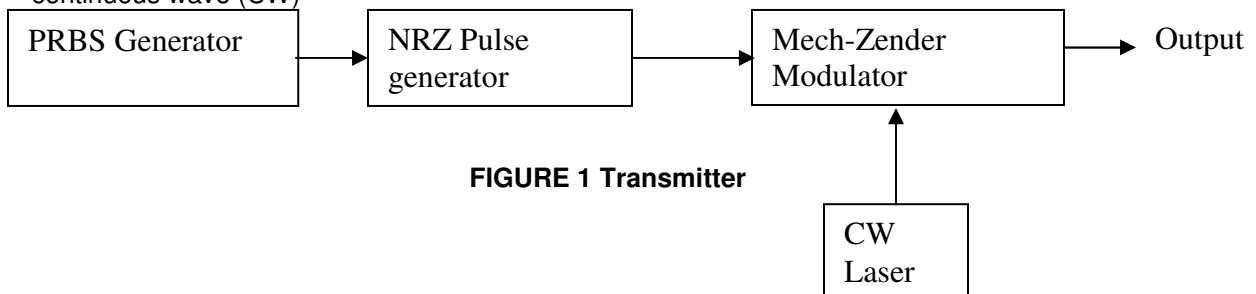


FIGURE 1 Transmitter

laser with phasor noise with overshoot and undershoot value of 30%. The output is provided to sixty four channel DWDM multiplexer operating at wavelength of 1566nm with channel spacing of 0.5 nm with line-width of 0.1 MHz.

2.2 Fiber Link Design

The output from the 64 channel DWDM Multiplexer is given to Single Mode fiber (SMF) of lengths 60 Km ,120Km and 240Km with Post dispersion compensated fiber of 10 Km, 20Km and 30Km respectively.[5]

The value of dispersion is different for different wavelengths so the exact value of dispersion at each wavelength for SMF and DCF is computed by using the formula :

$$D= WS (1-(w/W)^4)/4 \quad (1)$$

where:

D= Required value of Dispersion

S= Dispersion Constant

w= Reference Wavelength of the Fiber.

W= Wavelength at which Dispersion has to be calculated

Here the value of 'w' lies between 1295 to 1322 nm while the value of 'S' lies below 0.095 ps/((nm)²*km).



FIGURE 2 Fiber Design

Fig 2 shows the fiber link design of 60 Km with 50km of SMF and 10km of DCF. Polarization mode Dispersion (PMD) and deterministic birefringence with differential group delay of 0.2ps/Km are used to simulate the fiber. Two EDFA amplifiers are used with respective gains of 10db and 5 db to overcome the effect of attenuation.[5]

2.3 Receiver Design

Firstly the output from fiber is given to 64 channel DWDM demultiplexer. The output from each channel is fed to Fiber Bragg Grating which is a periodic or aperiodic perturbation of the effective refractive index in the core of an optical fiber. Typically, the perturbation is approximately periodic over a certain length of e.g. a few millimeters or centimeters, and the period is of the order of hundreds of nanometers, or much longer for *long-period fiber gratings* . The fiber core has a periodically varying refractive index over some length. The typical dimensions are 125 μm cladding diameter and 8 μm core diameter; periods of the refractive index gratings vary in the range of hundreds of nanometers or (for long-period gratings) hundreds of micrometers.[6,7]

The refractive index perturbation leads to the reflection of light (propagating along the fiber) in a narrow range of wavelengths, for which a *Bragg condition* is satisfied (Bragg mirrors):

$$\frac{2\pi}{\Lambda} = 2 \cdot \frac{2\pi n_{\text{eff}}}{\lambda} \Rightarrow \lambda = 2n_{\text{eff}}\Lambda \quad (2)$$

where Λ is the grating period, λ is the vacuum wavelength, and n_{eff} is the effective refractive index of light in the fiber. Essentially, the condition means that the wave number of the grating matches the difference of the (opposite) wave vectors of the incident and reflected waves. In that case, the

complex amplitudes corresponding to reflected field contributions from different parts of the grating are all in phase so that they can add up constructively; this is a kind of phase matching. Even a weak index modulation (with an amplitude of e.g. 10^{-4}) is sufficient for achieving nearly total reflection, if the grating is sufficiently long (e.g. a few millimeters).

It is shown that the dispersion of the neighboring transmitted channel may be determined uniquely by using FBG which is IIR filter. When considering dispersion effects on neighboring channels, dispersion may dictate channel spacing. [7] Light at other wavelengths, not satisfying the Bragg condition, is nearly not affected by the Bragg grating, except for some side lobes which frequently occur in the reflection spectrum (but can be suppressed by apodization of the grating). As a result of it, desired BER value of 10^{-12} can be obtained with small channel spacing of 0.5nm.

The receiver design shown in Fig3 consists of FBG whose output is fed into PIN photodetector having -3db gain with dark current of 10nA and responsivity of 1A/W which performs optical to electrical conversion. The electrical signal is fed into fourth order Low Pass Bessel filter having bandwidth of 7.5Ghz and depth of 100db. It is followed by 3R regenerator which consists of data recovery component and a NRZ pulse generator. The first output of Regenerator is the bit sequence, the second one is the modulated NRZ signal and the last output is reference input signal. These three signals can be connected directly to BER analyzer.

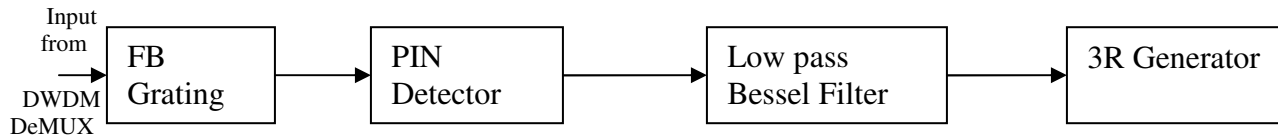


FIGURE 3 Receiver Design

Thus additional connections between transmitter and receiver stage are avoided. The output of BER analyzer gives the eye-diagram which gives the BER performance and Q factor for the system. The *MATLAB* code deals with the wireless part of this project. The output of the *OPTIWAVE* is a binary signal which is fed as an input to the *MATLAB* code for MIMO-OFDM system. [8]

3 MIMO-OFDM System

Traditionally, multiple antennas (at one side of the wireless link) have been used to perform interference cancellation and to realize diversity and array gain through coherent combining. The use of multiple antennas at both sides of the link offers an additional fundamental gain — spatial multiplexing gain, which results in increased spectral efficiency. A brief review of the techniques in a MIMO system is given in the following.

Spatial multiplexing yields a linear (in the minimum of the number of transmit and receive antennas) capacity increase, compared to systems, with a single antenna at one or both side of the link, at no additional power or bandwidth expenditure [3, 9]. The corresponding gain is available if the propagation channel exhibits rich scattering and can be realized by the simultaneous transmission of independent data streams in the same frequency band. The receiver exploits differences in the spatial signatures induced by the MIMO channel onto the multiplexed data streams to separate the different signals, thereby realizing a capacity gain.

Diversity leads to improved link reliability by rendering the channel “less fading” and by increasing the robustness to co-channel interference. Diversity gain is obtained by transmitting the data signal over multiple (ideally) independently fading dimensions in time, frequency, and space and by performing proper combining in the receiver. Spatial (i.e., antenna) diversity is particularly attractive when compared to time or frequency diversity, as it does not incur an

expenditure in transmission time or bandwidth, respectively. Space-time coding [2] realizes spatial diversity gain in systems with multiple transmit antennas without requiring channel knowledge at the transmitter. **Array gain** can be realized both at the transmitter and the receiver. It requires channel knowledge for coherent combining and results in an increase in average receive signal-to-noise ratio (SNR) and hence improved coverage.

3.1 Space time Coding

Diversity combining technique is implemented in the system by using space-time block codes. To get the idea of Space-Time Block Codes, it is comfortable to investigate the scheme for two transmit antennas. The information data is mapped to either PSK or QAM symbols and is divided in blocks of two symbols. To explain the functionality, two consecutive time steps of such a block are observed. (Fig-4)[10]

In the first time step the signals X_1 and X_2 are transmitted simultaneously from the first and the second antenna. In the next time step, the signals $-X_2^*$ and X_1^* are transmitted, so that we achieve the given code word matrix, which consists of orthogonal columns.

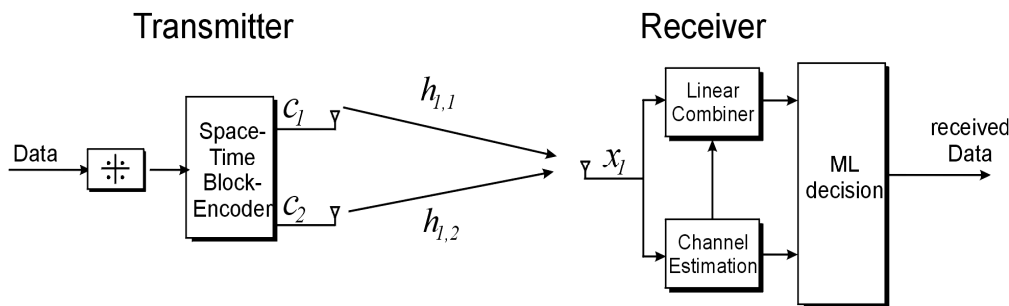


FIGURE 4 Space time block Coding

The received signals is described in the two timeslots by the given formula, which can be simplified by conjugating the term for describing the second received signal.

$$y[k] = h_0x_0[k]+h_1x_1[k]+n[k] \quad (3)$$

Three terms are received depending only on the fading gains, the transmitted signal and the noise. [2] The bandwidth efficiency challenge requires novel solutions in both the network and physical layers. The latter could include powerful coding and modulation methods, transmission adaptation techniques, and antenna configurations. MIMO communications based on multiple transmit and receive antenna is a very promising technique to increase bandwidth efficiency, and is seen as a potential key solution for fading channels with rich enough scattering. MIMO technology will predominantly be used in broadband systems that exhibit frequency-selective fading and, therefore, intersymbol interference (ISI). OFDM modulation turns the frequency-selective channel into a set of parallel flat fading channels and is, hence, an attractive way of coping with ISI. Fig 5 and 6 depicts the schematic of a MIMO-OFDM system. The basic principle that underlies OFDM is the insertion of a guard interval, called cyclic prefix (CP), which is a copy of the last part of the OFDM symbol, and has to be long enough to accommodate the delay spread of the channel.

The use of the CP turns the action of the channel on the transmitted signal from a linear convolution into a cyclic convolution, so that the resulting overall transfer function can be diagonalized through the use of an IFFT at the transmitter and an FFT at the receiver. Consequently, the overall frequency-selective channel is converted into a set of parallel flat fading channels, which drastically simplifies the equalization task. However, as the CP carries redundant information, it incurs a loss in spectral efficiency, which is usually kept at a maximum of 25 percent.[4,11]

In general, OFDM has tighter synchronization requirements than single-carrier (SC) modulation and direct-sequence spread spectrum (DSSS), is more susceptible to phase noise, and suffers from a larger peak-to-average power ratio.[10,11]

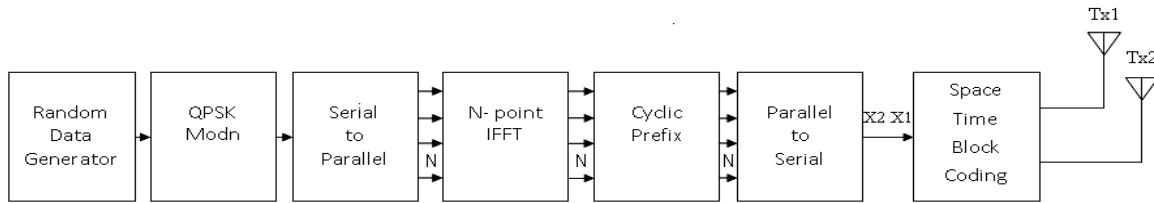


FIGURE 5 MIMO-OFDM Transmitter

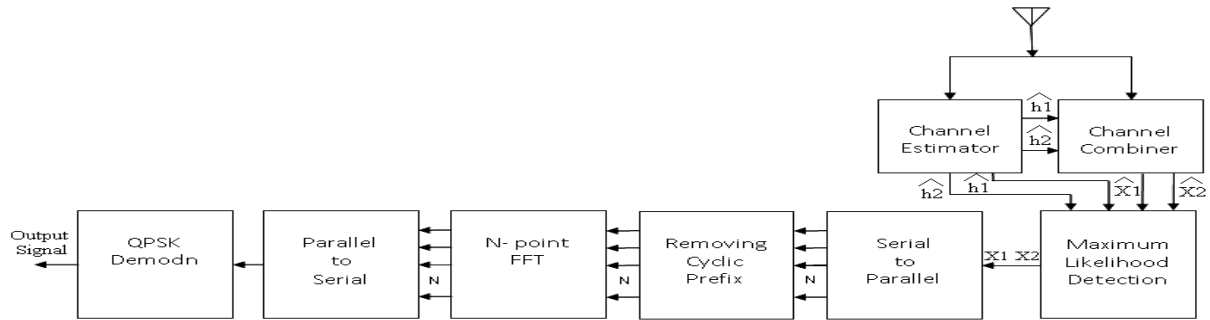


FIGURE 6 MIMO-OFDM Receiver

In MIMO-OFDM transmitter the binary signal is first converted into 4-ary signal so that the signal can be QPSK modulated. This modulated signal has to be converted into time domain. For this purpose the signal is first passed through a serial to parallel conversion block which takes 32 symbols at a time and converts them into parallel stream of data. Then this parallel data is used for taking IFFT which converts the signal into time domain.

In an OFDM signal the frequencies are orthogonal to each other. But still during transmission some amount of noise would be added to the signal due to multipath propagation. In order to reduce this Cyclic Prefix is used. Cyclic prefix are often used in conjunction with modulation in order to retain sinusoids properties in multipath channels. It is well known that sinusoidal signals are Eigen functions of linear and time-invariant systems. Therefore, if the channel is assumed to be linear and time-invariant then a sinusoid of infinite duration would be an Eigen function. However, in practice, this cannot be achieved, as real signals are always time-limited. So, to mimic the infinite behavior, prefixing the end of the symbol to the beginning makes the *linear convolution* of the channel appear as though it were *circular convolution*, and thus, preserve this property in the part of the symbol after the cyclic prefix.[12] After prefixing the signal is transmitted using Space Time Coding (*Almouti Scheme*) [2]. Here *spatial diversity* is used where a signal is passed through 2 antennas and both of them follow different path and the best path is chosen by the receiver resulting in 2:1 MIMO OFDM system. The same is the case when 4 antennas are used (4:1). [2]

The signal when passed through a channel is acted upon by noise. Two different channels are considered according to the type of noise that is added. The first one is *Rayleigh channel* where noise is in complex form and both the real and the imaginary part are *Gaussian* variables. The second is an ideal case i.e. *AWGN channel* which has a constant power spectral density.[13] The performances for these channels are seen by plotting graphs of *BER VS SNR*. After the signal is received Cyclic Prefix is removed. Then the signal is QPSK demodulated and converted from decimal to binary to obtain the original information.

4 Results and Discussion:

Bit error rate is calculated for NRZ system at three different distance of 60km , 120km and 240 km, at the wavelength of 1566nm Fig 7 shows the Q factor for NRZ signal at wavelength of 1566nm. It shows that Q factor of 8 is achieved at half the bit period .It is shown experimentally in [14] that Q factor of 7 to 8 is achieved at the data rate of 5 Gbps for a single channel system employing pre/post Dispersion compensation scheme with channel spacing of 1nm .Here the same performance is achieved with more stringent parameters i.e. data rate of 10Gbps for the channel spacing of 0.5nm. Fig 8 shows the bit error rate performance for the same system at 1566nm. It shows that BER of 10^{-16} is achieved at half the bit period

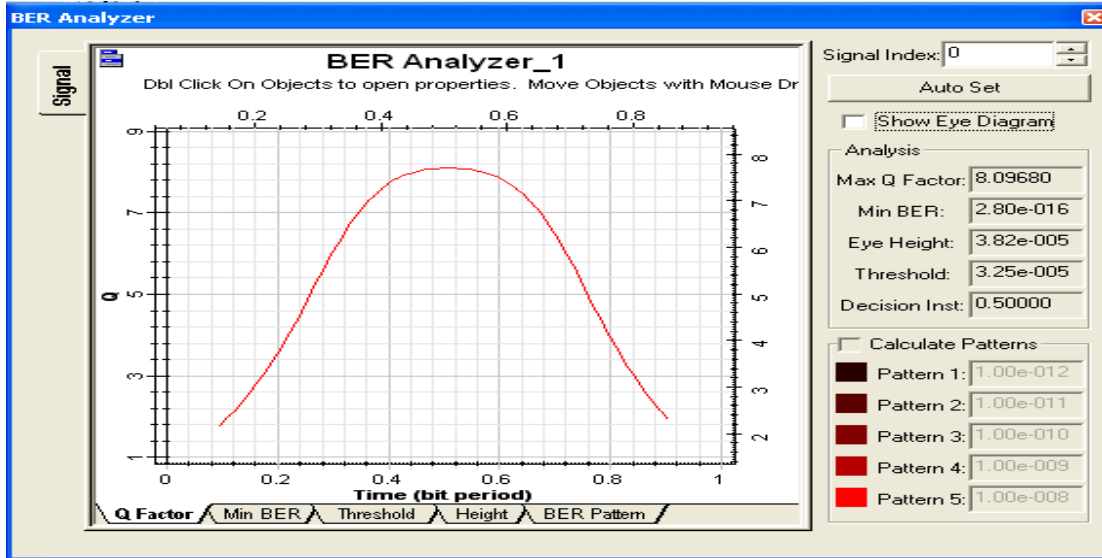


FIGURE 7 Q factor v/s bit period for NRZ signal at 1566nm

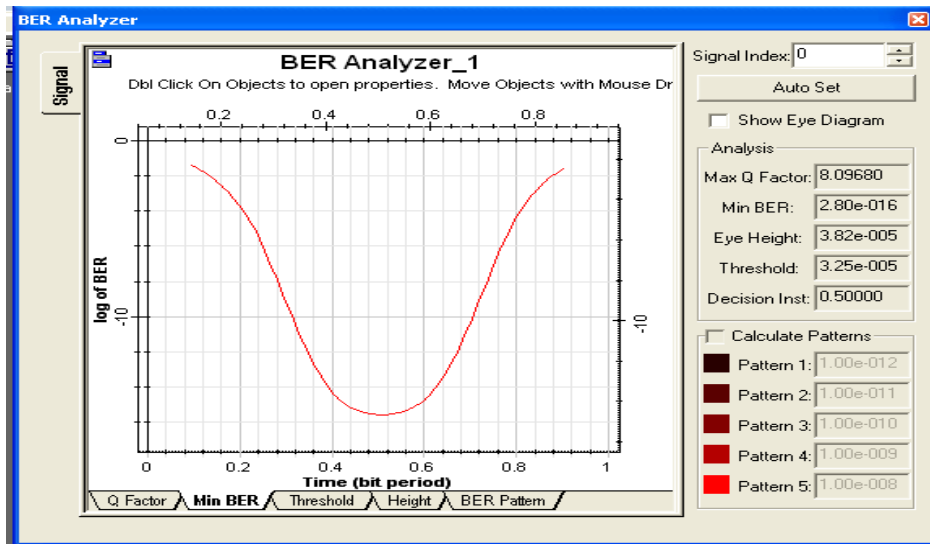


FIGURE 8: Log of BER v/s bit period for NRZ signal at 1566nm

Fig 9 shows the eye diagram with minimum BER value which is in the range of 10^{-16} . The eye opening shows that the signal is free of intersymbol Interference and can be detected at the centre of the bit period.

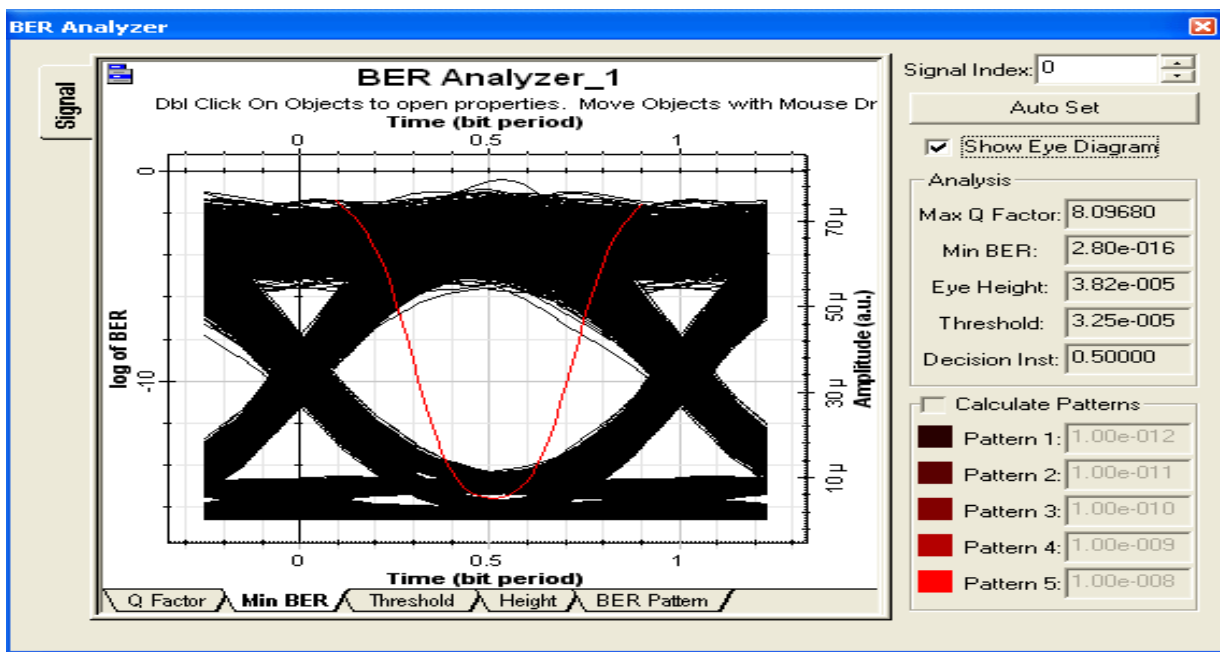


FIGURE 9: Eyediagram for NRZ at wavelength 1566nm at 60 KM

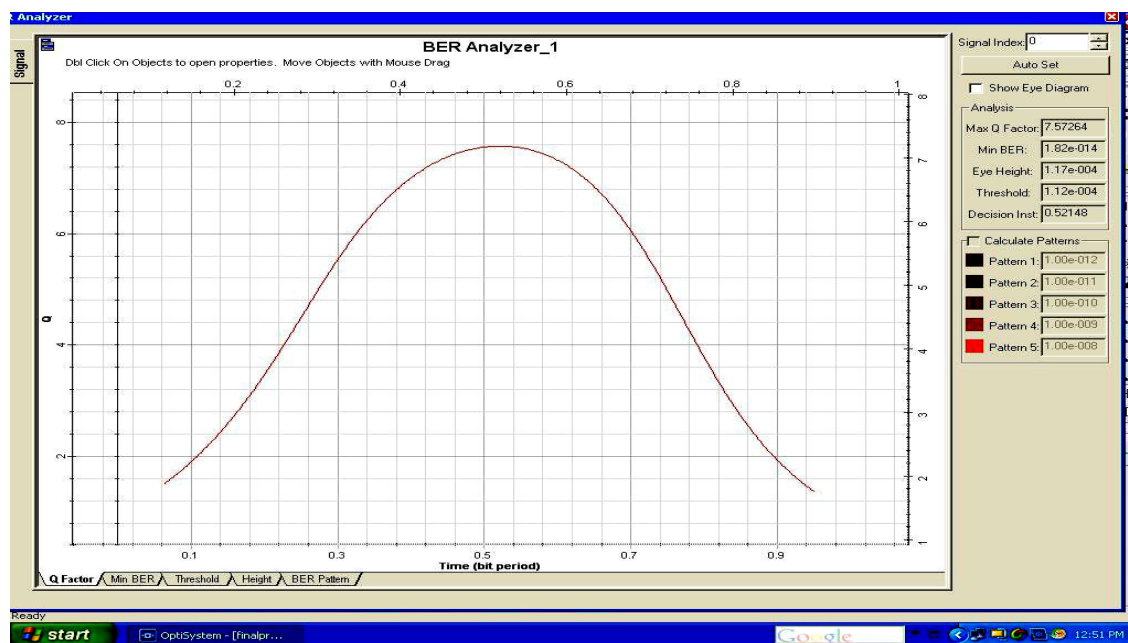


FIGURE 10 Q factor v/s Bit period for NRZ system at bandwidth of 1566nm at the distance of 120 KM.



FIGURE 11 Q factor v/s Bit period for NRZ system at bandwidth of 1566nm at the distance of 240KM

Second NRZ Output is taken at the distance of 120 and 240Km whose Q factor plots are shown in Fig 10 and 11. Q Factor of 7.5 and BER value of 10^{-14} is achieved for 120KM while maximum Q factor of 7 is achieved in the system with min BER of 10^{-11} for 240KM. Hence it can be concluded that as the distance increases, Q factor decreases. Improvement in Q factor can be achieved by increasing the transmit power of CW laser source.[14,15]

The received signal is then converted into a binary form and then modulated again so that it can be used as an input to the transmitters of the MIMO-OFDM system. After passing through desired channel model, the receiver then receives the signal and the signal is then checked for BER again at different values of Signal to Noise Ratio. The BER was found to be within acceptable limits. The BER v/s SNR curves for the systems with 2 transmit antennas and 4 transmit tested with the Rayleigh and AWGN channels are shown in fig12 to fig 15.

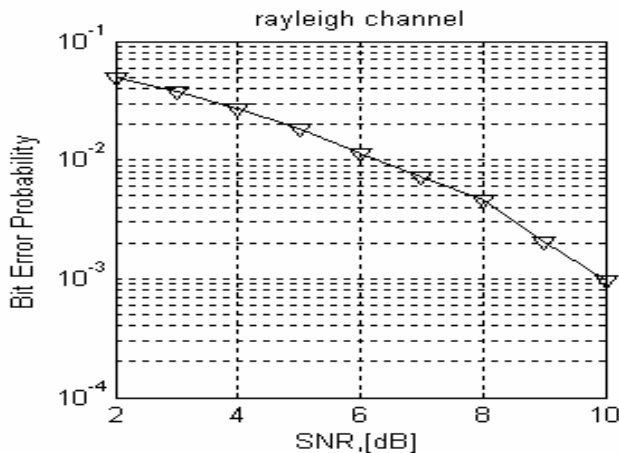


FIGURE 12 BER v/s SNR plot for rayleigh channel with 2 :1 MIMO OFDM system

From fig 12 and 13 , it can be seen that BER value of 10^{-3} is achieved at 10 db of SNR and AWGN channel performance is better as compared to Rayleigh Channel for two transmit antenna and one receive antenna MIMO-OFDM system.

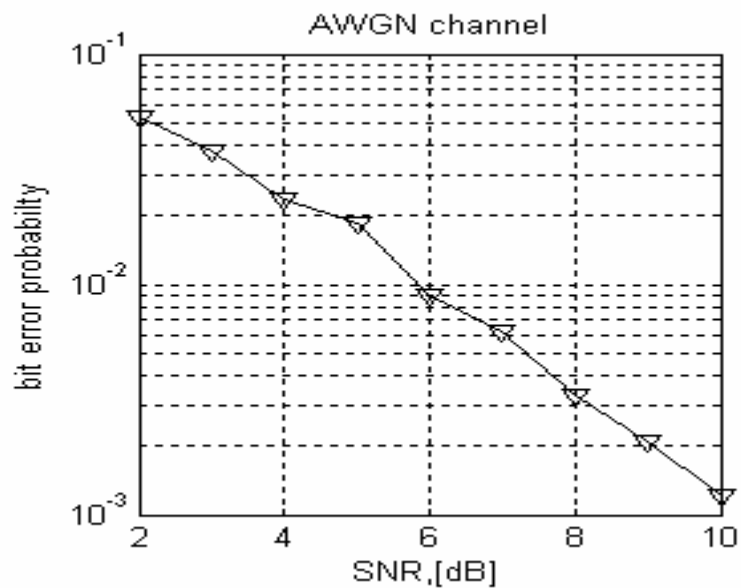


FIGURE 13 BERv/s SNR plot for AWGN channel with 2 :1 MIMO OFDM system

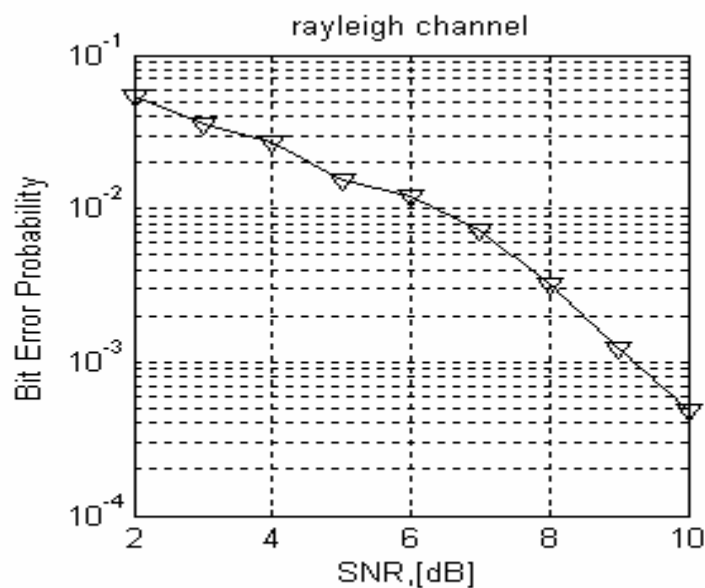


FIGURE 14 BERv/s SNR plot for rayleigh channel with 4 :1 MIMO OFDM system

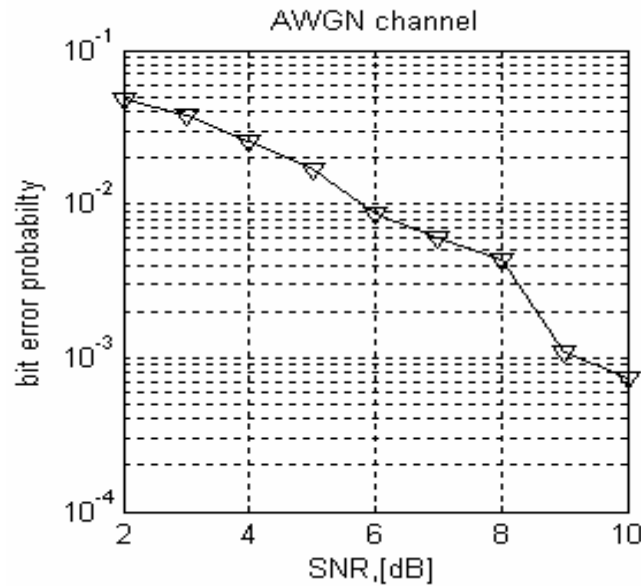


FIGURE 15 BER v/s SNR plot for AWGN channel with 4 :1 MIMO OFDM system

From Fig 14 and Fig 15 it can be seen that BER performance of $10^{-3.5}$ is achieved at SNR of 10db with four transmit and one receive antenna which is better as compared to two transmit and one receive antenna .

We have designed 64 channel DWDM system with data rate of 10Gbps and 0.5nm Channel spacing. BER and Q factor are observed for various fiber spans of 60km, 120km and 240km. It is obvious from Fig 9 and Fig 11 that for the fiber span of 60km BER achieved is 10^{-16} which is increased to 10^{-11} for the fiber span of 240km which is within the acceptable range.[14] For MIMO-OFDM system, analysis is performed for Rayleigh and AWGN Channel. Rayleigh Channel will provide a typical multipath environment in urban area for mobile communication while AWGN channel will provide additive white Gaussian noise. The simulation is performed for two transmit and one receive antenna and four transmit and one receive antenna. For both these cases, and under the different channel conditions it can be seen from fig 12-15 that BER of $10^{-3.5}$ is obtained for SNR of 10db. Hence it is obvious that for an integrated DWDM-MIMO-OFDM system an average BER of 10^{-12} is achieved for DWDM optical system and 10^{-4} is achieved with MIMO-OFDM system which is within the acceptable range for optical and wireless system respectively.[4,14]

5 Conclusion

Error Free DWDM transmission over 300 Km of Single mode fiber is simulated with 64 channel EML system with 0.5 nm channel spacing. Performance analysis for NRZ transmission is done. Minimum BER achieved with NRZ is 10^{-16} for 60 KM which is 10^{-11} for 240 KM of fiber length. This binary output is given to MIMO-OFDM system designed in MATLAB .Two MIMO-OFDM system is designed with two transmit antenna and one receive antenna and four transmit antenna and one receive antenna. Results which are plotted for various channels are shown above. These results are tabulated in Table 1. It can be inferred that the system with four transmit antennas gives better performance than two transmit antennas. This BER performance is achieved at the higher data rate of 10Gbps . The system performance can be further improved by incorporating Error correction technique in the system

Table1

Sr.NO	MIMO System	SNR	Rayleigh Channel	AWGN Channel
1	4:1	10	10-3.5	10-3.1
2	2:1	10	10-3	10-3

References

1. Aditya Goel and Ravi Shankar Mishra “Remote Data Acquisition using Wireless SCADA System”, International Journal of Engineering, Volume 3 (1),pp.58-65, 2009
2. S.Almouti, “ A simple transmit diversity technique for wireless communication” IEEE journal on select areas in communications, Oct 1998
3. Foschini, “Layered space time Architecture for wireless communication”, Bell labs Tech Journal Vol 1,1996
4. Markku Juntti, Mikko Vehkaperä, Jouko Leinonen, Zexian Li, and Djordje Tujkovic “MIMO MC-CDMA communications for future cellular systems”, IEEE Communication Magazine, Feb 2005.
5. D.Wake, L.Noel, D.G.Moodie, D.D.Marcenac, L.D.Westbrook and D.Neset “A 60 GHz 120 Mbps QPSK fiber-radio transmission experiment incorporating an electroabsorption modulator transceiver for a full duplex optical data path”*IEEE 1997 MTT-S Digest*
6. Volkan Kaman, Xuezhong Zheng, Shifu Yuan, Jim Klingshirn, Chandrasekhar Puserla, Roger Helkey, Olivier Jerphagnon, and John E. Bowers, “A 32 10 Gb/s DWDM Metropolitan Network Demonstration Using Wavelength-Selective Photonic Cross-Connects and Narrow-Band EDFAs” *IEEE Photonics Technology Letters*, volume 17(9) september 2005
7. G. Lenz, B. J. Eggleton, C. R. Giles, C. K. Madsen, and R. E. Slusher, “Dispersive Properties of Optical Filters for WDM Systems” IEEE journal of quantum electronics, volume 34(8) August 1998 pp1390-1402
8. Aditya Goel, R K Sethi “ Integrated Optical wireless network for next generation Wireless Systems” , Signal Processing – An International Journal Volume 3 (1), pp.1-13, 2009
9. V Tarokh, Hamid Jafarkhani and A Robert Calderbank , “Space time Block Coding for wireless Communications :Performance Results” IEEE Journal on Selected Area in Communications, volume 17(3) March 1999
10. A. J. Paulraj, R. U. Nabar, and D. A. Gore, Introduction to Space-Time Wireless Communications, Cambridge, UK: Cambridge Univ. Press, 2003.
12. Rane Manzoor, Regina Gani, Varun Jeoti, Nidal kamal, Muhammad Asif , “ Dwpt based FFT and its application to SNR estimation in OFDM systems” Signal Processing– An International Journal Volume 3 (2) , 2009
13. T. Rappaport “Wireless Communication” 2nd edition ,Prentice Hall Publication December 2001
14. Fariborz Mousavi Madani and Kazuro Kikuchi, “ Design Theory of Long-Distance WDM Dispersion-Managed Transmission System” ,Journal of lightwave technology, volume 17(8) pp1326-1335 , August 1999
15. Zhang Dechao, Li Xiaolin, Zhang Xiaoru, Wang Ziyu, Xu Anshi Chen Zhangyuan, Li Hongbin, Li Zhengbin, “43 Gb/s DWDM Optical Transmission System Using NRZ Format and Electro-absorption Modulation” *IEEE 2006*
16. Aditya Goel and A. Sharma, “Performance Analysis of Mobile Ad-hoc Network using AODV protocol”, International Journal of Computer Science and Security, Vol. 3(5),2009

A New Method for Pitch Tracking and Voicing Decision Based on Spectral Multi-scale Analysis

Mohamed Anouar Ben Messaoud

anouar.benmessaoud@yahoo.fr

*Electrical Engineering Department
National School of Engineers of Tunis
Le Belvédère BP. 37, 1002 Tunis, Tunisia*

Aïcha Bouzid

bouzidacha@yahoo.fr

*Electrical Engineering Department
National School of Engineers of Tunis
Le Belvédère BP. 37, 1002 Tunis, Tunisia*

Nouredine Ellouze

n.ellouze@enit.rnu.tn

*Electrical Engineering Department
National School of Engineers of Tunis
Le Belvédère BP. 37, 1002 Tunis, Tunisia*

Abstract

This paper proposes a new method for voicing detection and pitch estimation. This method is based on the spectral analysis of the speech multi-scale product. The multi-scale product (MP) consists of making the product of the speech signal wavelet transform coefficients. The wavelet used is the quadratic spline function. The spectrum of the multi-scale product analysis reveals rays corresponding to the fundamental frequency and its harmonics. We evaluate our approach on the Keele University database. The experimental results show the effectiveness of our method comparatively to the state-of-the-art algorithms.

Keywords: Speech, Wavelet transform, Multi-scale Product, Pitch, Voicing detection.

1. INTRODUCTION

Pre-processing of speech signal is very crucial in the applications where silence or background noise is completely undesirable. Applications like speech and speaker recognition [1] needs efficient feature extraction techniques from speech signal where most of the voiced part contains speech or speaker specific attributes. Silence removal is a well known technique adopted for many years for this and also for dimensionality reduction in speech that facilitates the system to be computationally more efficient. This type of classification of speech into voiced or silence/unvoiced sounds [2] finds other applications mainly in fundamental frequency estimation, formant extraction or syllable marking and so on.

More over, the fundamental frequency is an important parameter in the speech analysis and synthesis. It plays an eminent role in the speech production and perception. In application areas such as speech enhancement, analysis and prosody modeling, low-bit rate coding, and speaker recognition, a reliable pitch estimation is required [3].

A wide variety of sophisticated voicing classification and pitch detection algorithms have been proposed in the speech processing literature [4], [5], [6], [7], [8] and [13].

The voicing decision and pitch estimation from speech signal only are basically done by relying on different types of speech transformation. This transformation can be operated following three domains:

The first approach works in the time domain. The common transformation is the autocorrelation function like the YIN algorithm and the Praat Software application [9], [10] and [11].

The second approach works in the frequency domain. The frequently used transformation is the spectrum [12] and [13].

The third approach combines both time and frequency domains, using the Short Time Fourier Transform (STFT) or the Wavelet Transform (WT) [14].

In this paper, we propose and evaluate a new algorithm for voicing classification and pitch determination operated on a clean speech signal. We are motivated by the work developed in [15] and [16], where the multi-scale product-based approach constitutes an efficient method for glottal closure instant detection. These instants delimit the pitch period.

This paper is organized as follows: Section 2 reminds some properties of the continuous wavelet transform and the multi-scale product method for edge detection. In section 3, we detail our approach for voicing decision and pitch estimation. In section 4, experimental results are presented using the Keele University database. Finally, we conclude this work.

2. MULTI-SCALE ANALYSIS

Wavelet Transform [17], [18] is introduced as an alternative technique for analyzing non stationary signal. It provides a new way for representing the signal into well-behaved expression that yields useful properties. The wavelet is a square integrable function well localised in time and frequency, from which we can extract all basis functions by time shifting and scaling.

Dyadic wavelet transform, is a particular case of continuous wavelet transform when the scale parameter is discretized along the dyadic grid (2^j), $j \in \mathcal{Z}$.

The wavelet transform can be used for various applications as edge detection, noise reduction and parameter estimation. When the mother wavelet function is the n th derivative of a smoothing function, it acts as a differential operator. The number of wavelet vanishing moments gives the order of the differentiation. For an appropriately chosen wavelet, the wavelet transform modulus maxima denote the points of sharp variations of the signal [19]. Wavelet transform which is the first derivative of a smoothing function is proved to be convenient for discontinuity detection in a signal.

The wavelet transform is a multi-scale analysis which has been shown to be very well suited for speech processing in many applications as glottal closure instant (GCI) detection, pitch estimation [20], speech enhancement and recognition and so on.

To improve edge detection by wavelet transform, we use a non linear combination of wavelet transform coefficients. The multi-scale product (MP) consists of making the product of wavelet transform coefficients of the function $f(n)$ at some successive dyadic scales as follows [21]

$$p(n) = \prod_{j=j_0}^{j=j_L} w_{2^j} f(n) \quad (1)$$

Where $w_{2^j} f(n)$ is the wavelet transform of the function f at scale 2^j . This is distinctly a non linear function of the input time series $f(n)$.

Singularities produce cross-scale peaks in wavelet transform coefficients, which are reinforced in the product $p(n)$. Although particular smoothing levels may not be optimal, the non linear combination tends to reinforce the peaks while suppressing spurious noisy peaks. The signal

peaks align across scale for the first few scales, but not for all scales because increasing the amount of smoothing will spread the response and cause singularities separated in time to interact. Thus choosing scales too large will result in misaligned peaks in $p(n)$. Odd number of terms in $p(n)$ preserves the sign of maxima [22]. Choosing the product of three levels of wavelet decomposition is generally optimal and allows detection of small peaks.

This is intended to enhance multi-scale peaks due to edge, while suppressing noise, by exploiting the multi-scale correlation due to the presence of the desired signal. Bouzid et al. prove that the MP is very efficient for glottal closure and opening instants detection from speech signal only [16].

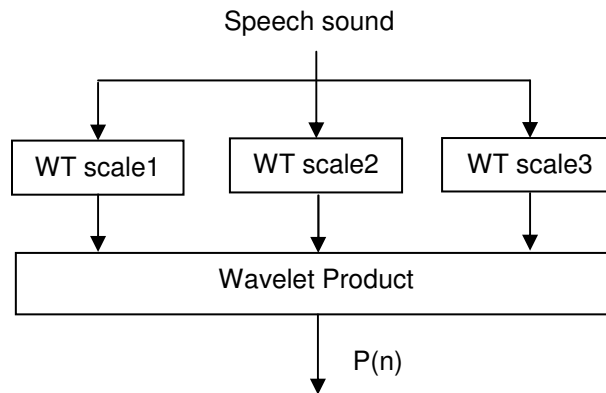


FIGURE 1: The Multi-scale Product scheme.

3. THE PROPOSED METHOD

We propose a new technique to localize voiced sounds with an estimation of the fundamental frequency in the case of a clean speech signal. The method is based on the spectral analysis of the speech multi-scale product (SPM).

Our method can be decomposed in four essential steps. The first step consists of computing the product of wavelet transform coefficients of the speech sound. The wavelet used in this multi-scale product analysis is the quadratic spline function at scales $s_1=2^{-1}$, $s_2=2^0$ and $s_3=2^1$.

The second step consists of calculating the fast Fourier transform (FFT) of the obtained signal over windows with a specific length of 4096 samples. In deed, the product is decomposed into frames of 1024 samples with an overlapping of 512 points at a sampling frequency of 20 kHz.

In fact, the product $p[n]$ is divided into frames of N length by multiplication with a sliding analysis window $w[n]$:

$$p_w[n, i] = p[n] w[n - i\Delta n] \quad (2)$$

Where i is the window index, and Δn the overlap. The weighting $w[n]$ is assumed to be non zero in the interval $[0, N-1]$. The frame length value N is chosen in such a way that, on the one hand, the parameters to be measured remain constant and, on the other hand, there are enough samples of $p[n]$ within the frame to guarantee reliable frequency parameter determination.

The third step consists of identifying voiced frames in a speech waveform. And the last step consists of giving an estimation of the pitch frequency for the detected voiced frames. Theses two steps will be detailed in the next subsections.

3.1 Voicing Decision

Figures 2 and 4 show the multi-scale product of the voiced speech signal and the unvoiced one respectively. For the first case, the MP has a periodic structure unlike the second case. The figure 3 shows the SMP corresponding to the voiced speech signal. However the figure 5 illustrates the

SMP corresponding to the unvoiced sound depicted in figure 4. We note clearly the difference between the two cases. So, a voicing detection approach can be derived.

After calculating the FFT of the speech MP in the i th frame, we localize all the peaks stored in the vector P_i . We eliminate ones that don't belong to the following frequency range $[F_{0min} F_{0max}]$. If there is no peaks, the frame is declared unvoiced, else we calculate the distance separating two successive peak positions $D_{ij}=P_{ij+1}-P_{ij}$ constituting the D_i vector elements. These elements are ranked in the growing order to compose the E_i vector. To make a voicing decision, we look for well defined groups constituted from the E_i vector.

The groups are sorted as follows:

If $E_{i1}-E_{i2}<10$, so E_{i1} and E_{i2} are in the same group G_{i1} and we calculate $E_{i1}-E_{i3}$, else, E_{i1} is in G_{i1} and E_{i2} is in G_{i2} . Then, we calculate $E_{i2}-E_{i3}$ and so on until reaching the last elements in the E_i vector.

Once the groups are formed, we look for their number N_i . If $N_i=1$, the i th frame is voiced. If $N_i=2$ and $(\text{card}(G_{i2})<2/3*\text{card}(G_{i1}))$, the i th frame is also declared voiced, else, the frame is unvoiced.

The voicing decision diagram is given in the figure 6.

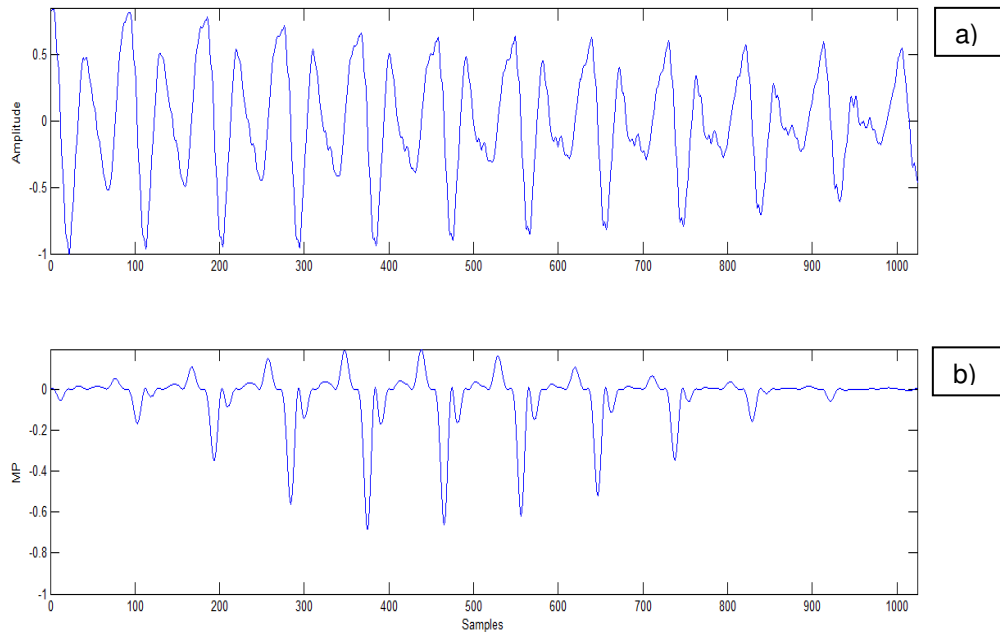


FIGURE 2: a) Voiced speech of a female speaker. b) its Multi-scale Product.

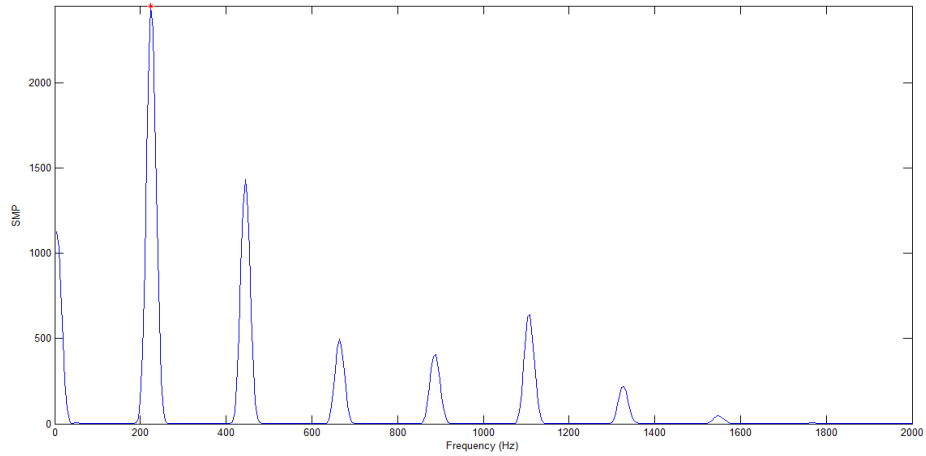


FIGURE 3: Spectral Multi-scale Product Analysis of the voiced speech signal 2(a).

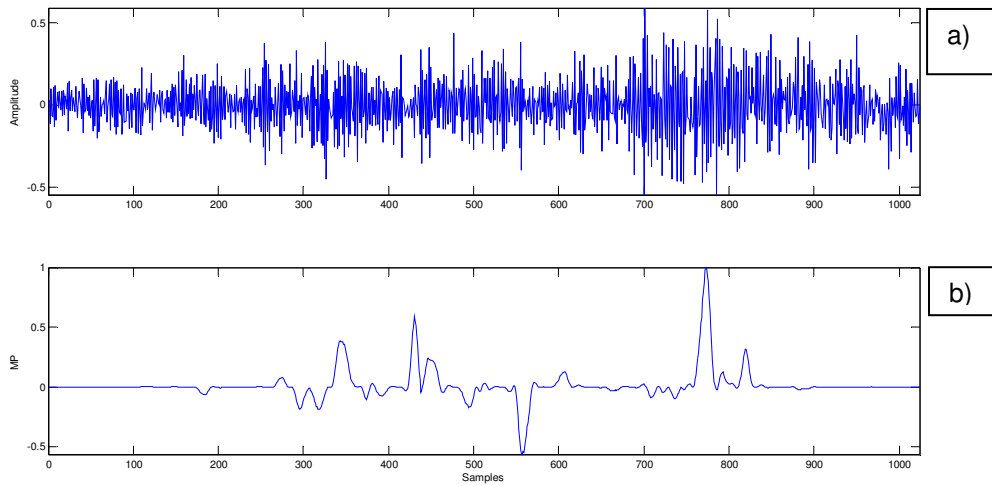


FIGURE 4: a) Unvoiced speech of a female speaker. b) its Multi-scale Product.

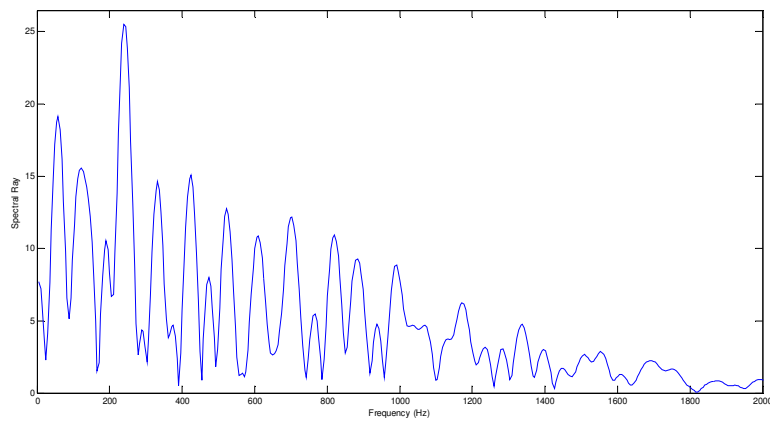


FIGURE 5: Spectral Multi-scale Product Analysis of the unvoiced speech signal 4(a).

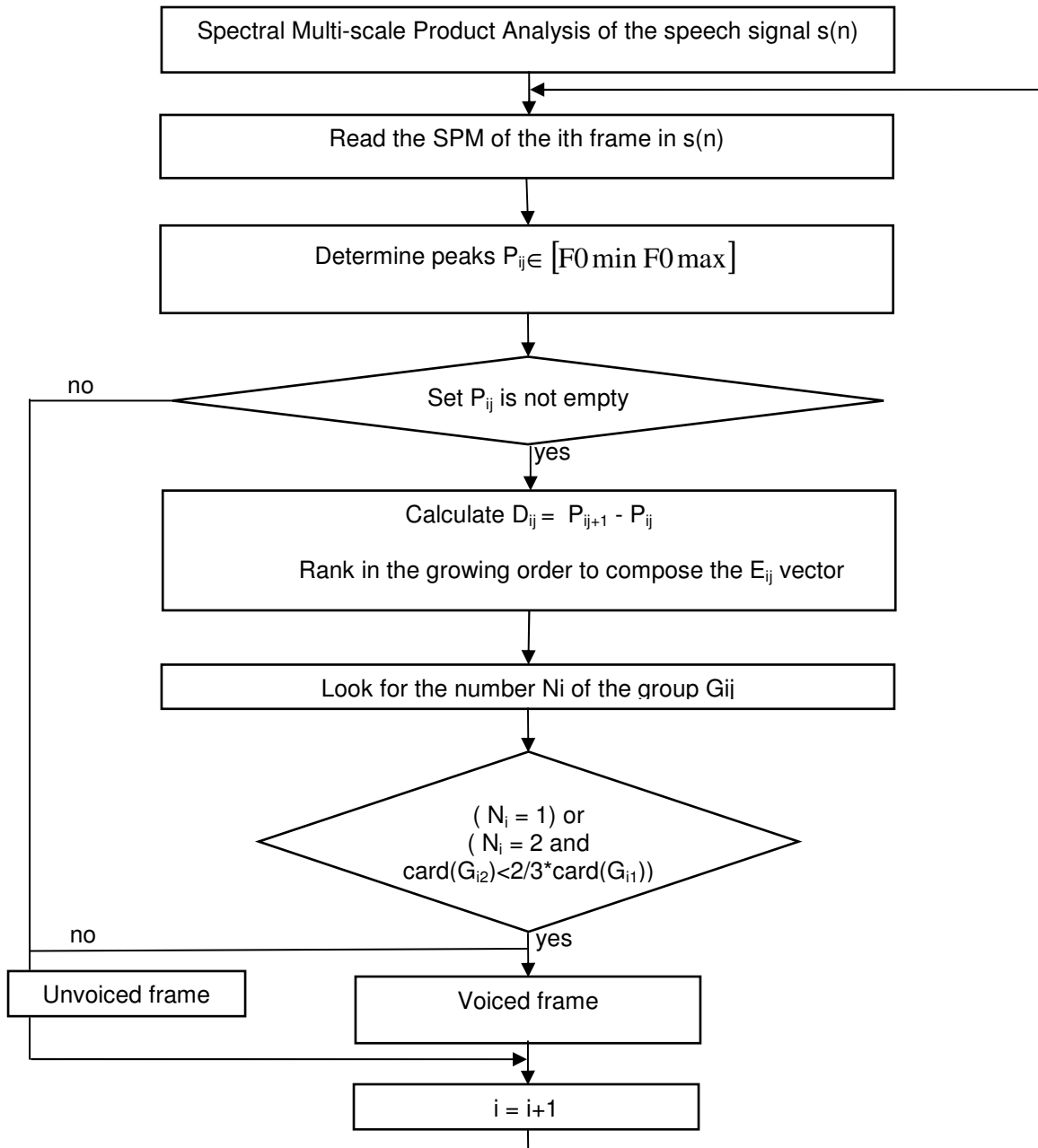


FIGURE 6: Algorithm of the proposed voicing classification approach.

3.2 Pitch Estimation

Pitch estimation is ensured on the i th voiced frame detected by the proposed approach. The fundamental frequency is the first element in the group G_{i1} described in the previous subsection.

4. Experiments and Results

To evaluate the performance of our algorithm, we use the Keele pitch reference database [19]. This database consists of speech signals of five male and five female English speakers each reading the same phonetically balanced text with varying duration between about 30 and 40 seconds. The Keele database includes reference files containing a voiced–unvoiced segmentation and a pitch estimation of 25.6 ms segments with 10 ms overlapping. The reference files also mark uncertain pitch and voicing decisions. The reference pitch estimation is based on a simultaneously recorded signal of a laryngograph. Uncertain frames are labelled using a negative flag.

For the evaluation of the voicing classification approach, we calculate the error decision probabilities that comprises unvoiced frames detected as voiced, and voiced frames detected as unvoiced as proposed in [9]. Table 1 reports evaluation results for voicing classification of the proposed method in a clean environment. We compare our method to other state-of-the-art algorithms [8], [24], [25] and [26] that are based on the same reference database. As can be seen, our method yields very good results in comparison with well known approaches with the lowest V-UV rate of 2.3%.

Methods	V-UV (%)
<i>Proposed Method</i>	2.3
RAPT [8]	3.2
NMF [24]	7.7
MLS [25]	7.0
Seg-HMM [26]	8.4

TABLE 1: Performance comparison of some methods for voicing classification.

For pitch estimation and according to Rabiner [27], the gross pitch error (GPER) denotes the percentage of frames at which the estimation and the reference pitch differ by more than 20%.

Table 2 lists the GPER of our proposed approach compared to others as PRAAT, YIN, and CEPSTRUM for male and female speakers and all the Keele database.

As can be seen, our approach yields good results encouraging us to use it in other hard environments. In fact, the SMP method shows a low GPE rate of 0.75% for all the database.

Methods	Cep	PRAAT	YIN	<i>Proposed</i>
	GPE (%)	GPE (%)	GPE (%)	GPE (%)
Female Speakers	4.2	3.3	1.2	0.4
Male Speakers	3.7	2.9	3.5	1.1
Total	3.95	3.1	2.35	0.75

TABLE 2: GPER for pitch estimation using Keele University database.

5. CONCLUSION

In this work, we propose a novel voicing decision and pitch estimation algorithm. This algorithm is based on the spectral analysis of the multi-scale product made by multiplying the wavelet transform coefficients of the speech signal.

The proposed approach can be summarised in four essential steps. First, we make the product of the speech wavelet transform coefficients at three successive dyadic scales (The wavelet is the quadratic spline function with a support of 0.8 ms). Second, we compute the short time Fourier transform of the speech multi-scale product. Thirdly, we select the entire peaks found in the frame spectrum. These peaks are gathered satisfying some criteria. Consequently a decision is made concerning the voicing state and a pitch estimation is given. The experimental results show the efficiency of our approach for clean speech in comparison with the state-of-the-art algorithms.

6. REFERENCES

1. J.P. Campbell. "Speaker Recognition : A Tutorial". In Proceedings of the IEEE, 85(9): 1437--1462, 1997
2. A. Martin, D. Charlet and L. Mauuary. "Robust Speech/ Non-speech Detection Using LDA Applied to MFCC". In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 1: 237--240, 2001
3. D. O. Shaughnessy. "Speech communications: human and machine". IEEE Press, NY, second edition, (2000)
4. D.G. Childers, M. Hahn and J.N. Larar. "Silence and Voiced/Unvoiced/Mixed Excitation Classification of Speech". IEEE Trans. On Acoust., Speech , Signal Process, 37(11):1771--1774, 1989
5. L. Liao and M. Gregory. "Algorithms for Speech Classification". In Proceedings of the 5th ISSPA, Brisbane, 1999
6. W. J. Hess. "Pitch and voicing determination", Marcel Dekker, Inc., pp. 3-48 (1992)
7. P. C. Bagshaw, S. M. Hiller and M. A. Jack. "Enhanced pitch tracking and the processing of f0 contours for computer aided intonation teaching". In Proceedings of the 3rd European Conference on Speech Communication and Technology, 1993
8. D. Talkin. "A robust algorithm for pitch tracking (RAPT)". In Speech Coding and Synthesis, W. B. Kleijn and K. K. Paliwal, Eds., Elsevier Science, pp. 497-518 (1995)
9. L. Rabiner. "On the use of autocorrelation analysis for pitch detection". IEEE Trans. Acoust., Speech, Signal Processing, 25(1): 24-33, 1977
10. D. A. Krubsack and R. J. Niederjohn. "An autocorrelation pitch detector and voicing decision with confidence measures developed for noise-corrupted speech". IEEE Trans. Acoust., Speech, Signal Processing, 39(1): 319-329, 1991
11. A. Cheveigné. "YIN, a fundamental frequency estimator for speech and music". Journal of the Acoustical Society of America, 111(4):1917-1930, 2002
12. P. Boersma. "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound". In Proceedings of the Institute of Phonetic Sciences, Amsterdam, 1993
13. A. M. Noll. "Cepstrum pitch determination". J. Acoust. SOC. Amer., 41: 293-309, 1967

14. T. Shimamura and H. Takagi. "Noise-Robust Fundamental Frequency Extraction Method Based on Exponentiated Band-Limited Amplitude Spectrum". In The 47th IEEE International Midwest Symposium on Circuits and Systems, 2004
15. A. Bouzid and N. Ellouze. "Electroglottographic measures based on GCI and GOI detection using multiscale product", International journal of computers, communications and control, 3(1): 21-32, 2008
16. A. Bouzid and N. Ellouze. "Open Quotient Measurements Based on Multiscale Product of Speech Signal Wavelet Transform", Research Letter in Signal Processing, 7: 1687-6911, 2008
17. C. S. Burrus, R. A. Gopinath and H. Guo. "Introduction to Wavelets and Wavelet Transform", A Primer. Prentice Hall, (1998)
18. S. Mallat. "A Wavelet Tour of Signal Processing", Academic Press, second edition, (1999)
19. Z. Berman and J. S. Baras. "Properties of the multiscale maxima and zero-crossings representations", IEEE Trans.on Signal Processing, 42(1):3216-3231, 1993
20. S. Kadambe and G. Faye Boudreaux-Bartels. "Application of the Wavelet Transform for Pitch Detection of Speech Signals". IEEE Trans. on Info. Theory, 38: 917-924, 1992
21. B. M. Sadler and A. Swami. "Analysis of multi-scale products for step detection and estimation". IEEE Trans. Inform. Theory, 1043-1051, 1999
22. B. M. Sadler, T. Pham and L. C. Sadler. "Optimal and wavelet-based shock wave detection and estimation". Journal of the Acoustical Society of America, 104: 955-963, 1998
23. G. Meyer, F. Plante and W. A. Ainsworth. "A pitch extraction reference database". EUROSPEECH,1995
24. F. Sha and L. K. Saul. "Real-time pitch determination of one or more voices by nonnegative matrix factorization", L. K. Saul, Y. Weiss, and L. Bottou, Eds., MIT Press, pp. 1233-1240 (2005)
25. F. Sha, J. A. Burgoyne and L. K. Saul. "Multiband statistical learning for F0 estimation in speech". In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Montreal, Canada, 2004
26. K. Achan, S. Roweis, A. Hertzmann and B. Frey. "A segment-based probabilistic generative model of speech". In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2005
27. L. R. Rabiner, M. J. Cheng, A. H. Rosenberg and C. A. McGonegal. "A comparative performance study of several pitch detection algorithms". IEEE Trans. Acoust., Speech, Signal Processing, 24(5): 399-417, 1976

Face Recognition using Neural Networks

P.Latha

*Selection .grade Lecturer,
Department of Electrical and Electronics Engineering,
Government College of Engineering,
Tirunelveli- 627007*

plathamuthuraj@gmail.com

Dr.L.Ganesan

*Assistant Professor,
Head of Computer Science & Engineering department,
Alagappa Chettiar College of Engineering & Technology,
Karaikudi- 630004*

Dr.S.Annadurai

*Additional Director, Directorate of Technical Education
Chennai-600025*

Abstract

Face recognition is one of biometric methods, to identify given face image using main features of face. In this paper, a neural based algorithm is presented, to detect frontal views of faces. The dimensionality of face image is reduced by the Principal component analysis (PCA) and the recognition is done by the Back propagation Neural Network (BPNN). Here 200 face images from Yale database is taken and some performance metrics like Acceptance ratio and Execution time are calculated. Neural based Face recognition is robust and has better performance of more than 90 % acceptance ratio.

Key words: Face recognition-Principal Component Analysis- Back Propagation Neural Network - Acceptance ratio-Execution time

1. INTRODUCTION

A face recognition system [6] is a computer vision and it automatically identifies a human face from database images. The face recognition problem is challenging as it needs to account for all possible appearance variation caused by change in illumination, facial features, occlusions, etc. This paper gives a Neural and PCA based algorithm for efficient and robust face recognition. Holistic approach, feature-based approach and hybrid approach are some of the approaches for face recognition. Here, a holistic approach is used in which the whole face region is taken into account as input data. This is based on principal component-analysis (PCA) technique, which is used to simplify a dataset into lower dimension while retaining the characteristics of dataset.

Pre-processing, Principal component analysis and Back Propagation Neural Algorithm are the major implementations of this paper. Pre-processing is done for two purposes

- (i) To reduce noise and possible convolute effects of interfering system,
- (ii) To transform the image into a different space where classification may prove easier by exploitation of certain features.

PCA is a common statistical technique for finding the patterns in high dimensional data's [1]. Feature extraction, also called Dimensionality Reduction, is done by PCA for a three main purposes like

- i) To reduce dimension of the data to more tractable limits

- ii) To capture salient class-specific features of the data,
- iii) To eliminate redundancy.

Here recognition is performed by both PCA and Back propagation Neural Networks [3]. BPNN mathematically models the behavior of the feature vectors by appropriate descriptions and then exploits the statistical behavior of the feature vectors to define decision regions corresponding to different classes. Any new pattern can be classified depending on which decision region it would be falling in. All these processes are implemented for Face Recognition, based on the basic block diagram as shown in fig 1.

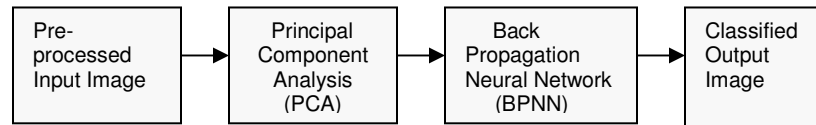


Fig. 1 Basic Block Diagram

The Algorithm for Face recognition using neural classifier is as follows:

- a) Pre-processing stage –Images are made zero-mean and unit-variance.
- b) Dimensionality Reduction stage: PCA - Input data is reduced to a lower dimension to facilitate classification.
- c) Classification stage - The reduced vectors from PCA are applied to train BPNN classifier to obtain the recognized image.

In this paper, Section 2 describes about Principal component analysis, Section 3 explains about Back Propagation Neural Networks, Section 4 demonstrates experimentation and results and subsequent chapters give conclusion and future development.

2. PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) [2] involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. PCA is a popular technique, to derive a set of features for both face recognition.

Any particular face can be

- (i) Economically represented along the eigen pictures coordinate space, and
- (ii) Approximately reconstructed using a small collection of Eigen pictures

To do this, a face image is projected to several face templates called eigenfaces which can be considered as a set of features that characterize the variation between face images. Once a set of eigenfaces is computed, a face image can be approximately reconstructed using a weighted combination of the eigenfaces. The projection weights form a feature vector for face representation and recognition. When a new test image is given, the weights are computed by projecting the image onto the eigen- face vectors. The classification is then carried out by comparing the distances between the weight vectors of the test image and the images from the database. Conversely, using all of the eigenfaces extracted from the original images, one can reconstruct the original image from the eigenfaces so that it matches the original image exactly.

2.1 PCA Algorithm

The algorithm used for principal component analysis is as follows.

- (i) Acquire an initial set of M face images (the training set) & Calculate the eigen-faces from the training set, keeping only M' eigenfaces that correspond to the highest eigenvalue.
- (ii) Calculate the corresponding distribution in M' -dimensional weight space for each known individual, and calculate a set of weights based on the input image
- (iii) Classify the weight pattern as either a known person or as unknown, according to its distance to the closest weight vector of a known person.

Let the training set of images be $\Gamma_1, \Gamma_2, \dots, \Gamma_M$. The average face of the set is defined by

$$\Psi = \frac{1}{M} \sum_{n=1}^M \Gamma_n \tag{1}$$

Each face differs from the average by vector

$$\Phi_i = \Gamma_i - \Psi \tag{2}$$

The co- variance matrix is formed by

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \cdot \Phi_n^T = A \cdot A^T \tag{3}$$

where the matrix $A = [\Phi_1, \Phi_2, \dots, \Phi_M]$.

This set of large vectors is then subject to principal component analysis, which seeks a set of M orthonormal vectors $u_1 \dots u_M$. To obtain a weight vector Ω of contributions of individual eigen-faces to a facial image Γ , the face image is transformed into its eigen-face components projected onto the face space by a simple operation

$$\omega_k = u_k^T (\Gamma - \Psi) \tag{4}$$

For $k=1, \dots, M'$, where $M' \leq M$ is the number of eigen-faces used for the recognition. The weights form vector $\Omega = [\omega_1, \omega_2, \dots, \omega_{M'}]$ that describes the contribution of each Eigen-face in representing the face image Γ , treating the eigen-faces as a basis set for face images. The simplest method for determining which face provides the best description of an unknown input facial image is to find the image k that minimizes the Euclidean distance ϵ_k .

$$\epsilon_k = \|(\Omega - \Omega_k)\|^2 \tag{5}$$

where Ω_k is a weight vector describing the k^{th} face from the training set. A face is classified as belonging to person k when the ' ϵ_k ' is below some chosen threshold Θ_ϵ otherwise, the face is classified as unknown.

The algorithm functions by projecting face images onto a feature space that spans the significant variations among known face images. The projection operation characterizes an individual face by a weighted sum of eigenfaces features, so to recognize a particular face, it is necessary only to compare these weights to those of known individuals. The input image is matched to the subject from the training set whose feature vector is the closest within acceptable thresholds.

Eigen faces have advantages over the other techniques available, such as speed and efficiency. For the system to work well in PCA, the faces must be seen from a frontal view under similar lighting.

3. NEURAL NETWORKS AND BACK PROPAGATION ALGORITHM

A successful face recognition methodology depends heavily on the particular choice of the features used by the pattern classifier. The Back-Propagation is the best known and widely used learning algorithm in training multilayer perceptrons (MLP) [5]. The MLP refer to the network consisting of a set of sensory units (source nodes) that constitute the input layer, one or more hidden layers of computation nodes, and an output layer of computation nodes. The input signal propagates through the network in a forward direction, from left to right and on a layer-by-layer basis.

Back propagation is a multi-layer feed forward, supervised learning network based on gradient descent learning rule. This BPNN provides a computationally efficient method for changing the weights in feed forward network, with differentiable activation function units, to learn a training set

of input-output data. Being a gradient descent method it minimizes the total squared error of the output computed by the net. The aim is to train the network to achieve a balance between the ability to respond correctly to the input patterns that are used for training and the ability to provide good response to the input that are similar.

3.1 Back Propagation Neural Networks Algorithm

A typical back propagation network [4] with Multi-layer, feed-forward supervised learning is as shown in the figure. 2. Here learning process in Back propagation requires pairs of input and target vectors. The output vector 'o' is compared with target vector 't'. In case of difference of 'o' and 't' vectors, the weights are adjusted to minimize the difference. Initially random weights and thresholds are assigned to the network. These weights are updated every iteration in order to minimize the mean square error between the output vector and the target vector.

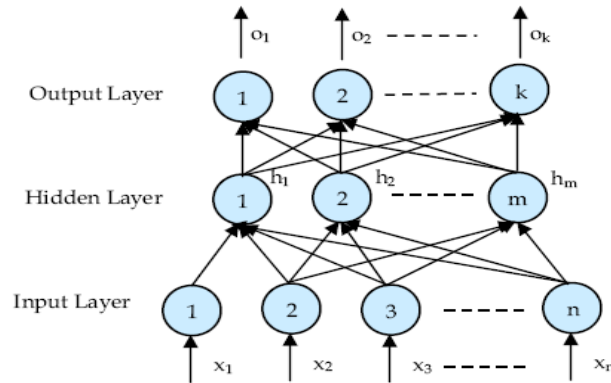


Fig. 2 Basic Block of Back propagation neural network

Input for hidden layer is given by

$$net_m = \sum_{z=1}^n x_z w_{mz} \quad \text{----- (6)}$$

The units of output vector of hidden layer after passing through the activation function are given by

$$h_m = \frac{1}{1 + \exp(-net_m)} \quad \text{----- (7)}$$

In same manner, input for output layer is given by

$$net_k = \sum_{z=1}^m h_z w_{kz} \quad \text{----- (8)}$$

and the units of output vector of output layer are given by

$$o_k = \frac{1}{1 + \exp(-net_k)} \quad \text{----- (9)}$$

For updating the weights, we need to calculate the error. This can be done by

$$E = \frac{1}{2} \sum_{i=1}^k (o_i - t_i)^2 \quad \text{----- (10)}$$

o_i and t_i represents the real output and target output at neuron i in the output layer respectively. If the error is minimum than a predefined limit, training process will stop; otherwise weights need to be updated. For weights between hidden layer and output layer, the change in weights is given by

$$\Delta w_{ij} = \alpha \delta_i h_j \quad \text{----- (11)}$$

where α is a training rate coefficient that is restricted to the range [0.01,1.0], h_{ajj} is the output of neuron j in the hidden layer, and δ_j can be obtained by

$$\delta_i = (t_i - o_i) o_i (l - o_i) \quad \text{----- (12)}$$

Similarly, the change of the weights between hidden layer and output layer, is given by

$$\Delta w_{ij} = \beta \delta_{Hi} x_j \quad \text{----- (13)}$$

where β is a training rate coefficient that is restricted to the range [0.01,1.0], x_j is the output of neuron j in the input layer, and δ_{Hi} can be obtained by

$$\delta_{Hi} = x_i (l - x_i) \sum_{j=1}^k \delta_j w_{ij} \quad \text{----- (14)}$$

x_i is the output at neuron i in the input layer, and summation term represents the weighted sum of all δ_j values corresponding to neurons in output layer that obtained in equation. After calculating the weight change in all layers, the weights can simply updated by

$$w_{ij} (new) = w_{ij} (old) + \Delta w_{ij} \quad \text{----- (15)}$$

This process is repeated, until the error reaches a minimum value

2.4.3 Selection of Training Parameters

For the efficient operation of the back propagation network it is necessary for the appropriate selection of the parameters used for training.

Initial Weights

This initial weight will influence whether the net reaches a global or local minima of the error and if so how rapidly it converges. To get the best result the initial weights are set to random numbers between -1 and 1.

Training a Net

The motivation for applying back propagation net is to achieve a balance between memorization and generalization; it is not necessarily advantageous to continue training until the error reaches a minimum value. The weight adjustments are based on the training patterns. As long as error the for validation decreases training continues. Whenever the error begins to increase, the net is starting to memorize the training patterns. At this point training is terminated.

Number of Hidden Units

If the activation function can vary with the function, then it can be seen that a n-input, m-output function requires at most $2n+1$ hidden units. If more number of hidden layers are present, then the calculation for the δ 's are repeated for each additional hidden layer present, summing all the δ 's for units present in the previous layer that is fed into the current layer for which δ is being calculated.

Learning rate

In BPN, the weight change is in a direction that is a combination of current gradient and the previous gradient. A small learning rate is used to avoid major disruption of the direction of learning when very unusual pair of training patterns is presented.

Various parameters assumed for this algorithm are as follows.

No.of Input unit	=	1 feature matrix
Accuracy	=	0.001
learning rate	=	0.4
No.of epochs	=	400
No. of hidden neurons	=	70
No.of output unit	=	1

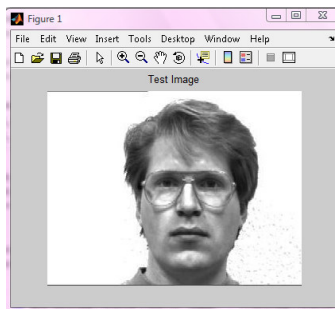
Main advantage of this back propagation algorithm is that it can identify the given image as a face image or non face image and then recognizes the given input image .Thus the back propagation neural network classifies the input image as recognized image.

4. Experimentation and Results

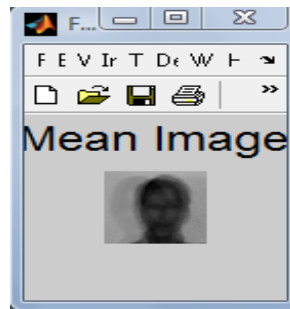
In this paper for experimentation, 200 images from Yale database are taken and a sample of 20 face images is as shown in fig 3. One of the images as shown in fig 4a is taken as the Input image. The mean image and reconstructed output image by PCA, is as shown in fig 4b and 4c. In BPNN, a training set of 50 images is as shown in fig 5a and the Eigen faces and recognized output image are as shown in fig 5b and 5c.



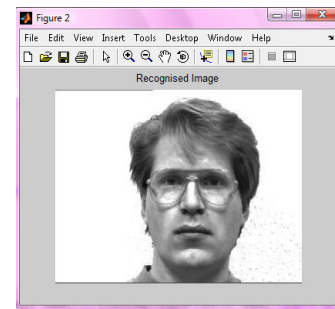
Fig 3. Sample Yale Database Images



4(a)

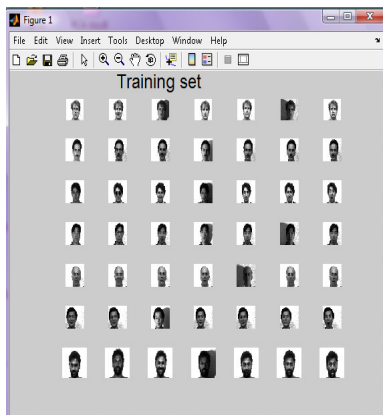


4(b)

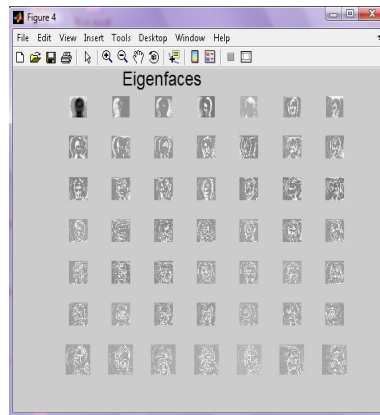


4 (c)

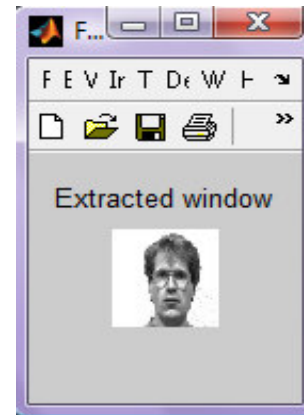
Fig 4.(a) Input Image , (b)Mean Image , (c) Recognized Image by PCA method



5(a)



5(b)



5(c)

Fig 5 (a) Training set, (b) Eigen faces , (c) Recognized Image by BPNN method

Table 1 shows the comparison of acceptance ratio and execution time values for 40, 80, 120,160 and 200 images of Yale database. Graphical analysis of the same is as shown in fig 6.

No .of	Acceptance ratio (%)	Execution Time (Seconds)
--------	----------------------	--------------------------

Images				
	PCA	PCA with BPNN	PCA	PCA with BPNN
40	92.4	96.5	38	36
60	90.6	94.3	46	43
120	87.9	92.8	55	50
160	85.7	90.2	67	58
200	83.5	87.1	74	67

Table 1 Comparison of acceptance ratio and execution time for Yale database images

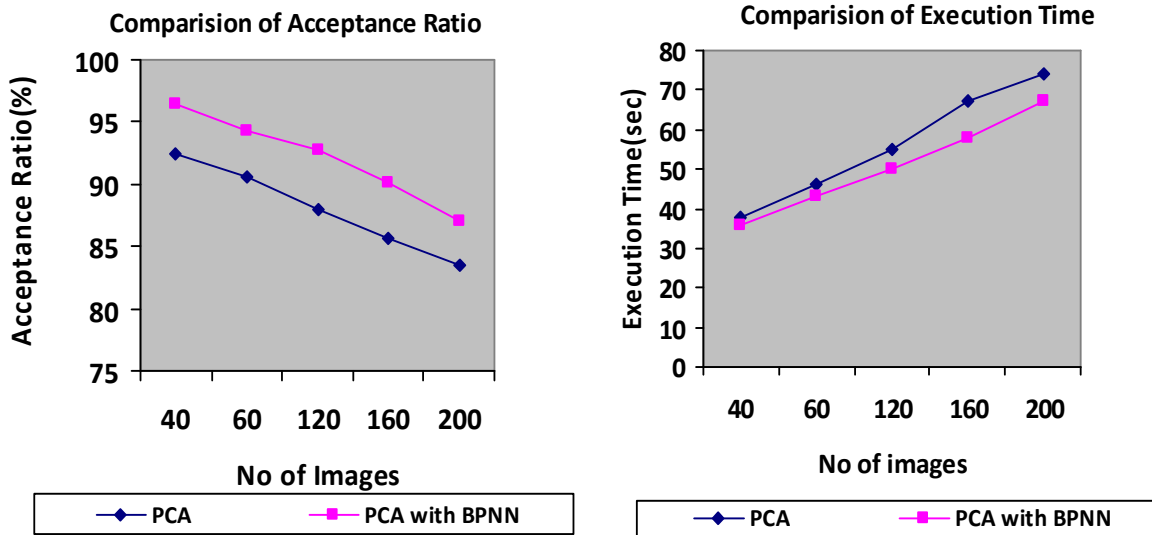


Fig.6: comparison of Acceptance ratio and execution time

5. CONCLUSION

Face recognition has received substantial attention from researches in biometrics, pattern recognition field and computer vision communities. In this paper, Face recognition using Eigen faces has been shown to be accurate and fast. When BPNN technique is combined with PCA, non linear face images can be recognized easily. Hence it is concluded that this method has the acceptance ratio is more than 90 % and execution time of only few seconds. Face recognition can be applied in Security measure at Air ports, Passport verification, Criminals list verification in police department, Visa processing , Verification of Electoral identification and Card Security measure at ATM's..

6. REFERENCES

- [1]. B.K.Gunturk,A.U.Batur, and Y.Altunbasak,(2003) "Eigenface-domain super-resolution for face recognition," IEEE Transactions of . Image Processing. vol.12, no.5.pp. 597-606.

- [2]. M.A.Turk and A.P.Petland, (1991) "Eigenfaces for Recognition," Journal of Cognitive Neuroscience. vol. 3, pp.71-86.
- [3]. T.Yahagi and H.Takano,(1994) "Face Recognition using neural networks with multiple combinations of categories," International Journal of Electronics Information and Communication Engineering., vol.J77-D-II, no.11, pp.2151-2159.
- [4]. S.Lawrence, C.L.Giles, A.C.Tsoi, and A.d.Back, (1993) "IEEE Transactions of Neural Networks. vol.8, no.1, pp.98-113.
- [5]. C.M.Bishop,(1995) "Neural Networks for Pattern Recognition" London, U.K.:Oxford University Press.
- [6]. Kailash J. Karande Sanjay N. Talbar "Independent Component Analysis of Edge Information for Face Recognition" International Journal of Image Processing Volume (3) : Issue (3) pp: 120 -131.

Arabic Phoneme Recognition using Hierarchical Neural Fuzzy Petri Net and LPC Feature Extraction

Ghassaq S. Mosa

*College of Engineering/Department of
Computer Engineering
University of Basrah
Basrah, Iraq.*

ghassaqsaeed@yahoo.com

Abduladhem Abdulkareem Ali

*College of Engineering/Department of
Computer Engineering
University of Basrah
Basrah, Iraq.*

aduladhem@compengbas.net

Abstract

The basic idea behind the proposed hierarchical phoneme recognition is that phonemes can be classified into specific phoneme types which can be organized within a hierarchical tree structure. The recognition principle is based on “divide and conquer” in which a large problem is divided into many smaller, easier to solve problems whose solutions can be combined to yield a solution to the complex problem. Fuzzy Petri net (FPN) is a powerful modeling tool for fuzzy production rules based knowledge systems. For building hierarchical classifier using Neural Fuzzy Petri net (NFPN), Each node of the hierarchical tree is represented by a NFPN. Every NFPN in the hierarchical tree is trained by repeatedly presenting a set of input patterns along with the class to which each particular pattern belongs. The feature vector used as input to the NFPN is the LPC parameters.

Keywords: Hierarchical networks, Linear predictive coding, Neural fuzzy Petri net, phoneme recognition, Speech recognition.

1. INTRODUCTION

The Arabic Language is one of the oldest living languages in the world. The bulk of classical Islamic literature was written in classical Arabic (CA), and the Holy Qur'an was revealed in the Classical Arabic language. Standard Arabic is the mother (spoken) tongue for more than 200 million people living in the vast geographical area known as the Arab world, which includes countries such as Iraq, Syria, Jordan, Egypt, Saudi Arabia, Morocco, and Sudan. Arabic is one of the world's oldest Semitic languages, and it is the fifth most widely used. Arabic is the language of communication in official discourse, teaching, religious activities, and in literature.

Many works have been done on the recognition of Arabic phonemes. These studies include the use of neural networks [1-3], Hidden Markov Model [4] and Fuzzy system [5].

Hierarchical approaches based on neural networks were employed in other languages with different techniques [5-7]. In this paper hierarchical Arabic phoneme recognition system is proposed based on LPC feature vector and neural Fuzzy Petri Net (NFPN). The principle is based on proposing a decision tree for the Arabic phonemes. NFPN is used as a decision network in each

node in the tree.

2. ARABIC LANGUAGE ALPHABET

Every language is typically partitioned into two broad categories: vowels and consonants. Vowels are produced without obstructing air flow through the vocal tract, while consonants involve significant obstruction, creating a noisier sound with weaker amplitude. The Arabic language consists of 28 letters, Arabic is written from right to left and letters take different forms depending on their position in a word; some letters are similar to others except for diacritical points placed above or beneath them. Arab linguists classify Arabic letters into two categories: sun and moon. Sun letters are indicated by an asterisk. When the sun letters are preceded by the prefix Alif-Laam in nouns, the Laam consonant is not pronounced. The Arabic language has six different vowels, three short and three long. The short vowels are fatha (a), short kasrah (i), and short dammah (u). No special letters are assigned to the short vowels; however special marks and diacritical notations above and beneath the consonants are used. The three long vowels are durational allophones of the above short vowels, as in mad, meet, and soon and correspond to long fatha, long kasrah, and long dammah respectively. Consonants can be also un-vowelised (not followed by a vowel); in this case a diacritic sakoon is placed above the Consonant. Vowels and their IPA (International Phonetic Alphabet) equivalents .

3. SPEECH RECOGNITION SYSTEM

The general model for speech recognition system here are five major phases recording & digitalizing speech signal, segmentation, pre-processing signal, feature extraction and decision-making, each phase will be explained in more details along with the approaches used to enhance the performance of the speech recognition systems.

A. A/D Conversion: The input speech signal is changed into an electrical signal by using a microphone. Before performing A/D conversion, a low pass filter is used to eliminate the aliasing effect during sampling. A continuous speech signal has a maximum frequency component at about 16 KHz.

B. Segmentation: Speech segmentation plays an important role in speech recognition in reducing the requirement for large memory and in minimizing the computation complexity in large vocabulary continuous speech recognition systems. [8].

C. Preprocessing: Preprocessing includes filtering and scaling of the incoming signal in order to reduce the noise and other external effect. Filtering speech signal before recognition task is an important process to remove noise related to speech signal which may be either low frequency or high frequency noise. Figure (1) shows the effect of preprocessing on signal.

D. Feature Extraction: The goal of feature extraction is to represent any speech signal by a finite number of measures (or features) of the signal. This is because the entirety of the information in the acoustic signal is too much to process, and not all of the information is relevant for specific tasks. In present ASR systems, the approach of feature extraction has generally been to find a representation that is relatively stable for different examples of the same speech sound, despite differences in the speaker or environmental characteristics, while keeping the part that represents the message in the speech signal relatively intact [9].

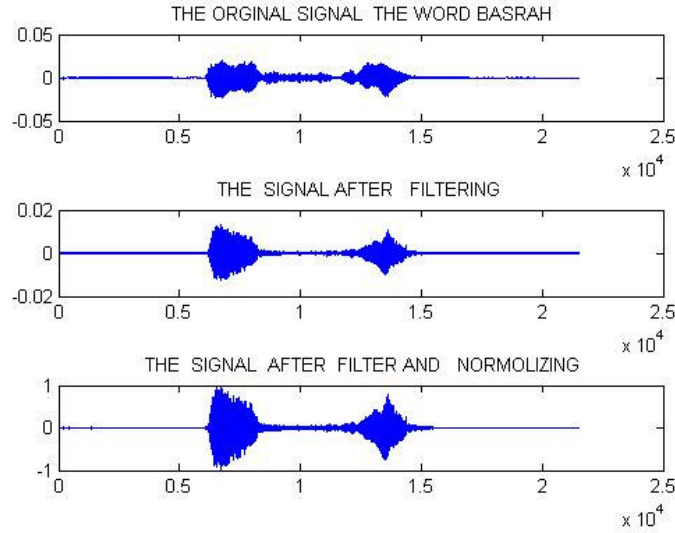


FIGURE 1: The word Basrah before and after filtering and normalizing.

4. LINEAR PRIDICTIVE CODING

Linear predictive analysis has been one of the most powerful speech analysis techniques since it was introduced in the early 1970s [10]. The LPC is a mode1 based on the vocal tract of human beings [11]. Figure (2) shows the block diagram of the LPC calculations.

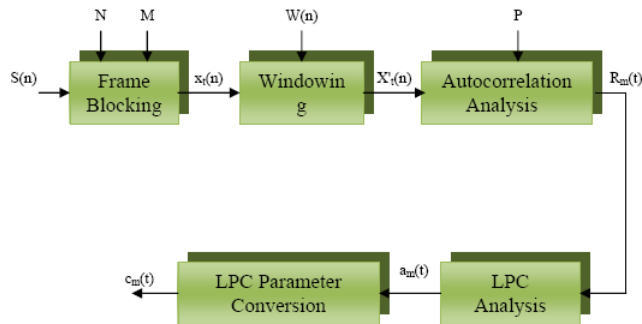


FIGURE 2: The LPC block diagram

A. **Frame Blocking:** The digitalized speech signal, $S(n)$, is blocked into frames of N samples, with adjacent frames being separated by M samples. If we denote the l th frame of speech by $x_l(n)$, and there are L frames within the entire speech signal, then

$$x_l(n) = s(M_l + n), \quad n = 0, 1, \dots, N - 1, l = 0, 1, \dots, L - 1$$

B. **Windowing:** To minimize the discontinuity and therefore preventing spectral leakage of a signal at the beginning and end of each frame, every frame is multiplied by a window function.

$$w(n) = 0.54 - 0.45 \cos\left(\frac{2\pi n}{N-1}\right) \tag{1}$$

$$0 \leq n \leq N-1$$

Figure (3) shows the effect of windowing.

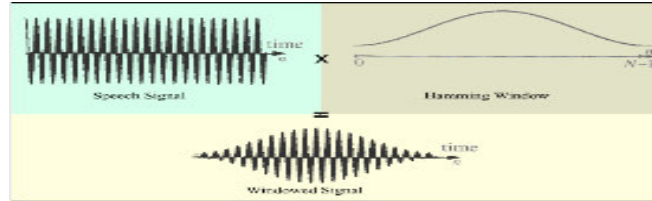


FIGURE 3: Effect of window on signal.

C. Autocorrelation Analysis: In this step, each frame of Windowed signal is auto correlated to give:

$$r_l = \sum_{n=0}^{N-1-m} x_l'(n)x_l'(n+m) \tag{2}$$

$m = 0, 1, 2, \dots, p$

Where $x_l'(n)$ is the windowed signal, Where highest correlation value, p , is the order of LPC analysis. Typically, values of p from 8 to 16 are used. It is interesting to note that the zeroth autocorrelation, $r_0(0)$, is the energy of l^{th} frame. The frame energy is an important parameter for speech-detection. In our case [10,11].

$$r_{xx}(l) = \sum_{n=-\infty}^{\infty} x(n)x(n-l) = \sum_{n=-\infty}^{\infty} x(n+l)x(n) \tag{3}$$

For $l=0, 1, 2, 3, \dots$

D. LPC Analysis: The next processing step is the LPC calculation, which convert each of the auto correlated frame into an "LPC parameter set", in which the set might be the LPC coefficients, the reflection (or PARCOR) coefficients, the log are ratio coefficients and the cepstral coefficients. The formal method for converting from autocorrelation coefficients to an LPC parameter set (for the LPC autocorrelation method) is known as Durbin's method and can formally given as the following algorithm for convenience omitting the subscript l or $r_l(m)$, [10].

$$E^0 = r(0) \tag{4}$$

$$K_i = \left\{ r(i) - \sum_{j=1}^{i-1} a^{(i-1)}_j r(i-j) \right\} / E^{(i-1)} \tag{5}$$

$$1 \leq i \leq p \tag{6}$$

$$a^{(i)}_i = K_i \tag{6}$$

$$a^{(i)}_j = a^{(i-1)}_j - K_i a^{(i-1)}_{i-j} \tag{7}$$

$$E^{(i)} = (1 - K_i^2) E^{(i-1)} \tag{8}$$

Where, the summation in Equation (5) is omitted for $i=1$, the set of Equation(4 -8) are solved recursively for $i=1, 2, \dots, P$, and the final solution is given as [11] :

$$a_m = \text{LPC coefficients} = a^{(P)}_m ,$$

$$1 \leq m \leq p \tag{9}$$

$$K_m = \text{PARCOR Coefficients} \tag{10}$$

$g_m = \text{Log area ratio coefficients}$

$$= \log \left[\frac{1 - K_m}{1 + K_m} \right] \tag{11}$$

D. LPC Parameter Conversion to Cepstral Coefficients: A very important LPC parameter set, which can be driven from the LPC coefficient set, is the LPC cepstral coefficients C_m , that is calculated by the following Equation:

$$C_m = a_m + \sum_{k=1}^{m-1} \left[\frac{k}{m} \right] c_k a_{m-k}$$

$$1 \leq m \leq p$$

Where a_m is LPC coefficients

5. FUZZY NEURAL PETRI NET

After extracting the desired features from the input data, they are applied to the decision making stage in order to make the appropriate decision on the specific class that the input data belongs to. In this work NFPN is used as the decision making network.

Petri nets, developed by Carl Adam Petri in his Ph.D. thesis in 1962, are generally considered as a tool for studying and modeling of systems. A Petri net (PN) is foremostly a mathematical description, but it is also a visual or graphical representation of a system. The application areas of Petri nets being vigorously investigated involve knowledge representation and discovery, robotics, process control, diagnostics, grid computation, traffic control, to name a few high representative domains [12]. Petri nets (PNs) are a mathematical tool to describe concurrent systems and model their behavior. They have been used to study the performance and dependability of a variety of systems.

Petri Nets essentially consist of three key components: places, transitions, and directed arcs [13]. The directed arcs connect the places to the transitions and the transitions to the places. There are no arcs connect transitions to transitions or places to places directly. Each place contains zero or more tokens. A vector representation of the number of tokens over all places defines the state of the Petri Net. A simple Petri Net graph is shown in Figure (4). The configuration of the Petri Net combined with the location of tokens in the net at any particular time is called the Petri Net s

Petri Net structure is formally described by the five-tuple (P, T, I, O, M),

Where P is the set of places {p1, ..., pn},

T is the set of transitions {t1, ..., tm},

I is the set of places connected via arcs as inputs to transitions,

O is the set of places connected via arcs as outputs from transitions,

and M is the set of places that contain tokens

et of transitions {t1, ..., tm}, Formal

Definition and State of Figure (4)

(P, T, I, O, M):

P = { p1, p2, p3, p4 }

T = { t1, t2 }

I = { { p1 }, { p2, p3 } }

O = { { p2, p3 }, { p4 } }

M = { 1, 0, 0, 0 }

The structure of the proposed Neural Fuzzy Petri Net is shown in figure (5) and (6). The network has the following three layers

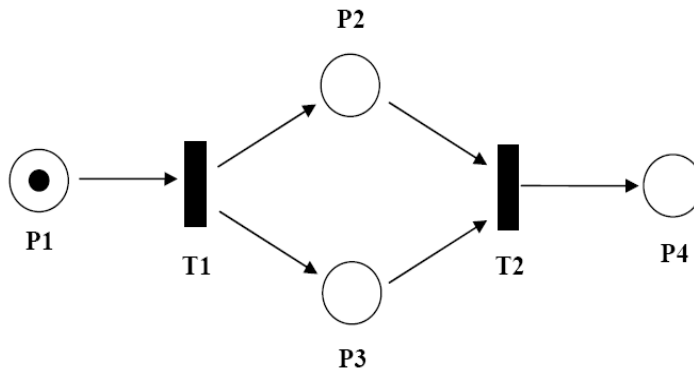


FIGURE 4: Petri Net graph.

:

- an input layer composed of n input
- a transition layer composed of hidden transitions;
- an output layer consisting of m output places.

The input place is marked by the value of the feature. The transitions act as processing units. The firing depends on the parameters of transitions, which are the thresholds, and the parameters of the arcs (connections), which are the weights. Each output place corresponds to a class of pattern. The marking of the output place reflects a level of membership of the pattern in the corresponding class [14].

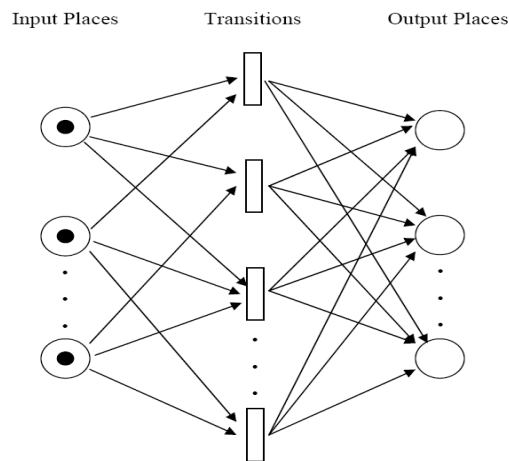


FIGURE 5: The structure of the Neural Fuzzy Petri Net.

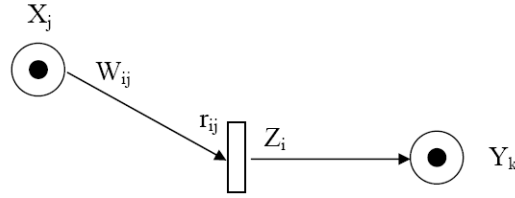


FIGURE 6: Section of the net outlines the notations.

The specifications of the network are as follows:

- X_j is the marking level of j -th input place produced by a triangular mapping function. The top of the triangular function is centered on the average point of the input values. The length of triangular base is calculated from the difference between the minimum and maximum values of the input. The height of the triangle is unity. This process keep the input of the network within the period $[0,1]$. This generalization of the Petri net will be in full agreement with the two-valued generic version of the Petri net [14].

$$x_j = f(input(j)) \tag{13}$$

Where f is a triangular mapping function

$$f(x) = \begin{cases} \frac{x - \min(x)}{\text{average}(x) - \min(x)} & , \text{ if } x < \text{average}(x) \\ \frac{\max(x) - x}{\max(x) - \text{average}(x)} & , \text{ if } x > \text{average}(x) \\ 1 & , \text{ if } x = \text{average}(x) \end{cases}$$

- W_{ij} is the weight between the i -th transition and the j -th input place;
- r_{ij} is a threshold level associated with the level of marking of the j -th input place and the i -th transition;
- Z_i is the activation level of i -th transition and defined as follows:

$$Z_i = \prod_{j=1}^n [W_{ij} S(r_{ij} \rightarrow X_j)] \tag{14}$$

, $j = 1, 2, \dots, n; i = 1, 2, \dots, \text{hidden}$

Y_k is the marking level of the k -th output place produced by the transition layer and performs a nonlinear mapping of the weighted sum of the activation levels of these transitions (Z_i) and the associated connections V_{jk}

$$Y_k = f\left(\sum_{i=1}^{\text{No.ofTransitions}} V_{ki} Z_i\right), \quad j = 1, 2, \dots, m \tag{15}$$

Where “ f ” is a nonlinear monotonically increasing function from R to $[0,1]$.

Learning Procedure

The learning process depends on minimizing certain performance index in order to optimize the network parameters (weights and thresholds). The performance index used is the standard sum of squared errors[4].

$$E = \frac{1}{2} \sum_{k=1}^m (t_k - y_k)^2 \tag{16}$$

Where t_k is the k-th target;
 y_k is the k-th output. The updates of the parameters are performed according to the gradient method

$$\text{param}(\text{iter} + 1) = \text{param}(\text{iter}) - \alpha \nabla_{\text{param}} E \tag{17}$$

Where $\nabla_{\text{param}} E$ is a gradient of the performance index E with respect to the network parameters, α is the learning rate coefficient, and iter is the iteration counter.

The nonlinear function associated with the output place is a standard sigmoid described as [14]:

$$y_k = \frac{1}{1 + \exp(-\sum Z_i V_{ki})} \tag{18}$$

In this paper we used the fuzzy Petri net in Arabic phoneme recognition and using LPC (linear predictive code) technique as feature extracting for speech signal.

6. EXPERIMENTAL RESULTS

For each phoneme, there are 24 recorded words. In eight of the 24 words, the target comes as the initial letter in word. In the second group, 8 phonemes come in the middle of the words. In other eight phonemes, the target phoneme comes at the end of the words. Manual segmentation is used to extract target phoneme from the recorded words. Half of the phonemes used in the training and the other half is used for testing the resulting system. Adobe Audition software 1.5 is used to record and save the data files. The recorded data is stored as (.WAV) files with 16-bit per sample precision and a sampling rate of 16 kHz.

A hierarchical tree is formed for the phonemes as shown in figure (7). Five classes exist in this tree. The first class is the fricative which contain both voiced and unvoiced, the second class is the stop which contain both voiced and unvoiced, the third class contain the semi vowel phoneme, the fourth class contain nasal, and fifth class contain affricative, lateral, and trail. Each node in the tree is recognized using a separate NFPN with LPC parameters as inputs to these networks. The recognition principle is based on divide and conquer. Class 1 is identified with a net consist of 18 input place, 56 transition and one output place. For class 2 is a net consist of 18 input place, 47 transitions and one output place. For class 3 the input place is 18, the hidden layer is 22 and one output place. For class 4 the input place is 2, the hidden layer is 22 and one output place. For class 5 is a net consist of 18 input place, 36 hidden layer and one output. Table 1 shows the recognition accuracy for each phoneme and class recognition. It is found that the total recognition accuracy reached 79.6378%.

7. Conclusion

Arabic phoneme recognition system is proposed in this work. The decision is based on LPC feature vector and hierarchical NFPN as a decision network. The experimental results shows that it is possible to use the hierarchical structure to recognize phonemes using NFPN.

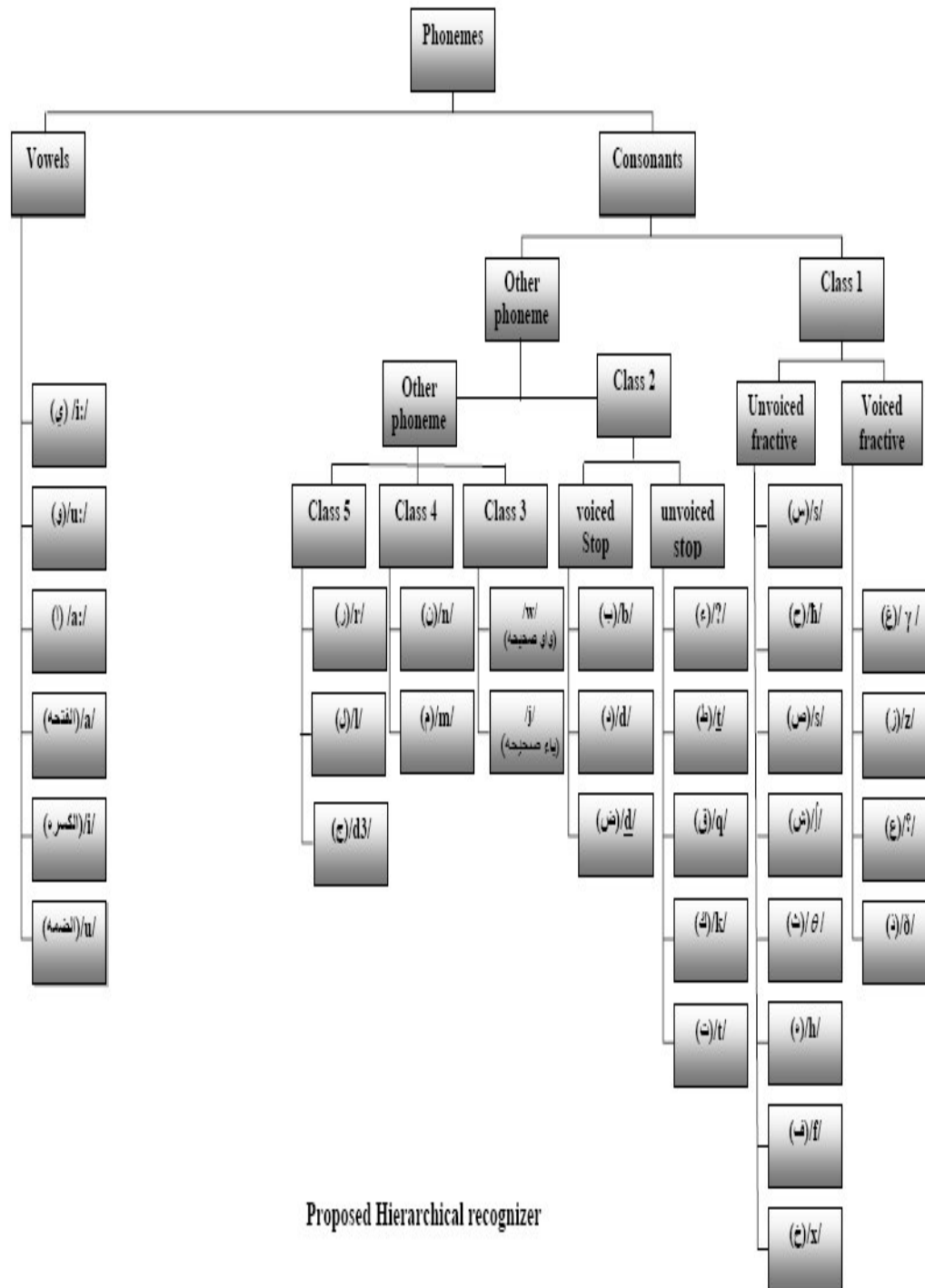


FIGURE 7: The proposed hierarchical tree.

NO	phonemes	IPA symbols	Phoneme recognition accuracy %	class recognition accuracy %
1	ء	/ʔ/	71.66	85.83
2	ب	/b/	66.67	87.5
3	ت	/t/	81.67	100
4	ث	/θ/	61.46	81.25
5	ج	/dʒ/	91.67	91.67
6	ح	/h/	87.5	100
7	خ	/x/	56	75
8	د	/d/	69.44	83.33
9	ذ	/ð/	60.41	75
10	ر	/r/	58.8	75
11	ز	/z/	86.45	90.27
12	س	/s/	91.67	100
13	ش	/ʃ/	100	100
14	ص	/s/	86.85	100
15	ض	/d/	58.33	83.33
16	ط	/t/	56.67	75
17	ع	/ʕ/	90.97	100
18	غ	/ɣ/	91.67	91.67
19	ف	/f/	75	87.5
20	ق	/q/	71.67	87.5
21	ك	/k/	90.33	91.67
22	ل	/l/	70.83	91.67
23	م	/m/	83.33	83.33
24	ن	/n/	66.67	83.33
25	ه	/h/	89.58	91.67
26	و	/w/	83.33	100
27	ي	/j/	91.667	100
28	ا عله	/a:/	81.25	87.5
29	و عله	/u:/	100	100
30	ي عله	/i:/	91.5833	100
31	فتحه	/a/	86.667	91.667
32	ضمه	/u/	86.667	91.667
33	كسره	/i/	91.5833	100
Total accuracy %			79.6378	90.3744

Recognition accuracy for the phonemes in the Hierarchical based LPC.

TABLE 1: The phonemes recognition accuracy

8. REFERENCES

1. S. Ismail, and A. Ahmad, "Recurrent neural network with back propagation through time algorithm for Arabic recognition," In Proceedings of the 18th ESM Magdeburg, Germany, 13-16 June 2004.
2. S. Al-Sayegh, and A. AbedEl-Kader, "Arabic phoneme recognizer based on neural network", In Proceedings of International Conf. Intelligent Knowledge Systems (IKS-2004), August 16-20,2004.
3. Y. Alotaibi, S. Selouani, and D. O'Shaughnessy, "Experiments on Automatic Recognition of Nonnative Arabic Speech", EURASIP Journal on Audio, Speech, and Music Processing, Vol. 2008, pp.1-6, 2008.
4. M. Awais, and Habib-ur-Rehman, "Recognition of Arabic phonemes using fuzzy rule base system", In Proceedings of 7th Int. Multi Topic Conf. INMIC-2003, pp.367-370, 8-9 Dec. 2003.
5. P. Schwarz, P. Matejka, and J. Cernocky, "Hierarchical Structures of Neural Networks for Phoneme Recognition", In Proceedings of IEEE Int. Conf. Acoustics, Speech and Signal Processing, ICSP-2006, 14-19 May 2006.
6. J. Pinto and H. Hermansky, "Combining Evidence from a Generative and a Discriminative Model in Phoneme Recognition", In Proceedings of Interspeech. Brisbane, Australia 22-26 September 2008
7. M. Scholz, and R. Vigario, " Nonlinear PCA: a new hierarchical approach", In Proceedings of European Symposium on Artificial Neural Networks ESANN-2002, pp. 439-444, Bruges, Belgium, 24-26 April 2002.
8. Y. Suh and Y. Lee, "Phoneme Segmentation of Continuous Speech using Multi-Layer perceptron", In Proceedings of 4th Int. Conf. Spoken Language, ICSLP-96,3, pp.1297-1300 , 1996.
9. Y.Gong, "Speech Recognition in Noisy Environments: A Survey", *Speech Communication* ,16, p:261-291, 1995.
10. N. Awasthy, J.P.Saini and D.S.Chauhan "Spectral Analysis of Speech: A New Technique", *Int. J. Signal Processing* ,2(1), p: 19-28, 2005.
11. L.Rabinar and R.W.Schafar "Fundamental of Speech Recognition ",Prentice Hall, 1993.
12. S. I. Ahson "Petri net models of fuzzy neural networks," *IEEE Trans. Syst. Man Cybern.*, 25(6), pp. 926–932, Jun. 1995.
13. A. Seely, "Petri Net Implementation of Neural Network Elements", M.Sc. Thesis, Nova Southeastern University, 2002.
14. H. M. Abdul-Ridha, " ECG Signal Classification using Neural, Neural Fuzzy and Neural Fuzzy Petri Networks" Ph.D. Thesis, Department of Electrical Engineering, University of Basrah, 2007.

COMPUTER SCIENCE JOURNALS SDN BHD
M-3-19, PLAZA DAMAS
SRI HARTAMAS
50480, KUALA LUMPUR
MALAYSIA