

Different Techniques and Algorithms for Biomedical Signal Processing

Sh-Hussain Salleh

*Centre for Biomedical Engineering,
Faculty of Electrical Engineering,
UTM Skudai, Johor, Malaysia
E-mail: hussain@fke.utm.my*

ABSTRACT

This paper is intended to give a broad overview of the complex area of biomedical and their use in signal processing. It contains sufficient theoretical materials to provide some understanding of the techniques involved for the researcher in the field. This paper consists of two parts: feature extraction and pattern recognition. The first part provides a basic understanding as to how the time domain signal of patient are converted to the frequency domain for analysis. The second part provides basic for understanding the theoretical and practical approaches to the development of neural network models and their implementation in modeling biological system

Keywords: Biomedical Signal Processing, LPC, IEFEE, Time Frequency, Neural Network.

INTRODUCTION

Auscultation, the act of listening to the sounds of internal organs, is a valuable medical diagnostic tool. Auscultation methods provide the information about a vast variety of internal body sounds originated from the heart, lungs, bowel and vascular disorders. The information acquired by a traditional stethoscope is, however, subjective and qualitative in nature since the differentiation of signals picked up by the sensor during manual interpretation is limited by human perception abilities and varies with personal aptitude and training. This may result in inaccurate or insufficient information due to the inability of the user to discern certain complex, low-level, short duration or rarely encountered abnormal sounds. It is, thus, desirable to enhance the diagnostic ability by processing the auscultation signals electronically and providing a visual display and automatic analysis to the physician for a better comparative study. During the last decade, efforts have been made in this direction and electronic stethoscopes are now available commercially with phono-cardiograph display. The usage of such devices, however, is less common due to the involvement of cumbersome instrumentation and complex/additional skills. Our research at CBE, UTM

encompasses a more user friendly and quantitative approach together with automatic signal analysis using intelligent algorithms.

FEATURE EXTRACTION OF BIOMEDICAL SIGNAL

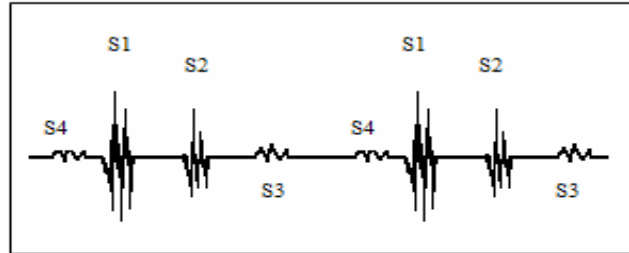


Figure 1: Location of S1, S2, S3 and S4

The characteristics of the heart sounds can be modeled as a time-varying discrete time signal

$$\begin{aligned}
 x(n) &= c(n) \cos(j\phi(n)) + w(n) \\
 &= c(n) \cos\left(2\pi \sum_{\lambda=-\infty}^n f_i(\lambda)\right) + w(n) \\
 0 \leq n \leq N - 1
 \end{aligned} \tag{1}$$

Time Frequency Analysis

Time Frequency Signal Processing is a method and technique developed for representation, analysis and processing of non-stationary signals, which time and frequency are interrelated. Time representation, $s(t)$, and frequency representation, $S(f)$, are related via the Fourier Transform.

$$s(t) = \int_{-\infty}^{\infty} S(f) e^{j2\pi ft} df \stackrel{FT}{\longleftrightarrow} S(f) = \int_{-\infty}^{\infty} S(t) e^{j2\pi ft} .dt \tag{2}$$

Time-frequency representation represent signals in time and frequency which all signals information is accessible and provides a distribution of signal energy versus t and f simultaneously which called time frequency distribution (TFD). TFD reveals the number of signal components, the time variation of frequency content of the components and order of appearance in time of the different frequencies present.

The B-Distribution is a Quadratic or bilinear Time frequency Distribution algorithm which capable to minimizing cross term [6] and it defined as:

$$B(t, f) = \iint (|\tau| / \cosh^2(u - \tau))^\beta z(u + \tau/2) z^*(u - \tau/2) e^{-j2\pi f \tau} du d\tau \quad (3)$$

Singular Value Decomposition-SVD

SVD based technique is introduced to reduce the effect of noise from TFD which deal with singular matrices, or once which are very close to being singular. They are an extension of eigen decomposition to suit non-square matrices. Any matrix may be decomposed into a set of characteristic eigenvector pairs called the component factors, and their associated eigenvalues called the singular value.

In order to extract data dynamically which decompose the X , TF distribution matrix of the power disturbance signals, which $m \times n$ (*time x frequency*) into a set of characteristic. A singular value decomposition of an $m \times n$ matrix X is any factorization of the form:

$$X = U \Sigma V^T \quad (4)$$

where U is an $m \times m$ orthogonal matrix; i.e. has orthonormal columns, V is an $n \times n$ orthogonal matrix and Σ is $m \times n$ an diagonal matrix of singular values with components $\sigma_{ij}=0$ if and $i \neq j$; (for convenience we refer to the i -th singular value $\sigma_i = \sigma_{ii}$). Furthermore it can be shown that there exist non-unique matrices U and V such that singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$. The columns of the orthogonal matrices U and V are called the left and right singular vectors respectively; an important property of U and V is that they mutually orthogonal [4].

Principle Component Analysis-PCA

PCA is generally used when the research purpose is data reduction (to reduce the information in many measured variables into a smaller set of components) and it can minimize the reconstruction error in the sense of least square error then find out the most representative feature. PCA seeks a linear combination of variables such that maximum variance is extracted from the variables. It then removes this variance and seeks a second linear combination which explains the maximum proportion of the remaining variance, and so on. This is called the principle axis method and results in orthogonal (uncorrelated) factors. PCA analyzes total (common and unique) variance. We can see that the SVD is in fact closely related to the PCA. In fact the matrix product $U \Sigma$ is analogous to the matrix Y defined for PCA as:

$$Y = XV = U\Sigma \quad (5)$$

Because both the singular vectors defined for an SVD are square and have orthonormal columns their inverses are given by their transposes. Now the relation in Equation 4 can be expressed $X=U\Sigma V^T$ which is the definition of an SVD. The pairs of eigenvectors are the row in U and the column in V [14].

Instantaneous Energy and Frequency Estimate

To estimate the instantaneous energy and frequency, the equation (1) must be in complex or discrete-analytical form. So that the real and imaginary part can be separated and calculated independently.

$$z(n) = x(n) + jH[x(n)] \quad (6)$$

Where $H[\cdot]$ is the Hilbert transform[6] of $x(n)$ which has approximately unity gain and introduce a $\pi/2$ phase shift with respect to the original signal. The Hilbert transform calculation method is

$$jX_i[n] = \frac{1}{N} \sum_{m=0}^{N-1} X_R[m] \otimes V_N[n-m] \quad (7)$$

Where $X_R[m]$ is the DFS[6] of $x(n)$ and the \otimes denote the circular convolution technique. Thus, the complex form of equation (1) is

$$z(n) = c(n) \exp \left(j2\pi \sum_{\lambda=-\infty}^n f_i(\lambda) \right) + w(n)$$

$$0 < n < N-1. \quad (8)$$

Instantaneous Energy- Another characteristic of analytical signal is the relationship between the signal amplitude and the instantaneous energy. The instantaneous energy is basically representing the temporal strength of a time varying signal. So, refer to the analytical equation (4), the strength is refer to the amplitude of the signal. Thus, using the signal definition in equation (8), the instantaneous energy is

$$E_z = z(n)z^*(n) = c(n)c^*(n). \quad (9)$$

Instantaneous Frequency- The instantaneous frequency is estimated using Central Finite Difference Frequency Estimate method (CFD-FE). Assuming high signal-to-noise ratio, then the instantaneous frequency is obtained by [7] is

$$f_i(n) = \frac{1}{4\pi} [\phi(n+1) - \phi(n-1)] \quad (10)$$

The parameters of the heart sound signal are estimated using instantaneous energy and frequency estimation method. There are basically 10 of them.

- 1) E_2/E_1 -Energy ratio of S2 to S1
- 2) E_3/E_1 -Energy ratio of S3 to S1
- 3) E_4/E_1 - Energy ratio of S4 to S1
- 4) E_{12}/E_1 - Energy ratio of S12 to S1
- 5) E_{21}/E_1 - Energy ratio of S21 to S1
- 6) F_2/F_1 - Frequency ratio of S2 to S1
- 7) F_3/F_1 - Frequency ratio of S3 to S1
- 8) F_4/F_1 - Frequency ratio of S4 to S1
- 9) F_{12}/F_1 - Frequency ratio of S12 to S1
- 10) F_{21}/F_1 - Frequency ratio of S21 to S1

Linear Predictive Coding-LPC

The heart signal is segmented into frames and LPC analysis is performed into each frame. Segmentation is to reduce the non-stationary characteristic of the heart sound signal. The Durbin's method [8] was used to calculate the LPC coefficient. Detail LPC derivation can be found in [8, 11]. LPC encodes a signal by finding a set of weights on earlier signal values that can predict the next signal value:

$$y(n) = a(1)y(n-1) + a(2)y(n-1) + a(3)y(n-3) + e(n) \quad (11)$$

If values for $a(1..3)$ can be found such that $e(n)$ is very small for a stretch of speech (say one analysis window), then we can transmit only $a(1..3)$ instead of the signal values in the window. The result of LPC analysis then is a set of coefficients $a(1..k)$ and an error signal $e(n)$, the error signal will be as small as possible and represents the difference between the predicted signal and the original.

The LPC output approximation $\hat{y}(n)$, depending only on past output samples, is

$$\hat{y}(n) = -\sum_{i=1}^p a(i)y(n-i) \quad (12)$$

p is the prediction order, $a(i)$ are the model parameters called the predictor coefficients, and $y(n-i)$ are past outputs.

CLASSIFICATION OF BIOMEDICAL SIGNAL

Neural Network

The feature of pattern to be recognized is feed into input layer and distributed to network through hidden layer and finally output layer. Each node in the network operates by function below:

$$out_i = f(net_i) = f\left(\sum_j W_{ij}out_j + \theta_i\right) \quad (13)$$

W_{ij} is a weighting vector from layer under consideration, i to preceding layer j . θ_i is a network threshold and $f(\)$ is a non-linear activation function. The activation function used is sigmoid function. The network was train by back propagation. The error measure with the mean square (MSE) error will be used. MSE calculates the distance between the desired output and the actual outputs of the networks. Learning in back propagation is by modifying the weight connection between neuron to reduce learning error.

Optimization of NN Back Propagation Algorithm

Purpose: To optimize convergence performance of a system, i.e: to overcome the problem of local minima get trapped in the valley.

Steepest Gradient Descent Method

The origin of this method is from a standard method called the steepest descent or also called basic dynamic gradient system. The direction of the descent is derived as

$$d_k = -\nabla E(x^{(k)}) \quad (14)$$

Conjugate Gradient Method

The Conjugate Gradient Method (CGM) only requires a slight modification of the discrete-time steepest descent method but often enable to improve the convergence rate dramatically. In CGM, the global search is performed along the conjugate direction; in turn this algorithm generally produces faster convergence compared to the steepest descent. The conjugate gradient algorithm starts with the search in the steepest descent direction (negative of the gradient) in the first iteration. The first search direction is denoted as P_o , where

$$P_o = -g_o$$

The simplest form of CGM algorithm is $x^{(k+1)} = x^{(k)} + \eta^{(k)} P_k$ in which it represent the line search to determine the optimal distance to move along the current search direction, where

$$\eta_k = \arg \min_{\eta \geq 0} E(x^{(k)} + \eta P_k) \quad (15)$$

and P_k is the new search direction. It is combination of steepest descent and the previous search direction. Thus,

$$P_k = -\nabla E(x^{(k)}) + \beta_k P_{k-1} \quad \text{where } \beta_k = \frac{\|\nabla E(x^{(k)})\|_2^2}{\|\nabla E(x^{(k-1)})\|_2^2}. \quad (16)$$

The conjugate gradient algorithm can minimize a quadratic function in n or less iteration than the steepest gradient descent method.

Line Search

Conjugate gradient require that a line search is performed to locate the minimum point of the approximated function. The Golden Section Search method is used to locate the initial interval where the minimum of the function occurs. This search is accomplished by evaluating the performance at a sequence of points, starting at a distance, ε . The distance is doubled at each step along the search direction. This first stage search is stopped when there is an increase in their performance value between two successive iterations; then the minimum is bracketed.

Golden Section Search

In this search, the step size is reduced within the interval containing the minimum point (the optimal global point). So, two new points are located within the initial interval. The values of these two points determine a section of the interval that can be discarded and replace them with a new pair of points within the new interval. The procedure continues until the interval is reduced to a width less than tolerance ($b_{k+1} - a_{k+1} < tol$). In order to find the interval locations, the interval step size is assumed as: $\varepsilon = 0.075$.

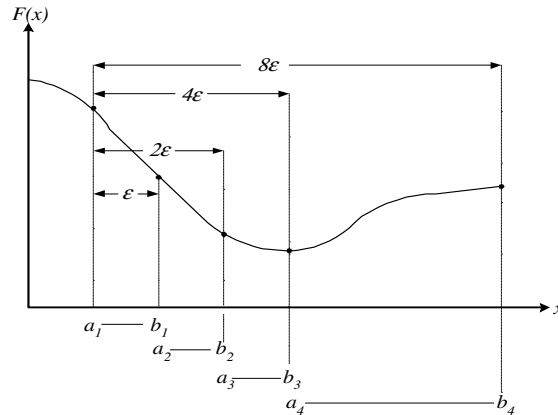


Figure 2:Interval locations for conjugate gradient global minimum search (Golden Section Search)

A comparison of the convergence of back-propagation with steepest gradient descent, Quasi Newton method, and Conjugate gradient method is made. The test is performed for a number of subjects and the overall trend showed is as in Figure 2. The steepest gradient descent method (sgm) starts at a lower SSE but settled at a higher error than the CGM. The convergence of CGM has less SSE compared especially after 2000 epochs. The oscillation of the sum squared error in the CGM (during the first 800 epochs) is due to the rapid adjustment made of the weights between layer connections in the golden section region.

From the literature review, the recognition rate when using the conjugate gradient descent is not greatly improved compared to using the steepest gradient descent, only that the convergence rate is faster when using the conjugate gradient descent algorithm. In Figure 2, the curves tell how the search for optimal global minimum behaved for each type of the gradient search. In comparison of the three curves: the steepest gradient descent seems to reach the convergence at the fastest rate, but not to the optimal point.

EXPERIMENTAL RESULT

Results on LPC

The LPC coefficient, size of window and overlapping were obtained through try and error. A set of experiments had been carried out to determine the best combination of LPC coefficients and size of window. Neural network architecture with 4 hidden nodes, eight output nodes with learning rate of 0.1 and momentum term of 0.9 was used to train the heart diagnostic system. A series of experiments had been carried out to determine the best neural network parameters such as learning rate, momentum term and number of hidden node. The first experiment was carried out to determine the number of LPC coefficients for optimal accuracy. The neural network was fixed with 4 hidden nodes and learning rate and momentum term of 0.1 and 0.9. The best performance came from a window size of 18 with 18 LPC coefficients and overall of system performance of 96.7%.

TABLE 1: Experiment on Window size and LPC coefficient for heart sounds

No	Window Size	Overlap	LPC coeff	Input No	Accuracy (%)
1	5	8	2	98	77.42
2			4	196	77.42
3			6	294	83.871
4			8	392	77.42
5		10	2	96	67.741
6			4	192	87.1
7			6	288	83.87
13	10	16	2	48	77.42
14			4	96	83.87
15			6	144	77.42
16		20	2	46	80.645
17			4	96	93.548
18			6	144	77.42
19	18	24	16	208	51.613
20			18	234	93.548
21			20	260	87.1
22		28	18	234	96.774
23			20	260	80.645
24	24	36	26	234	90.323

25			28	252	90.323
26			29	261	90.323
27		48	26	234	67.742
28			28	252	87.1
29			30	270	83.87

Result on Instantaneous Energy and Frequency Estimate

The second experiment is carried out using IEFE method. This method converts the signal into energy and frequency representation with the respect to the original signal. From the signal, 10 coefficients we selected based on the important event of human heart which is the pumping activity of the heart. This coefficient is fed to neural network for classification. A series of experiment has been conducted to investigate the parameter of hidden node, learning rate and momentum term.

Input Node:10, Learning rate = 0.1, Momentum term = 0.9

TABLE 4: Result based on IEFE method

No	Hidden node	Accuracy (%)
1	2	77.419
2	4	93.548
3	6	93.548
4	8	83.871

The result shows the recognition accuracy of 93.5%. IEFE represent the heart sound signal in term of strength and frequency intensity of the sound. The accuracy is slightly lower due to noise interference because this technique is exposed to a lot of noise especially when the acquisition is not properly done.

Result on Time-Frequency Analysis

In the following experiment, the heart sound sample is transformed using TFD utilizing B-Distribution as a transformation kernel. This will result a time and frequency representation with the high resolution plane. In addition to an appropriate process, it is often essential that dimensionality reduction be performed on the TFD, so that the information is sufficiently compact for presentation to a classifier. The main goal of SVD technique is to ensure that as much relevant information as possible is preserved in as few dimensions

as possible. By using the time-frequency method, the accuracy obtained is 90% which is slightly lower compared to LPC and IEFÉ.

TABLE 5 Result based on Time-Frequency Method

Item	TFD
Feature Dimension	256
Training Data	50
Testing Data	50
Hidden Layer	20
Learning Rate	0.05
Momentum Term	0.1
Accuracy	90%

CONCLUSIONS

Biomedical signals are apparently random or aperiodic in time. Studies have shown that random behavior can arise in deterministic nonlinear system with just a few degrees of freedom. This give hope to provide simple mathematical model for analyzing physiological system. There are many challenges ahead and challenges questions to answer. Which mathematical model are we supposed to choose to manage this imperfect knowledge?. How best is our knowledge presentation related to the given problem?. Biomedical signals that have been analyzed using techniques in this paper include heart rate, heart sound, ECG waveform. Abnormalities in the temporal durations of the segments between the deflections or of intervals between waves as well as their relatives heights serve to expose and highlight physiological dysfunction. In Biomedical engineering as well as in other domains imperfect knowledge can never be avoided. It is a tedious and challenging task to develop an automatic system to provide classification or pattern recognition tools to help specialist make a decision.

ACKNOWLEDGMENT

Special thanks to Rubita Sudirman and Kamarulafizam for their help and support.

REFERENCES

- Barkat B. and Boashash B. 2001. A High-Resolution Quadratic Time–Frequency Distribution for Multicomponent Signals Analysis. *IEEE Transactions on Signal Processing*. **49**(10). October 2001. 2232-2239
- B, Tovar-Corona, MD Hind, JN Torry, R Vincent, 2001. Effects of respiration on heart sounds using time-frequency, Analysis. *Computers in Cardiology*; **28**; 457-460
- Boashash, B. 1992. *Time Frequency Signal Analysis: Methods and Applications*. Melbourne, Australia: Longman Cheshire
- Boadshah,B., 1992. Interpreting and Estimating the Instantaneous Frequency of a Signal-Part II: Algorithms, *Proceeding of IEEE*, April, pages 512-538
- Boashash, B. .1992. Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals, *Proceedings of the IEEE* , 80(4), 520 – 538
- Cohen L. 1989. Time-Frequency Distributions: A Review. *Proceedings of the IEEE*. **77**(7). July 1989. 941-981.
- Daliman, S. and Sha’ameri, A. Z. 2003. Time-Frequency Analysis of Heart Sounds using Windowed and Smooth Windowed Wigner-Ville Distribution. *Proceedings Seventh International Symposium on Signal Processing and its Application*. 2. July 1-4. 625-626.
- Davis, S. B. and Mermelstein, P. 1980 Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. *IEEE Transactions on Acoustics, Speech and Signal Processing*. **28**(4). August 1980. 357-366.
- Jozef Wartak,M.D., 1972. *Phonocardiology, integrated study of heart sounds and murmurs,*” Harper & Row Publishers
- Molau, S., Pitz, M., Schluter, R. and Ney, H. 2001. Computing Mel-Frequency Cepstral Coefficients on the Power Spectrum. Proceedings ICASSP '01) 2001 *IEEE International Conference on Acoustics, Speech and Signal Processing*. May 7-11. 73-76.

- Oppenheim A. V., Schaffer R.J., 1978. *Discrete-time Signal Processing*”, Prentice Hall, New Jersey
- Oppenheim A. V., 1978. Schaffer R.J.,*Discrete-time Signal Processing*, Prentice Hall, New Jersey
- O'Shaughnessy, D. 1988, Linear predictive coding, *IEEE Potentials* , 7(1), 29 –32.
- Picone, J. W. 1993. Signal Modeling Techniques in Speech Recognition. *Proceedings of the IEEE*. 81(9). September 1993. 1215-1247.
- Rabiner L.; Juang B. H. 1993. *Fundamentals of Speech Recognition*, Englewood Cliffs, New Jersey: Prentice Hall,
- Sucic, V., Barkat B. and Boashash, B. 1999. Performance Evaluation of the B-Distribution. *Fifth International Symposium on Signal Processing and its Applications*. August 22-25. Brisbane, Australia: IEEE, 267-270.
- Zhenyu Guo, Moulder C., Durand L. G., Murray Loew, 1998. Development of a Virtual instrument for data acquisition and analysis of the phonocardiogram, *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 20(1)