E.W. Lang, P. Gruber University Regensburg, Germany A. Martins da Silva HGSA,ICBAS/IBMC, Portugal

ABSTRACT

In this work we will present a method based on singular spectrum analysis (SSA), Ghil et al. (1) and Golyandina et al. (2), to remove ocular artifacts from an Electroencephalogram (EEG). After embedding the EEG signals in a feature space of time-delayed coordinates, the principal directions are computed. Using the projections of the embedded signals onto the eigenvectors corresponding to the largest eigenvalues, the Electrooculogram (EOG) signal is extracted. Thereby it is tacitly assumed that, as the EOG artifact represents a large-amplitude signal, it should be associated with the largest eigenvalues. We incorporate a Minimum Description Length (MDL) criterion based on information theory to determine the appropriate number of eigenvalues which correspond to the EOG signal. The extracted EOG signal is subtracted from the original EEG signal to obtain the artifact-free signal we are interested in.

Keywords: Singular Spectrum Analysis, Denoising, Singular Value Decomposition, EOG removal.

INTRODUCTION

Ocular activity creates artifacts in Electroencephalogram(EEG) signals, especially those recorded from frontal channels. When eyes move (blinking or other movements) the electric field around the eyes changes. producing an electrical signal known as Electrooculogram (EOG). As the signal propagates over the scalp, it also appears in the recorded Electroencephalogram (EEG). Figure 1 shows a segment of two EEG channels. The large amplitude artifactual EEG signal is more visible in the frontal channel (Fp1-Cz) than in the derivation of the occipital region (O1-Cz). In some studies, like single-trial event-related potentials (ERP), data from frontal electrodes located near the eyes are often not used in the analysis. Another strategy is to reject all segments of signals (epochs) that possibly contain artifacts. The task then is to exclude, manually or semi-automatically, all epochs which contain signals larger than a given arbitrary threshold. This is a tedious work and results in a substantial data loss. A variety of automatic procedures have been addressed to correct or remove ocular artifacts from EEG recordings: Wallstrom et al. (3) propose a method based on regression analysis; Vigário (4), Jung et al. (5, 6) apply independent component analysis(ICA) and He et al. (7) use adptative filters. Jung et al. (5)

show that ICA has an higher performance when compared with PCA, where some remnants of the Electrooculogram (EOG) are still visible in the corrected data. However the identification of the components related with the ocular artifacts needs to be done manually in order to reconstruct the data without artifacts. It is also to be noted

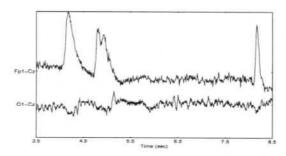


Figure 1: A segment of 2 EEG signals recorded from Fp1-Cz and O1-Cz, respectively

that those techniques were applied to the whole set of recorded EEG signals, which all included the EOG electrodes. In this work we will present a method based on singular spectrum analysis, Ghil et al. (1) and Golyandina et al. (2), to remove ocular artifacts from EEG recordings. We will try to extract the EOG signal using the information from a single channel of the EEG recording only. The SSA or multichannel singular spectrum analysis (MSSA) methods mainly comprise the following steps:

- · embedding
- singular value decomposition (SVD) or principal component analysis (PCA),
- · grouping and diagonal averaging.

In these steps free parameters are involved that must be adapted to the problem under study:

- · the dimension of the embedding and
- the grouping strategy.

In this work we suggest the use of an MDL criterion to separate the signal into two groups. Further on we propose a clustering step after the embedding. SVD and grouping are then performed locally in the clusters as it is proposed in Gruber et al. (8). The next section will present the basic steps of the SSA or MSSA technique and the principles of the MDL criterion. Then some numerical simulations are presented which corroborate the consistency of the MDL criterion in the determination of the signal subspace dimension when compared with other strategies. The removal of the EOG from the EEG signal is also presented for different variants of the SSA method discussed.

METHODS

Singular Spectrum Analysis (SSA) is a technique applied in time series analysis. The method has been widely used in climatic, meteorologic and geophysics time series(Allen and Robertson (9), Ghil et al. (1), Golyandina et al. (2)). The general purpose of the SSA analysis is the decomposition of the time series of interest into several additive components that typically can be interpreted as "trend" components (slowly varying parts of the series), various "oscillatory" components and "noise" components. The next sections present basic principles of the methods used in the experimental set-up. Starting by a brief description of SSA or MSSA methods, then the MDL criterion and finally a modified version of SSA that incorporates a clustering step are also described.

Singular Spectrum Analysis

Basic SSA has two main steps: decomposition and reconstruction. Each one comprising also two steps as described by Moskvina and Schmidt (10) and Golyandina et al. (2). A related technique is Multichannel Singular Spectrum Analysis (MSSA) which is an extension of SSA to more than one time series (Golyandina et al. (2)).

Decomposition

Embedding Embedding can be regarded as a mapping that transfers a one-dimensional time series $x_i = (x_i[0], x_i[1]...x_i[N-1])$ to a multidimensional sequence of lagged vectors. Let M be an integer (window length) with M < N. The embedding procedure forms K = N - M + 1 lagged vectors that constitute the columns of the trajectory matrix

$$\mathbf{X_i} = \begin{bmatrix} x_i[M-1] & x_i[M] & \dots & x_i[N-1] \\ x_i[M-2] & x_i[M-1] & \dots & x_i[N-2] \\ x_i[M-3] & x_i[M-2] & \dots & x_i[N-3] \\ \vdots & \vdots & \ddots & \vdots \\ x_i[0] & x_i[1] & \dots & x_i[N-M] \end{bmatrix}$$

Note that the matrix has the same entries along the diagonals.

MSSA is an extension of SSA to multidimensional time series, i.e. $\mathbf{x}[n] = [x_1[n], x_2[n]...x_L[n]]^T$. In that case

the global trajectory matrix is a concatenation of elementary trajectory matrices formed for each $x_i[n], i=1,...,L$

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_L]^T \tag{2}$$

with dimension $LM \times (N-M+1)$, where L is the number of sub-matrices with the structure described by (1).

Singular Value Decomposition This step comprises the singular value decomposition (SVD) of the trajectory matrix. Let $\mathbf{S} = \langle \mathbf{X}\mathbf{X}^T \rangle$ be the correlation matrix of zero-mean data, the eigenvalues of \mathbf{S} can be taken in decreasing order of magnitude $(\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_{LM} \geqslant 0)$ and let $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{LM})$ be the corresponding eigenvectors. Further let $d = \max\{l|\lambda_l>0\} \leqslant LM$. If

$$\mathbf{v}_{l} = \mathbf{X}^{T} \mathbf{u}_{l}(\lambda_{l})^{-1/2}, (l = 1...d)$$
 (3)

then the SVD of the trajectory matrix X can be written as

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_d \tag{4}$$

where

$$\mathbf{X}_{l} = (\sqrt{\lambda_{l}}\mathbf{u}_{l})\mathbf{v}_{l}^{T} = \mathbf{u}_{l}\mathbf{u}_{l}^{T}\mathbf{X}$$
 (5)

are the elementary matrices.

Reconstruction

Grouping Once the expansion is achieved, the grouping procedure partitions the set of indices $I = \{1, ..., d\}$ into m disjoint subsets $I = I_1, ..., I_m$ and the expansion (4) leads to the decomposition

$$X = X_{I_1} + X_{I_2} + ... + X_{I_m}$$
 (6)

where X_{I_m} corresponds to the sum of elementary matrices whose indices belong to subset I_m . The procedure of choosing the sets $I_1, ..., I_m$ is called eigentriple grouping (Golyandina et al. (2)). This step is dependent on the goal of the time series analysis.

In this work we consider a denoising task where a large-amplitude EOG signal has to be extracted. So only two groups m=2 are considered, where one of the groups comprises the largest eigenvalues which must be related to the large amplitude EOG signal. The grouping strategy principle will be discussed in more detail in the next section.

Diagonal averaging The last step is to transform each matrix of the grouped decomposition (6) into a sequence with N samples. None of the grouped matrices \mathbf{X}_{I_k} has the structure of a trajectory matrix. Hence the first step consists in substituting all the elements in each diagonal by its mean such that the resulting matrix will be a trajectory matrix of the desired sequence. The averaging procedure guaranties that the Frobenius norm of the difference between the original matrix and the transformed

matrix has minimum value among all possible solutions to get a matrix with all diagonals equal. In MSSA the diagonal averaging is done in the L separated sub-matrices of the global trajectory matrix.

Selection of signal subspace

In multidimensional signal processing analysis, the covariance matrix is modelled by the following equation

$$S = \Psi + \sigma^2 I$$

where Ψ is a semi-positive definite symmetric matrix of rank k (lower than its dimension p=LM) and represents the covariance of the signals, while $\sigma^2\mathbf{I}$ represents the covariance of the noise (assumed to be gaussian white noise). The largest k eigenvalues of \mathbf{S} are then related to the signals, while the remaining eigenvalues are related with the noise (ideally all of value σ^2).

The most widely used criterion to determine the dimension of the signal subspace k are based on the application of information theoretic principles like the MDL criterion (see Wax and Kailath (11), Fishler and Messer (12)). That determination is based on the application of a maximum likelihood estimation of the parameter vector of S which is $\theta = (\lambda_1, \lambda_2, \dots, \lambda_k, \sigma^2, \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k)$. Where λ_i , $i = 1, \dots, k$ are the k largest eigenvalues of S, and \mathbf{u}_i the corresponding eigenvectors and σ^2 corresponds to the mean of the discarded eigenvalues. Using a maximum likelihood estimate of $\hat{\theta}$, then k will be the value that minimizes the following expression

$$MDL(k) = -\ln f(\mathbf{X}|\hat{\theta}) + \frac{1}{2}K \ln N, \ k = 0,..., p-1$$
(7)

where N is the number of observations available to estimate the covariance matrix \mathbf{S} and $f(\mathbf{X}|\widehat{\boldsymbol{\theta}})$ denotes the conditional probability density parameterized by $\widehat{\boldsymbol{\theta}}$. The log likelihood function $L(\widehat{\boldsymbol{\theta}}) = \ln f(\mathbf{X}|\widehat{\boldsymbol{\theta}})$ represents the accuracy of the representation of the data with the parameter vector $\widehat{\boldsymbol{\theta}}$ and depends on the discarded eigenvalues

$$L(\widehat{\boldsymbol{\theta}}) = N(p - k) \ln \left[\frac{\prod_{i=k+1}^{p} \lambda_i^{1/(p-k)}}{\frac{1}{p-k} \sum_{i=k+1}^{p} \lambda_i} \right]$$
(8)

The negative log-likelihood $-L(\widehat{\boldsymbol{\theta}})$ is recognized to be a standard measure of the training error. However it has been reported that the simple maximization of this term tends to result in the phenomenon of over-fitting. Thus the second term in eqn. (7) was added as a regularization term to penalize complexity. The value of K is related with the number of parameters in θ and the complexity of its estimation. Considering real valued signals, the value of K is computed according to

$$K = k+1+pk-k^2/2-k/2 = pk-k^2/2+k/2+1$$
 (9)

Leonowicz et al. (13), Liavas and Regalia (14) report the behavior of the MDL criterion (and other criteria like Akaike's information criterion (AIC)) for the estimation of the dimension of the signal subspace considering different applications (like model order selection in autoregressive models). In what concerns the MDL criterion, it has been concluded that the smaller eigenvalues must be clustered together and must exhibit a gap to the larger eigenvalues in order to result in a reliable estimation of the signal subspace dimension.

Clustering and SSA

In Gruber et al. (8) a denoising technique is presented which uses a clustering step of the columns of the trajectory matrix. So after the embedding, the clustering is performed and the SVD is computed for the data in each cluster. After choosing the principal directions of each cluster, the data are reconstructed as described before. Finally, the clustering of the reconstructed data is reversed to obtain the reconstructed trajectory matrix and a diagonal averaging is performed.

In this work the algorithm K-means, as described in Bishop (15), was used to perform the clustering in a fixed number c of clusters according to the user choice.

NUMERICAL SIMULATIONS

In this section we will present numerical simulations concerning the denoising of one EEG signal using SSA and the use of MDL incorporated in SSA and MSSA applications as grouping strategy. A further study will present a set of artificial mixed noisy signals embedded in time-delayed coordinates. The EEG denoising is presented for one frontal channel using SSA and the MDL criterion. However, after embedding and before the grouping stage steps, we also use a pre-processing step based on clustering. Thereby we try to distinguish different amplitude levels of the input signal. With this pre-processing step, we then simplify the grouping step by joining only the largest eigenvalue.

On the use of the MDL criterion

Artificial signals (s) are mixed with a random matrix (A) and gaussian noise is added, i.e, $\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n] + \mathbf{e}[n]$. Using a segment of N samples of a L dimensional signal, $\mathbf{x}[n]$ is embedded in a space of delayed coordinates of dimension M as described in . The input space then has dimension LM and 1000 simulations where realized varying the mixing matrix and adding random noise such that the SNR keeps the same level.

We use different criteria to determine the dimension of the signal subspace k: MDL, Akaike's information and the percentage of data variance to be preserved. The Akaike (AIC) criterion is very similar to the MDL criterion; its definition can be found in Liavas and Regalia (14), Wax and Kailath (11). The variance criteria (VE) are used in pattern recognition applications to preserve a fixed percentage of the data variance (95% is the most

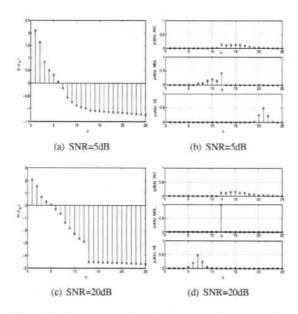


Figure 2: Parameters of the simulation: N = 5000, $A = 5 \times 3$, M = 5 (a) and (c) $\ln(\lambda_k)$ versus k, (b) and (d) pdf(k) of AIC, MDL and VE versus k (top-bottom)

common value).

Fig. 2 shows the main conclusions of these experiments. In that case the dimension of the global trajectory matrix is LM = 25 and N = 5000, and we can see (fig. 2 b and d) that the estimated probability density function of the signal subspace k exhibits a pronounced peak when the MDL criterion is applied. For this criterion, when the SNR is high ($\geq 20dB$) the probability has a value close to 1 for k = 12. It has to be noticed that there is a clustering of the lower eigenvalues (see fig. 2-c) as well as a clear gap. For lower SNR (5dB) the clustering is not so evident, and there is no gap between larger and smaller eigenvalues (see fig. 2-a). However, the largest value of the pdf occurs also for k = 12. Using the AIC, the estimated values for k yield values in the range (12-16) with almost the same probability both for low and high SNRs. The percentage of variance (VE) results in completely different estimates around k = 21 when SNR = 5dBand k = 7 if SNR = 20dB. This has to be expected, because a constant threshold is used, hence when the noise level increases it is as important as another signal. In conclusion this study shows that the MDL criterion gives a more consistent estimate of k when compared with the other criteria under the same conditions.

Denoising a single channel EEG signal

A frontal channel (Fp1-Cz) sampled at 128Hz was used. A segment of the signal with N=1664 samples ($\simeq 13s$) is considered. The one-dimensional signal (L=1) was embedded using M=40 in all the 3 experiments described next. The results will be compared visually in the time domain, but in some cases the power spectral density

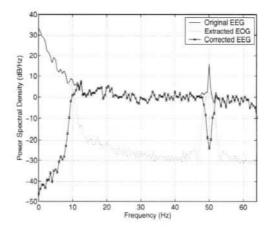


Figure 3: Power Spectral Density (dB/Hz) versus Frequency (Hz) of Original EEG, Extracted EOG and Corrected EEG

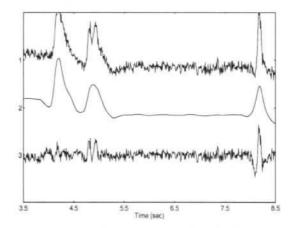


Figure 4: Signals obtained using 6 clusters and reconstruction with highest eigenvalue: original EEG, extracted EOG and corrected EEG.

computed by the Welch method (Oppenheim and Schafer (16)). EEG studies consist in many cases in studying the frequency contents of a set of bands: theta(3.5-7.5Hz), alpha (7.5-13Hz) and beta(13-25Hz). We also compared instantaneous measures of energy in some of those bands just to evaluate the differences between the corrected EEG and original EEG.

Denoising Using SSA and MDL

In this section we show the results with the basic SSA algorithm using the MDL criterion to compute the signal subspace dimension to reconstruct the EOG signal (extracted EOG). In this case the MDL criterion picks up the 9 largest eigenvalues. Then the extracted EOG is obtained using these 9 principal directions of the SVD of the trajectory matrix. The corrected EEG is then obtained by subtracting the extracted EOG from the original EEG

signal. A visual inspection of the extracted EOG signal shows that the 50Hz interference is also present. The power spectral density of the corrected signal (see fig. 3) is very low in the lower frequency band (corresponding to theta and alpha bands < 10Hz) whereas it is altered in the original signal by the presence of the EOG artifact. But the beta band is very similar to the original EEG except in the proximity of 50Hz line noise.

Denoising Using Clustering and SSA

Like in Gruber et al. (17) a clustering step precedes the PCA step, and a denoising of a high-amplitude signal was performed reconstructing the signal using only the largest eigenvalue. Then the columns of the trajectory matrix were clustered in 6 cluster. The data in each cluster were reconstructed using only the largest eigenvalue. To reconstruct the trajectory matrix, the clustering simply had to be reversed. The diagonal averaging will provide the extracted EOG signal. In figure 4 we can see that the extracted signal provides the ocular movements but the corrected EEG still contains some EOG artifacts. The power spectral density plot (not shown here) confirms that result, the plot of the corrected EEG is very similar to the original EEG.

Denoising Using Clustering, SSA and MDL

A strategy similar to the previous experiment is applied but an MDL criterion is used to choose the dimension of the signal subspace in each cluster. In figure 6 the eigenvalues of each cluster are plotted and the vertical trace indicates the estimated dimension of the subspace. It can be seen that the MDL estimate corresponds to the leveling off of the eigenvalues which, hence, form a distinct cluster related with noise signals. The EOG artifact is extracted as described above, and then the corrected EEG is obtained. All signals are shown in figure 5. The mean squared error between the extracted EEG using

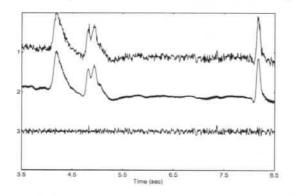


Figure 5: Signals obtained using 6 clusters and reconstruction with the subspace dimension given by MDL: original EEG, extracted EOG and corrected EEG.

MDL and clustering and the one obtained without cluster-

ing amounts to 6.76. Further it can be seen (figure 3 trace with symbol \square) that there is an increase of frequency content in lower bands. In fact, figure 7 shows that the fre-

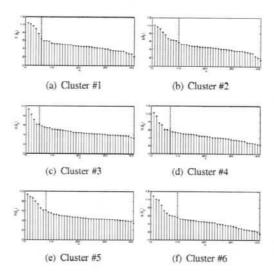


Figure 6: $ln(\lambda_k)$ versus k for each cluster. The vertical indicates the dimension estimated by MDL

quency contents of the lower band (< 10Hz) increases with the number of clusters. It should be noted that for practical reasons that increase is limited by the number of data vectors in a cluster, i.e, the number of vectors in cluster must be larger than M. We also tried to evaluate the frequency contents of theta and beta bands along the data segment. The lower band was chosen just because it is reported by Jung et al. (6) that PCA methods remove the lower frequency contents. The energy was estimated

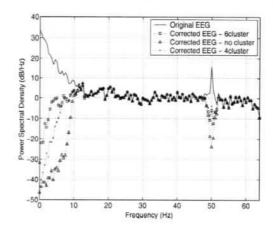


Figure 7: Power Spectral Density (dB/Hz) versus Frequency (Hz) of the original signal (\star) , the Corrected EEG Signal using 6 Clusters (\Box) , 4 Clusters (\bullet) and no Clustering (\triangle)

in segments of 2s, windowed with an Hamming window, corresponding to a frequency resolution of 0.5Hz with

an overlap of 50%. The figure 8 shows both results. The beta band is the dominant band in a frontal electrode and does not appear altered in the energy density estimated for the whole segment. In the lower bands using clustering (c=6), the energy contents in extracted EEG is more close to original EEG and it is more close when there is no ocular movement as is the case in the interval 5.5s-7.5s.

CONCLUSIONS

In this work we study the viability of SSA based methods to remove ocular artifacts from EEG signals. Preliminary results show that SSA with a clustering step provide promising results. Due to the nature of the noise extracted, the MDL criterion for grouping the eigenvalues proves to be effective. But there are also parameters (like the embedding dimension and the number of clusters) which had to be determined experimentally. Concerning the embedding, traditional SSA applications, (Golyandina et al. (2)) use an embedding dimension close to half of the length of the data segment (N). However, using the experimental data set, we also conclude that the choice of the subspace dimension k is less consistent among the different approaches when the number of samples was small. In particular, when N = 2 * LM, the pdf(k) function of the MDL criterion does not show a clear peak. Considering the EEG data, if the embedding dimension is very high (M > 150), the choice of the subspace dimension by the MDL criterion coincides with the dimension of the input space (which amounts to 150 in this case). For similar reasons, the number of clusters cannot increase in a way that the number of data vectors in one cluster is close to the dimension of the embedding space. If there are clusters, where the subspace dimension k coincides with the dimension of the input space, then there are segments of the corrected EEG which convey no information, having no frequency content up to 20Hz. All information then is contained in the extracted EOG. The experimental study is currently extended in our group to a larger set of signals with different characteristics.

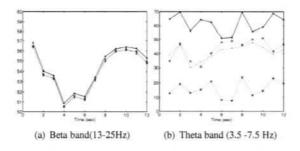


Figure 8: Energy along the segment: Original EEG (*); Corrected EEG with 6 clusters and mdl (\Box -); Corrected EEG with 4 clusters and MDL (•); Corrected EEG: no Cluster and MDL ∇)

References

- Ghil, M., Allen, M., Dettinger, M. D., Ide, K., and al, e., 2002, Reviews of Geophysics, 40(1), 3.1–3.41
- Golyandina, N., Nekrutkin, V., and Zhigljavsky, A., 2001, Analysis of Time Series Structure: SSA and Related Techniques, Chapman & HALL/CRC
- Wallstrom, G. L., Kass, R. E., Miller, A., Cohn, J. F., and Fox, N. A., 2004, Internat. Journ. of Psychophysiology,53,105–119
- Vigário, R.N., 1997, Electroenceph. Clin. Neurophys., 103, 395–404
- Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., and Sejnowski, T. J., 2000, Clinical Neurophysiology, 111, 1745–1758
- Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., Mckeown, M. J., Iragui, V., and Sejnowski, T. j., 2000, Psychophysiology, 37, 163–178
- He, P., Wilson, G., and Russel, C., 2004, <u>Medical & Biological Engineering & Computing</u>, 42, 407–412
- Gruber, P., Theis, F. J., Stadlthanner, K., Lang, E. W., and Tomé, A. M., 2004, in IEEE, editor, Intern. J. Conf. On Neural Networks, 1959–1964, Budapest
- Allen, M. R. and Robertson, A. W., 1996, Climate Dynamics, 12, 775–784
- Moskvina, V. and Schmidt, K. M., 2003, SIAM Journal Mat. Anal. Appl., 24(4), 932–942
- Wax, M. and Kailath, T., 1985,
 IEEE Trans. Acoust., Speech and Signal Processing,
 33, 387–392
- Fishler, E. and Messer, H., 2000, IEEE Trans. Signal Processing, 48(8), 2242–2247
- Leonowicz, Z., Karvanen, J., Tanaka, T., and Rezmer, J., 2004, in International Workshop Computational Problems of Electrical Engineering
- Liavas, A. P. and Regalia, P. A., 2001, IEEE Trans. on Signal Processing, 49(8)
- Bishop, C. M., 1995, Neural Networks for Pattern Recognition, Oxford University Press, Oxford
- Oppenheim, A. V. and Schafer, R. W., 1975, Digital Signal Processing, Prentice-Hall
- Gruber, P., Stadlthanner, K., Tomé, A. M., Teixeira, A. R., Theis, F. J., Puntonet, C. G., and Lang, E. W., 2004, in LNCS 3195, Proc. ICA' 2004, 993–1000, Springer, Granada

Proceedings of the

2nd International Conference on Computational Intelligence in Medical and Healthcare

CIMED2005

29 June - 1 July 2005 Costa da Caparica, Portugal

Organised by the IEE Healthcare Technologies

Professional Network Technically co-sponsored by the IEEE EMBSukri Chapter

Edited by Professor José Manuel Fonseca

Disclaimers

Author Disclaimer

"While the author and publisher believe that the information and guidance given in this work are correct, all parties must rely upon their own skill and judgement when making use of it. Neither the author nor the publisher assume any liability to anyone for any loss or damage caused by any error or omission in the work, whether such error or omission is the result of negligence or any other cause. Any and all such liability is disclaimed."

Copyright and Copying

All Rights Reserved. No part of this publication may be produced, stored in a retrieval system, or transmitted in any form or by any means - electronic, mechanical, photocopying, recording or otherwise - without the prior written permission of the publisher.

Copyright © 2005 by the IEE

Edited by Professor José Manuel Fonseca

No partial or complete copy of the material in this proceeding is allowed without specific concern of the publishers.

ISBN: 0-86341-520-2