31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Seizure Prediction for Epilepsy using a Multi-Stage Phase Synchrony based System

Christopher J. James*, *Senior Member, IEEE*, and Disha Gupta, *Student Member, IEEE*

***Abstract*—Seizure onset prediction in epilepsy is a challenge which is under investigation using many and varied signal processing techniques. Here we present a multi-stage phase synchrony based system that brings to bear the advantages of many techniques in each substage. The 1st stage of the system unmixes continuous long-term (2-4 days) multichannel scalp EEG using spatially constrained Independent Component Analysis and estimates the long term significant phase synchrony dynamics of narrowband (2-8 Hz and 8-14 Hz) seizure components. It then projects multidimensional features onto a 2-D map using Neuroscale and evaluates the probability of predictive events using Gaussian Mixture Models. We show the possibility of seizure onset prediction within a prediction window of 35-65 minutes with a sensitivity of 65-100% and specificity of 65-80% across epileptic patients.**

I. Introduction

THE unpredictable nature of epileptic seizures have for long made epilepsy a debilitating and challenging brain disorder for patients across the world. The efforts of suppressing the seizures with anti-epileptic medications and surgery usually prove useful in about 75% of the patients. Ongoing research on seizure prediction is also advancing with state-of-the-art signal processing approaches, motivated by the need to establish a warning or therapeutic intervention system to suppress seizures. Of the many approaches in the literature, we present here an analysis of the synchronization dynamics of the Electroencephalogram (EEG) which has recently attracted much attention [1],[2]. Previous research in this field has posed the hypothesis that during a seizure the seizure focus entrains neighbouring (not necessarily spatial) areas of the brain, leading to a hyper-synchronous state. This is preceded by the neighbouring areas losing their synchronization with the other cortical areas around them. This detachment causes the seizure focus to become isolated from the involvement of the rest of the brain dynamics as well as making available an idle population of neurons [3],[1]. Phase synchronization has been shown to be a sensitive indicator of nonlinear interactions between neurophysiological signals [1] which can be useful for seizure prediction as signals appear to lack morphological or topographical cues about an oncoming seizure. However, most of this literature has focused on phase synchronization analysis with 'raw' EEG and on short data sets. Here we show an advanced multi-stage phase synchronization based system for seizure prediction with long-term EEG. The system involves prior unmixing of scalp EEG through the blind source separation (BSS) technique of Independent Component Analysis (ICA). This allows an objective and fairly automatic processing of long-term continuous multichannel EEG. We have previously shown in [5] how we can use ICA with phase synchronization analysis for seizure prediction on long term EEG. Interesting cyclical patterns in the synchrony dynamics were observed and analysed subjectively. However, there was a need for an objective measure – a feature extraction system to determine the features that discriminate between the naturally occurring interictal and pre-ictal groups and to statistically analyze the possibility of seizure prediction. We show such a system here, based on a feature extraction and dimensional reduction method, followed by probabilistic density modelling. The system aims to elicit features from the synchrony dynamics that are indicative of an impending seizure and then track the probability of occurrence of a predictive event. The multi-dimensional feature space is reduced to a 2-dimensional visualisation space with a semi-supervised Radial Basis Network based technique called Neuroscale [6]-[11]. This low dimensional data is then used for probabilistic modelling with a Gaussian Mixture Model (GMM) [12]. A multitude of features are extracted from the complex phase synchrony values, themselves extracted from the de-noised (and umixed) EEG signals.

A. Synchrony dynamics of unmixed long-term EEG

Unmixing EEG with ICA: The unmixing of multichannel scalp EEG is performed with the help of ICA. ICA is a technique that performs BSS on linearly, instantaneously, squarely mixed, stationary and statistically independent sources. ICA essentially helps in unmixing the measured signals (scalp EEG in this case) into fairly independent or least dependent components (LDCs) as well as obtains the mixing information of each, as described in further detail in [13],[14]. However, EEG being a non-stationary signal; requires ICA to be performed on relatively short moving windows (generally about two minutes long), which makes tracking a physiologically relevant source across time quite challenging. To overcome this problem, spatially constrained ICA (ScICA) [15] has been used in this work, seeking relevant LDCs with a relatively known spatial

*C. J. James is with the Signal Processing and Control Group, ISVR, University of Southampton, SO171 BJ, UK. (Tel: +44 (0) 2380 593043; Fax: +44 (0) 23 8059 3190; e-mail: C.James@soton.ac.uk). D. Gupta is with the Signal Processing and Control Group, ISVR, University of Southampton, SO171 BJ, UK. e-mail: disha@soton.ac.uk).

distribution (in this case we use the location most active/ relevant to the seizure focus of the patient under test). ***Phase synchrony with PLV-d:*** The next step is extracting and tracking the synchronization dynamics from the unmixed spatially constrained seizure sources, in two relevant frequency bands (in this case, 2 Hz–8 Hz and 8 Hz–14 Hz). Phase synchrony measures the coupling between signals (irrespective of their amplitudes). It is useful in measuring non-linear relationships, even in chaotic, non-periodic signals [16]. It is based on the estimation of prolonged instants (non-overlapping windows of 5 seconds) of constant phase differences. The instantaneous phases are found with the Hilbert Transform (HT), which necessitates the use of the narrowband signals we describe, as shown in [17]. The index used to assess the strength of synchronization is called the Phase Locking Value (PLV) [1]. The significance of the PLV is tested with the Phase Locking Statistic (PLS), calculated using 100 surrogate series (phase randomized) and a bootstrap distribution. Additional processing is then performed to observe concurrent variations in phase synchrony in the two frequency bands, by calculating the difference between the PLS curves of the two bands, referred to as the PLV-d curves. A moving average filter of the order 300 PLV-d samples was then applied (equivalent to 1500 seconds), smoothing the curves over longer time spans (which are relevant to the dynamics we are testing for).

B. Feature extraction with Neuroscale

Multidimensional feature maps: A set of thirteen features were formulated, extracted from one hour moving windows with a 75% overlap of the PLV-d curves. The features used were: standard deviation, mean, ratio of peak to the preceding window mean, ratio of minimum amplitude to preceding window mean, gradient, gradient of previous window, mean of previous window, absolute maxima, absolute minima, energy, fluctuations of curves about the mean, dominant frequency trend and the number of zero level crossings of the PLV-d curve.

Neuroscale: A novel topographic feature extraction and dimensional reduction technique, Neuroscale, was used for visualising the high-dimensional datasets. Topographic extraction refers to a transformation of the data such that the geometric structure of the data is optimally preserved in the transformed space. This implies that the inter point distance in the feature space closely match the distances in the data space. The algorithm is based on a Radial Basis Function (RBF) Neural Network which is trained by adjusting the network parameters in order to minimize a topographic error measure known as the Sammon Stress metric*:* $E = \sum_{p}^{N} \sum_{q>p}^{N} (d_{qp}^{*} - d_{qp})^2$, where d_{qp}^{*} are the inter point Euclidian distances in the data space. In a supervised training 'mode', the Neuroscale algorithm is provided with additional data class information. These have been provided here in the form of a binary distance metric, where the distance between points within the same class is zero while between two different classes is one. The extent of supervision with this additional information is controlled by a parameter α, as the inter-point distance in the input space d_{qp}^{*} is replaced by $\delta_{qp} = (1-\alpha).d_{qp}^{*} + \alpha.s_{qp}$.

C. Probabilistic modelling of seizure prediction

The Neuroscale feature mapping is then followed by the semi-parametric probability density estimation technique using GMMs [12]. In a mixture model, a probability density function is defined as a linear combination of basis functions. Mixture models like the GMM output a weighted sum of their parametric mixture components, estimating the mixture coefficients and the parameters of the individual components. A Gaussian basis for a set of data points is parameterized by the mean, a covariance matrix and the weights of all component densities; represented as: $\theta = \left\{ \mu_j, \sum_j , P_j \right\}$, where j =1, 2,…, M.

The GMM has previously been used in unsupervised, semi-supervised or supervised manner for modeling various statistical distributions, including nonlinear distributions [16[1],[18]. Here it is being used in the supervised and the semi-supervised form. In a supervised form the GMM trains separately on the pre-ictal and interictal data sets while in a semi-supervised form, it trains on the labelled training data and additionally uses the structure learned from the unlabelled test data to enhance the classifier learning. This helps to incorporate information about the data/class distribution that a sparse training (labelled) data may not provide.

Formulation of the model: The two dimensional Neuroscale feature map is divided into a *training* set (labelled as pre-ictal or interictal) including at least 2-3 seizures, and a *test* (unlabelled) set. In the supervised mode (shown in Figure 1), the multimodal GMMs are first trained on each of the pre-ictal and interictal training sets. The probability density contours are then obtained from these trained GMMs. The learned set of parameters is then used to project the test data on to the GMM density maps obtained. In the case of semi-supervised GMM, initial training remains the same as in supervised GMM, additionally the GMM is trained on the unlabeled data and finally on the unlabeled data while using the information from the trained GMM of the labelled data. A few labeled data points included in the training can be helpful in obtaining a good generalization performance, which may achieve a low classification error [19]. The GMM training involves the expectation maximization algorithm [12] solving for the GMM parameters and alternating between providing soft labels to the unlabeled data. The classification accuracy is measured using the known label information of the test data in both cases.

II. RESULTS

Data*:** The data used is long term continuous scalp EEG from nine epileptic patients undergoing pre-surgical evaluation for possible epileptic surgery. The data is recorded continuously for 2-4 days using the 10-20 international electrode placement system. It is sampled at 200 Hz. Five datasets are used in feature extraction and four datasets are used to evaluate the algorithm on unseen patients. The results of only two patients are shown here (see Figure 1: Patient 1 and Patient 2) as examples. ***General observations from PLV-d curves*:* There were some general patterns observed from the PLV-d curves across patients, irrespective of the type of epilepsies. These are: (1) In general all PLV-d curves show a low frequency oscillatory pattern. This pattern fluctuates across the zero level. The zero level implies that the difference in the PLV's of the two frequency bands is zero which means that either the PLV's of the two frequency bands are equally strong and cancel each other out or are simultaneously zero. (2) The fluctuating rhythm appears to have a pattern that follows the day and night timings. (3) The cyclical pattern of the PLV-d curve is also seen to show a 'set-reset' pattern [2] before a seizure onset, which may be useful in indicating an oncoming seizure. The set-reset pattern appears to be indicative of the long term dynamics of synchrony. ***General observations from feature extraction:*** A clear distinction of pre-ictal and interictal feature points was not obtained and was not expected either. This was because the pre-ictal data windows had been defined as the windows that were about one hour prior to the seizure onset. It was quite possible that the data reflected interictal features when further away from the seizure. However, the GMM was able to model the training sets of interictal and pre-ictal feature points quite effectively, such that the test feature points were found to be near their respective centers. This gave a probability map of pre-ictal events. The probability was then scanned with various prediction windows (0 to 4 hours in steps of 5 minutes) and with varying threshold levels (0 to 1 in steps of 0.01) to find the prediction window that shows maximal predictions (high sensitivity of prediction) and minimal false alarms (high specificity of prediction). The ROC curves were obtained to find the prediction window that shows high sensitivity and high specificity. On average in the five patients, a prediction window of 35 minutes shows a prediction sensitivity of 86% and a specificity of 80% on average. In the four unseen datasets, a sensitivity of 65-100% and specificity of 65-80% is found with a 35-65 minute prediction window. The classification was also tested against a random predictor for various prediction windows. The ICA-PLVd-Neuroscale-GMM predictor demonstrated an ability to predict that was better than a random predictor.

III. CONCLUSIONS

We show here that a phase synchrony based system applied to unmixed epileptic EEG can show the underlying long term synchrony dynamics for seizure onset prediction. The overall performance of the algorithm cICA-PLV-d-supervised Neuroscale-GMM has been very encouraging. It has successfully demonstrated the existence of a predictive space based on synchrony dynamics, more suitably acquired through unmixed multichannel EEG signals. A prediction performance with the test data was found to show a specificity of 65%-100% and 65%-80% sensitivity with a prediction window of 35-65 minutes in general. It was also found to perform much better than a random predictor.

REFERENCES

[1] J. P. Lachaux, E. Rodriguez, J. Martinerie, F. J. Varela, "Measuring phase synchrony in brain signals," *Human Brain Mapping*, vol. 8(4), pp. 194-208, 1999.

[2] F. Mormann, K. Lehnertz, P. David and C. E. Elger, "Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients", *Physica D*, 114, pp. 358-369, 2000.

[3] M. Le. Van Quyen, "Anticipating epileptic seizures: from mathematics to clinical applications," *C R Biol.*, vol. 328(2), pp. 187-198, 2005.

[4] K. Ansari-Asl, L. Senhadji, JJ Bellanger, F. Wendling, "Quantitative evaluation of linear and nonlinear methods characterizing interdependencies between brain signals", *Phys Rev E Stat Nonlin Soft Matter Phys*. Vol 74, 2006.

[5] D. Gupta, C. J. James and W. P. Gray, "Phase synchronization with ICA for epileptic seizure onset prediction in the long term EEG," in 2008 *Proc. of MEDSIP Intl. Conf.*

[6] N. C. Noel, "Dynamical Embedding and Feature extraction of electroencephalographic data ," M.S. thesis, Aston Univ., UK., 1998.

[7] C. J. James and D. D'Alimonte, "Tracking multisource brain activity in multi-channel EEG with a probabilistic model through ICA," in *2004 Proc. MEDICON2004 Conf.*

[8] P. N. Hughes and L. Tarassenko, "Novel signal shape descriptors through wavelet transforms and dimensionality reduction," in 2003 *Proc. SPIE: Intl.. Soc. for Optical Engg*., pp. 763-773.

[9] M. E. Tipping, "Topographic Mappings and Feed-Forward Neural Networks," Ph.D. dissertation Aston University, U.K., 1996.

[10] D. Lowe, "Feed-forward neural networks and topographic mappings for exploratory data analysis," *Neural. Computing. and Application.*, vol.. 4, pp. 83-95, 1996.

[11] D. Lowe, "Feature space embeddings for extracting structure from single channel wake EEG using RBF networks," Neural Networks for Signal Processing VIII, IEEE Sig. Proc. Soc. , 31, 428-437, 1998.

[12] I. T. Nabney, *NETLAB Algorithms for Pattern Recognition*. Springer-Verlag London, 2004, pp. 79-116.

[13] C. J. James and C. W. Hesse, "Independent Component Analysis for biomedical signals," *Phys. Meas.,* vol. 26, pp. 15-39, 2005.

[14] A. Hyvärinen, J. .Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley Sons, USA, 2001.

[15] C. W. Hesse and C. J. James, "The FastICA algorithm with spatial constraints," *IEEE Signal Processing Letters*, vol. 12, 2005.

[16] M. G, Rosenblum, A. S. Pikovsky and J. Kurths, "Phase synchronization of chaotic oscillations," *Phys.Rev. Lett*., vol.76, pp. 1804-1807, 1996.

[17] D. Gupta and C. J. James, "Narrowband vs. broadband phase synchronization analysis applied to Independent Components of Ictal and Interictal EEG," in 2007 *Proc. of IEEE-EMBS 29th Intl. Conf.*

[18] W. Pieczynski, J. Bouvrais and C. Michel, "Estimation of generalized mixture in the case of correlated sensors*," IEEE Trans. on Image Proces*., vol. 9, pp. 308-311, 2000.

[19] M. O. Szummer, "Learning from Partially Labeled Data," Ph.D. dissertation, Massachusetts Institute of Technology, USA, 2002.

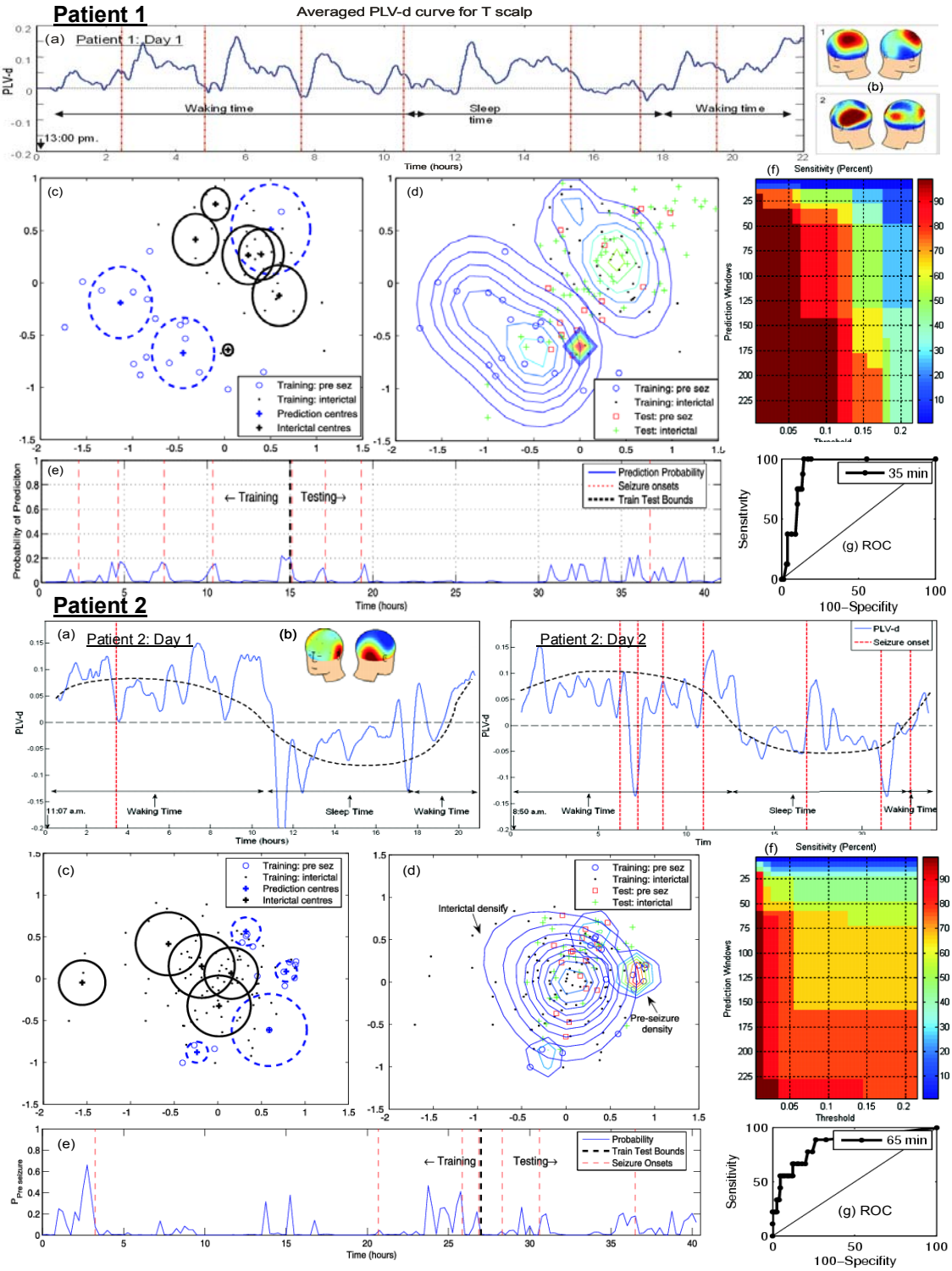

Figure 1: (a) The PLV-d curve of both patients show a cyclical pattern of long term synchrony (20 hours each). The vertical lines mark the seizure onset times, and the sleep and wake times are marked from the data annotations. Prominent peaks can be seen before seizure onsets indicating a set-reset cycle [2]. Contralateral spatial constraints used for spatially constrained ICA are shown in (b). 2-D maps obtained by Neuroscale from the multi-dimensional feature data are shown in (c). (d) Shows the GMM probability contours for training pre-ictal and interictal feature data sets along with the projected test feature data. Resulting probability vs. time curve for the pre-ictal events is shown in (e) (training-test boundary marked). Sensitivity of the occurrence of a predictive event is shown in (f) for various threshold levels and prediction windows (0-4 hours in steps of 5 minutes) imposed on 2-D probability maps. ROC curves (g) show a predictive sensitivity of 100% and specificity of 80% for predictive window of 35 minutes (patient 1) and 80% sensitivity and 80% specificity for a prediction window of 65 minutes (patient 2).