A Neural Network Approach to Distinguish Parkinsonian Tremor from Advanced Essential Tremor

A. Hossen[1], M. Muthuraman[2], J. Raethjen[2], G. Deuschl[2], and U. Heute[3]

[1] Department of Electrical and Computer Engineering, Sultan Qaboos University, Oman
abhossen@squ.edu.om, {m.muthuraman,j.raethjen, g.deutschl}@neurologie.uni-kiel.de,
[2] Department of Neurology, College of Medicine, University of Kiel, Germany
[3] Institute of Circuit and System Theory, Faculty of Engineering, University of Kiel, Germany
uh@tf.uni-kiel.de

Abstract. A new technique for discrimination between Parkinsonian tremor and essential tremor is investigated in this paper. The method is based on spectral analysis of both accelerometer and surface EMG signals with neural networks. The discrimination system consists of two parts: feature extraction part and classification (distinguishing) part. The feature extraction part uses the method of approximate spectral density estimation of the data by implementing the wavelet-based soft decision technique. In the classification part, a machine learning approach is implemented using back-propagation supervised neural network. The data has been recorded for diagnostic purposes in the Department of Neurology of the University of Kiel, Germany. Two sets of data are used. The training set, which consists of 21 essential-tremor (ET) subjects and 19 Parkinson-disease (PD) subjects, is used to obtain the important features used for distinguishing between the two subjects. The test data set, which consists of 20 ET and 20 PD subjects, is used to test the technique and evaluate its performance.

Keywords: Wavelet-Decomposition, Soft-Decision Technique, Parkinsonian Tremor, Essential Tremor, EMG, Accelerometer, Artificial Neural Networks.

1 Introduction

Essential tremor (ET) and the tremor in Parkinson's disease (PD) are the two most common pathological tremor forms encountered in clinical neurology [1]. Differential diagnosis between the two tremors is usually achieved clinically. But there is a certain overlap in the clinical presentation between the two diseases that can make the differentiation on purely clinical grounds difficult [2]. In such cases, functional imaging of the dopaminergic deficit as the hallmark of PD is considered the diagnostic gold standard [3-4]. However, this requires SPECT (Single Photon Emission Computer Tomography) technology, injection of a radioactivity-labeled

dopamine transporter ligand into the patients (DAT-Scan), and needs a considerable amount of time. Thus more readily available and easier diagnostic tests are desirable [5]. Spectral analysis of tremor time-series recorded by accelerometry and surface EMG is a common approach [6]. It has proven useful to distinguish between physiological and pathological tremor [7], but is not superior to the clinical assessment in the distinction of ET from PD in its present form [8]. Therefore methods beyond the standard spectral analysis of the recorded tremor time-series have been applied to safely separate ET and PD [9-14]. A new approach of spectral analysis is investigated in [15]. This approach is based on a soft-decision wavelet-decomposition technique and it succeeds in obtaining 85% accuracy of discrimination of ET from PD. In this paper the wavelet-based soft-decision approach is combined with a neural network. The organization of the paper is as follows:

In section 2, both the trial data and test data are described. Section 3 contains the main idea of the soft-decision wavelet-based technique. The results of its implementation on test data using neural networks and discussion of the results are given in section 4. Conclusions are given in section 5.

2 Data

In this study, a total of 39 PD and 41 ET subjects were analyzed. The training set consists of 21 ET and 19 PD subjects, while the test set consists of 20 ET and 20 PD subjects. The mean age, sex and disease duration of the PD patients were compared with the ET patients for the trial and test data in Tables 1 & 2, respectively.

Table 1. Description of trial data of both PD and ET subjects

	PD	**ET**
Number of Patients	19	21
Mean Age (Range)	64.54 (40-90) Years	63.24 (27-94) Years
Gender (Male/Female)	11/8	12/9
Mean Disease Duration	16.4 Years	34 Years

Table 2. Description of test data T of both PD and ET subjects

	PD	ET
Number of Patients	20	20
Mean Age (Range)	68.22 (52-85) Years	64.52 (32-86) Years
Gender (Male/Female)	12/8	11/9
Mean Disease Duration	15.3 Years	29 Years

A piezoelectric accelerometer of about 2g was fixed to the dorsum of the more affected hand (selected clinically) in the middle of the third metacarpal bone, and bipolar surface-EMG recordings with silver-silver-chloride electrodes from forearm flexors (EMG1) and extensors (EMG2) were taken. All data were sampled at 800 Hz. The EMG was band-pass filtered between 50 and 350 Hz and full-wave rectified. The relatively high sampling frequency was useful for the EMG recordings as within the bursts there are frequency components up to 350 Hz and can only be fully picked up with such a sampling frequency to satisfy the Nyquist theorem.

3 Soft-Decision Wavelet-Decomposition

3.1 Wavelet-Decomposition

The block-diagram of a one-stage wavelet-decomposition is shown in Fig.1.

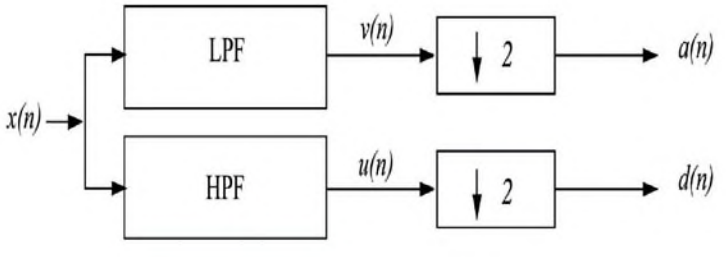

Fig. 1. A Single Stage of Wavelet Decomposition

The input signal *x(n)* of length-*N* is first filtered by low-pass (LPF) and high-pass (HPF) filters and then down-sampled by a factor of 2 to produce both the "approximation" *a(n)* and the "details" *d(n)*. Assuming Hadamard-filters are used, a*(n)* and *d(n)* can be obtained by:

$$a(n) = \frac{1}{2}[x(2n) + x(2n+1)],$$

$$d(n) = \frac{1}{2}[x(2n) - x(2n+1)] \tag{1}$$

If there is no information about the energy distribution of the input sequence, a band-selection algorithm [16] can be used to decide (as a hard decision) which band is to be computed or kept for more processing. This method depends on the energy comparison between the low- and high-frequency subsequences after the down sampling in Fig.1.

$$B = \sum_{n=0}^{\frac{N}{2}-1} (a(n))^2 - (d(n))^2 \tag{2}$$

According to the sign of B, the decision is taken: If B is positive, the low-frequency band is considered, and if B is negative, the high-frequency band is considered. Since we are not interested in the value of B, but only in its sign, a more-simpler equation than Eq.2 can be obtained approximately as [17]:

$$\mathrm{sgn}(B) = \mathrm{sgn} \sum_{n=0}^{\frac{N}{2}-1} (|a(n)| - |d(n)|) \tag{3}$$

3.2 Soft-Decision Algorithm

The soft-decision algorithm can be summarized in the following steps:

1) The one stage decomposition (Fig.1) is computed with all branches up to a certain pre-selected stage.
2) All estimator results up to this stage are stored, and their outputs are given a probabilistic interpretation by assigning a probability measure to each path.
3) If J(L) is the assigned probability of the input signal being primarily low-pass, the number J(H) = 1- J(L) is the probability that the signal is primarily high-pass.

At the following stage, the resulting estimate can be interpreted as the conditional probability of the new input sequence containing primarily low (high) frequency components, given that the previous branch was predominantly low (high)-pass. Using this reasoning and laws of probability, the assignments for the probability measure of the resulting subbands should be made equal to the product of the previous branch probability and the conditional probability estimated at a given stage (see Fig.2). The above probabilities derived from the estimator outputs may be interpreted themselves as a coarse measurement of the power spectral density

[17]. The higher the probability value of a particular band, the higher is its power-spectral content! So, after m-stage decomposition, a staircase approximation of the PSD is obtained, when the 2^m probabilities are plotted. For m=8, 256-subbands are resulted, each covering 400/256 Hz of the spectrum (0-400) Hz.

4 The Artificial Neural Networks

Before discussing the classification network used in this work, let us define the key features used in classification. The power spectral density of the first 16 bands out of 256 bands of Accelerometer, EMG1 and EMG2 signals is used as key features of the neural network. The selection of the 16 bands was on a prior information that the frequency of both tremors and their important harmonics are allocated in those bands.

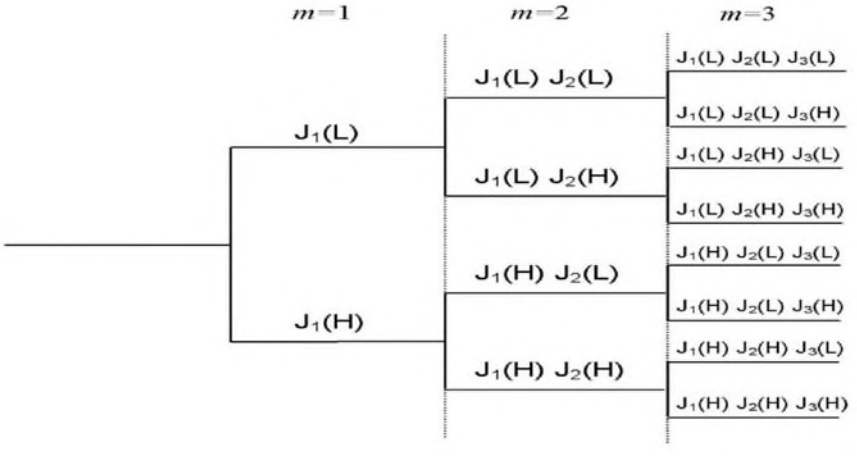

Fig. 2. The Soft-Decision Algorithm

4.1 Supervised Neural Network

In supervised learning, the training input data is composed of feature vector and the corresponding target output values. This approach is commonly described as learning with a teacher, since the desired output of a given input vectors is known and used during the learning process. The implementation of the supervised classification network is done according to the following two steps:

4.1.1 Training Stage of the Supervised ANN

A neural network of the type feed-forward back-propagation [18] (referred to as a multi-layer perceptron) is used in this approach. This network consists of three layers. The first layer (input layer) accepts 16 input signals (B1 to B16) from the

outside world and redistributes these signals to all neurons in the second layer. Actually, the input layer does not include computing neurons. The second layer (hidden layer) with a size of (3, 1) has three hyperbolic tangent sigmoid "*tansig*" neurons. The selection of "*tansig*" neurons was due to the nature of the difference of input features (PSD of the 16 bands) between the two classes (PD and ET) under investigation. Any PSD value is located between a maximum and a minimum, which can be easily simulated by a "*tansig*" function. The Neurons in the hidden layer detect the features; the weights of the neurons represent the features hidden in the input patterns. These features are then used by the third layer (output layer) in determining the output pattern. This third layer has one linear "*purelin*" neuron in our approach since the output is one out of two cases (PD or ET). The whole network has a single output that corresponds to one out of the two types under classification (ET or PD). The features extracted are the outputs of the properly designed hidden layer.

Fig.3 shows the three-layer back-propagation neural network used in the training. Since the neural network needs large set of data for training, a 2000 data set (1000 PD and 1000 ET) is simulated randomly to satisfy the spectral ranges of the 16 bands obtained from 21 ET and 19 PD (the original training data). Any new data is simulated by assignments of 16 PSD values corresponding to the 16 frequency bands in such a way that the PSD of any region is selected randomly between the minimum PSD and the maximum PSD of that region in the original set of data. The assigned PSD value is a random number having a mean value equals the average between the minimum PSD and the maximum PSD and a standard deviation with 10% to 20% of the maximum PSD.

4.1.2 Testing Stage of the Supervised ANN

At this stage the power spectral densities of the test data (1000 ET and 1000 PD), simulated randomly to satisfy the spectral ranges of the 16 bands obtained from 20 ET and 20 PD subjects (of the original test data set), are fed to the 16 inputs of the neural network. The assumed output is either 1 for PD or 2 for ET. In this work, binary classification is considered, e.g. classification between two different cases, positive (PD) and negative cases (ET). The performance of a classifier is evaluated by three main metrics: Specificity, Sensitivity and Accuracy as follows [19]:

$$Specificity\ (\%) = \frac{TN}{TN + FP}.100 \tag{4}$$

$$Sensitivity\ (\%) = \frac{TP}{TP + FN}.100 \tag{5}$$

$$Accuracy\,(\%) = \frac{TP + TN}{T}.100 \tag{6}$$

where the entities in the above equations are defined in confusion matrix shown in Table 3, and T is the total number of data under test.

Table 3. The Confusion Matrix

	Predicted Class [a]	
Actual Class [a]	**Positive (P)**	**Negative (N)**
Positive (P)	TP	FN
Negative (N)	FP	TN

[a]Positive = PD, Negative = ET, T=True, F=False

Results are shown in Tables 4 and 5 using each signal (Accel., EMG1, EMG2) individually and using all three signals with 48-inputs to the neural network . In Table 5, the training and testing sets are interchanged to test the consistency of the algorithm and its data independency.

Table 4. Results obtained from test data

Signal	**Specificity**	**Sensitivity**	**Accuracy**
Accel.	59%	70.6%	64%
EMG1	98.5%	98.3%	98.4%
EMG2	55.2%	95.6%	75.4%
All Signals	88.2%	95%	91.6%

Table 5. Results obtained from training data

Signal	**Specificity**	**Sensitivity**	**Accuracy**
Accel.	85.8%	86.9%	86.3%
EMG1	80%	89.6%	84.8%
EMG2	92.5%	76.9%	84.7%
All Signals	94.9%	94%	94.4%

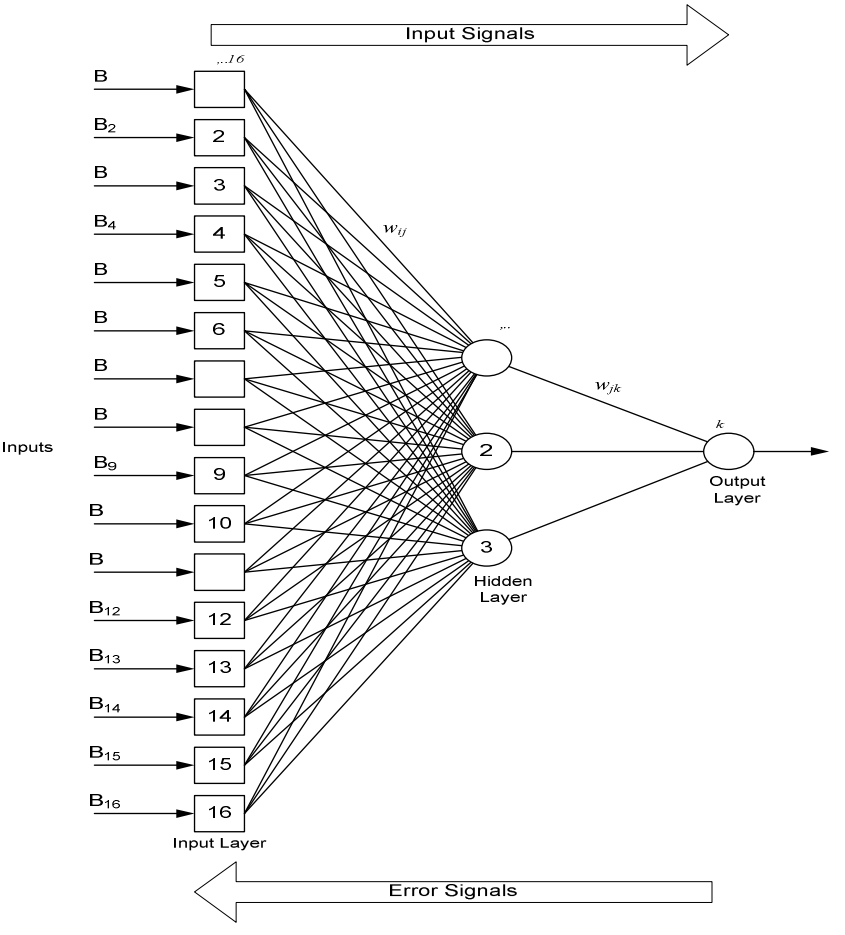

Fig. 3. The Artificial Neural Network Structure

5 Conclusions

A new identification method for PD from ET subjects is investigated. The method is based on the soft-decision wavelet-decomposition power spectral estimation and neural network. The first 16-bands out of 256 bands are used to represent the power-spectral density that forms the classification features. The accuracy of classification approaches 91.6% and 94.4% by testing the designed supervised neural network on test data set and training data set, respectively. In brief, the technique used in this paper, is a complete distinguishing system between ET and PD, that is data-independent, simple, efficient with automatic results.

References

[1] Harris, M.K., Shneyder, N., Borazanci, A., Korniychuk, E., Kelley, E., Minagar, A.: Movement Disorders. Med. Clin. North America 93, 371–388 (2009)

[2] Deuschl, G., Bain, P., Brin, M.: Consensus statement of the movement disorder society on tremor. Ad Hoc Scientific Committee, Mov. Disorder 13(suppl. 3), 2–23 (1998)

[3] Marshall, V., Reininger, C.B., Marquardt, M., Patterson, J., Hadley, D.M., Oertel, W.H., Benamer, H.T., Kemp, P., Burn, D., Tolosa, E., Kulisevsky, J., Cunha, L., Costa, D., Booij, J., Tatsch, K., Chaudhuri, K.R., Ulm, G., Pogarell, O., Hoffken, H., Gerstner, A., Grosset, D.G.: Parkinson's disease is overdiagnosed clinically at baseline in diagnostically uncertain cases: A 3-year European multicenter study with repeat [(123)I] FP-CIT SPECT. Mov. Disord. 24, 499–507 (2008)

[4] Djaldetti, R., Nageris, B.I., Lorberboym, M., Treves, T.A., Melamed, E., Yaniv, E.: [(123)I]- FP-CIT SPECT and oldfaction test in patients with combined postural and rest tremor. J. Neural Transm. 115, 469–472 (2008)

[5] Antonini, A., Berto, P., Lopatriello, S., Tamma, F., Annemans, L., Chambers, M.: Cost-effectiveness of [(123)I]-FP-CIT- SPECT in the differential diagnosis of essential tremor and Parkinson's disease in Italy. Mov. Disord. 23, 2202–2209 (2008)

[6] Deuschl, G., Krack, P., Lauk, M., Timmer, J.: Clinical neurophysiology of tremor. J. Clin. Neurophysiol. 13, 110–121 (1996)

[7] Raethjen, J., Lauk, M., Koster, B., Fietzek, U., Friege, L., Timmer, J., Lucking, C.H., Deuschl, D.: Tremor analysis in two normal cohorts. Clin. Neurophysiol. 115, 2151–2156 (2004)

[8] Bain, P., Brin, M., Deuschl, G., Elble, R., Jankovic, J., Findley, L., Koller, W.C., Pahwa, R.: Criteria for the diagnosis of essential tremor. Neurology 54, S7 (2000)

[9] Deuschl, G., Lauk, M., Timmer, J.: Tremor classification and tremor time series analysis. CHAOS 5(1), 48–51 (1995)

[10] Spieker, S., Jentgens, C., Boose, A., Dichgans, J.: Reliability, specificity and sensitivity of long-term tremor recordings. Electroencephalography and Clinical Neurophysiology 97, 326–331 (1995)

[11] Sapir, N., Karasik, R., Havlin, S., Simon, E., Hausdorff, J.M.: Detecting scaling in the period of dynamics of multimodal signals: Application to Parkinsonian tremor. Physical Review E 67(031903), 1–8 (2003)

[12] Breit, S., Spieker, S., Schulz, J.B., et al.: Long-term EMG recordings differentiate between parkinsonian and essential tremor. J. Neurol. 255, 103–111 (2008)

[13] Elble, R.J.: Essential tremor frequency decreases with time. Neurology 55, 1547–1551 (2000)

[14] Cohen, O., Pullman, S., Jurewicz, E., et al.: Rest tremor in patients with essential tremor: prevalence, clinical correlates, and electrophysiologic characteristics. Arch. Neurol. 60, 405 (2003)

[15] Hossen, A., Muthuraman, M., Raethjen, J., Deuschl, G., Heute, U.: Discrimination of Parkinsonian tremor from Essential tremor by implementation of a wavelet-based soft-decision technique on EMG and Accelerometer signals. Biomedical Signal Processing and Control 5, 181–188 (2010)

[16] Hossen, A., Heute, U.: Fully Adaptive Evaluation of SB-DFT. In: Proceedings of IEEE Int. Symp. on Circuits and Systems, Chicago, Illinois (1993)
[17] Hossen, A.: Power spectral density estimation via wavelet decomposition. Electronics Letters 40(17), 1055–1056 (2004)
[18] Principe, J., Euliano, N., Lefebvre, W.: Neural and Adaptive Systems: Fundamentals Through Simulations. John Wiley & Sons (2000)
[19] Rangayyan, R.M.: Biomedical Signal Analysis: A Case-Study Approach. IEEE Press, NJ (2001)