34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Automatic Diagnosis of ADHD based on Multichannel Nonlinear Analysis of Actimetry Registries

Pablo Casaseca-de-la-Higuera*, Diego Martín-Martínez, Susana Alberola-López, Jesús María Andrés-de-Llano, José Antonio López-Villalobos, José Ramón-Garmendia Leiza, Carlos Alberola-López

***Abstract*— Attention-Deficit Hyperactivity Disorder (ADHD) is the most common mental health problem in childhood and adolescence. It is commonly diagnosed by means of subjective methods which tend to overestimate the severity of the pathology. A number of objective methods also exist, but they are either expensive or time-consuming. Some recent proposals based on nonlinear processing of activity registries have deserved special attention. Since they rely on actigraphy measurements, they are both inexpensive and non-invasive. Among these methods, those shown to have higher reliability are based on single-channel complexity assessment of the activity patterns. This way, potentially useful information related to the interaction between the different channels is discarded. In this paper we propose a new methodology for ADHD diagnosis based on joint complexity assessment of multichannel activity registries. Results on real data show that the proposed method constitute a useful diagnostic aid tool reaching** 87.10% **sensitivity and** 84.38% **specificity. The combination of ADHD indicators extracted with the proposed method with single-channel complexity-based indices previously proposed lead to sensitivity and specifity values above** 90%.

***Index Terms*— ADHD, Actimetry, activity/rest analysis, Central Tendency Measure, Multichannel Processing.**

I. INTRODUCTION

Attention-deficit/hyperactivity disorder (ADHD) is the most common neurobehavioral disorder in the school age population [1]. However, even though the treatment of ADHD is well-defined for the known types —inattentive, hyperactive-impulsivity and combined—, a simple standard diagnostic method does not exist, mainly due to the fact that its etiology is not completely understood [2]. Current diagnostic methods can be broadly classified in two categories, namely, subjective and objective methods. Both methods have considerable drawbacks: subjective methods depend on the observer and need many people involved (parents/relatives, teacher/fellow worker, etc.) whereas objective methods are either too expensive or not reliable enough.

Considering objective diagnosis, both activity and sleep patterns have been studied by means of actimetric analysis [3], [4]. Actimetry registries constitute a simple and economic alternative for the objective diagnosis of ADHD. Most of the actigraphy-based studies reporting differences between ADHD diagnosed and healthy children are focused in the observation of specific amplitude patterns in the sleep registries. In order to extract meaningful information from these rest epochs, logs of, at least, 7 days duration are required [5]. Additionally, the fact of considering only rest (sleep) epochs involves discarding meaningful information related to normal activity.

In [6], [7] we proposed an inexpensive method to objectively detect the combined type of ADHD in which only the involvement of the patient is needed and it is as simple as wearing a small device that does not prevent the patient from carrying out usual habits during the test. The method is based on the nonlinear analysis of 24-hour-long actigraphic registries to characterize both activity and sleep patterns. These data sets are obtained with actimeters wrapped on the wrist of the dominant hand of the patient; these devices provide motion signals in each of the three spatial coordinates (x,y,z). Our method outperformed those so far proposed also relying on actigraphy registries. The obtained results were in line with those reported in the literature for more expensive and less convenient objective methods either based on polysomnography (PSG) or Magnetic Resonance Imaging (MRI).

The analysis performed in [6], [7] did not focus on the extraction of information related to the interaction between the acquired channels. That is, the ADHD-informative indices were individually extracted from each actigraphic channel (x,y,z and global motion $r=\sqrt{x^2+y^2+z^2}$) and further combined to achieve higher diagnostic capability. In this paper, we focus on the evaluation of new indices based on the joint evolution of the acquired vector signal. By simultaneously processing the multichannel signal, additional information based on cross-channel interaction can be obtained. By combining this new information with the one provided by single-channel-based indices, further improvements in the diagnostic capability can be achieved.

Manuscript received March 16, 2012; revised June 4, 2012. This work was partially supported by the Ministerio de Ciencia e Innovación and the Fondo Europeo de Desarrollo Regional (FEDER) under Research Grant TEC2010–17982 and by the Centro para el Desarrollo Tecnológico Industrial (CDTI) under the cvREMOD (CEN–20091044) project. The work was also funded by the Junta de Castilla y León under Grants VA376A11–2 and VA039A10–2. *Asterisk indicates corresponding author.*

Laboratory of Image Processing (LPI), E.T.S. Ingenieros de Telecomunicación, University of Valladolid. Paseo Bel'en 15, 47011, Valladolid, Spain (email: casaseca@lpi.tel.uva.es, dmarmar@lpi.tel.uva.es, salberola@ono.com, jmandres@ono.com, villalobos@cop.es, garbi69@garbi6.jazztel.es, caralb@tel.uva.es).

II. MATERIALS AND METHODS

A. Subjects and Data

The study is conducted as a case-control analysis, where the two groups respectively consist of 31 and 32 6-year-old children. Children included in the case group were diagnosed as having the combined kind of ADHD according to the DSM-IV criteria [8] and none of them suffered any type of sleep disorder. The Helsinki protocol has been followed. Registries were acquired with the ActiGraph GT3x (Actigraph Inc. Pensacola, FL, USA) device wrapped on the subject's dominant wrist at $f_s = 1$ sample per second during 24 hours.The activity measurements (x,y,z) from the three channels were recorded for processing. We will refer to the vector signal as $\mathbf{x}[n] = [x_x[n], x_y[n], x_z[n]]^T, n = 0,\ldots,N-1$, with N, the length of the acquired time series. We additionally consider the vector time series $\mathbf{x}_{sph}[n] = [x_r[n], x_\theta[n], x_\phi[n]]^T$ obtained by transforming $\mathbf{x}[n]$ to spherical coordinates:

$$\begin{bmatrix} x_r[n] \\ x_\theta[n] \\ x_\phi[n] \end{bmatrix} = \begin{bmatrix} \sqrt{x_x^2[n]+x_y^2[n]+x_z^2[n]} \\ \arccos\left(\frac{x_z[n]}{\sqrt{x_x^2[n]+x_y^2[n]+x_z^2[n]}}\right) \\ \arctan\left(\frac{x_y[n]}{x_x[n]}\right) \end{bmatrix}. \quad (1)$$

B. Methods

1) Preprocessing: The preprocessing stage consists of the two following procedures:

a) Identification of Activity and Rest Intervals: The performed analysis is three-fold, i.e., we have independently analyzed the complete signal, as well as both the activity and the rest intervals. To that end, we define the rest interval as the interval of time spent in bed at night time; the remaining time is considered as the activity time. The identification of both intervals has been performed automatically with the method proposed by the authors in [6].

b) Registries Decimation: Decimation of the registries is a necessary stage for the enhancement of the activity-related information hidden by the large amount of zero-valued samples found in the original signal. Authors of sleep scoring algorithms [9], [10], which dealt with the same problem, solved it by sample accumulation. We propose to create a new signal consisting of a sequence of averaged intervals as follows:

$$\mathbf{x}_{dec}[k] = \frac{1}{M}\sum_{i=0}^{M-1}\mathbf{x}[kM+i], 0 \le k \le \left\lfloor \frac{N}{M} \right\rfloor \quad (2)$$

with N the overall number of samples. Each sample of $\mathbf{x}_{dec}$ summarizes information of M samples of $\mathbf{x}[n]$, i.e., samples of $\mathbf{x}_{dec}$ are epochs of M/f_s seconds. Parameter M has been chosen from the five options (15, 30 seconds, 1,5 and 15 minutes). This choice lets us deal with short epochs (of the order of seconds) that represent the activity of movements and large epochs (several minutes) that represent the activity of tasks.

2) Central Tendency Measure for Vector Signals: The Central Tendency Measure (CTM) is a non-linear magnitude which quantifies the complexity observed in a second-order difference plot constructed from a time series. In order to compute it over vector signals, we first approximate the tangent vector of the trajectory in the phase space as [11]:

$$\mathbf{y}[k] = \mathbf{x}_{dec}[k+1] - \mathbf{x}_{dec}[k] \quad (3)$$

The angle between the tangent vectors can be expressed by its cosine value, obtained from the inner product definition:

$$A[k] = \frac{< \mathbf{y}[k+1], \mathbf{y}[k] >}{||\mathbf{y}[k+1]|| \cdot ||\mathbf{y}[k]||} \quad (4)$$

where $< \cdot,\cdot >$ stands for the inner product, and $||\cdot||$ is the ℓ^2 norm of the vector. Compared to the angle itself, the cosine value presents higher robustness to noise and artifacts. The second order difference plot express the change rate of the tangent vectors angle. The diagram is constructed by employing $A[k+2]-A[k+1]$ and $A[k+1]-A[k]$ as axes. According to [12], we compute the CTM over the series of cosine values by selecting a circular region of radius ρ, around the origin, counting the number of points that fall within the region, and dividing by the total number of points:

$$CTM = \frac{1}{\left\lfloor \frac{N}{M} \right\rfloor - 4} \sum_{k=0}^{\left\lfloor \frac{N}{M} \right\rfloor - 5} \delta_k, \quad (5)$$

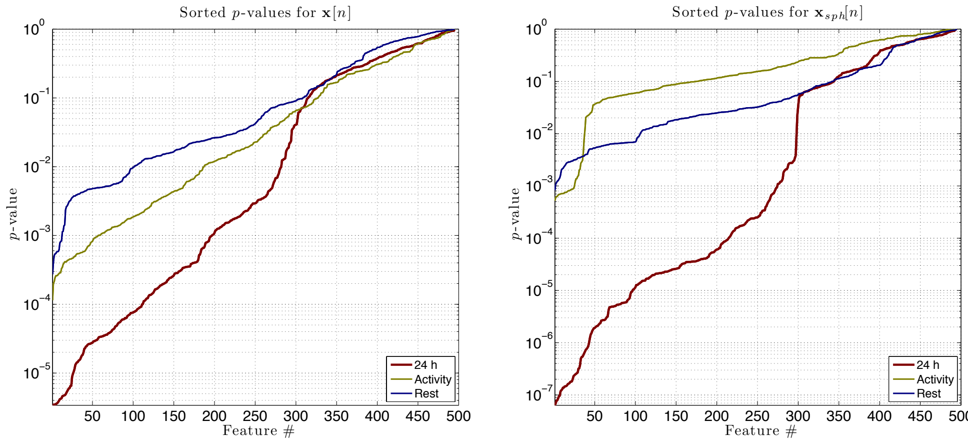
where

$$\delta_k = \begin{cases} 1 & \text{if } \sqrt{\begin{matrix}(A[k+2]-A[k+1])^2 + \\ (A[k+1]-A[k])^2\end{matrix}} \le \rho \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

Higher CTM values are obtained from trajectories with soft changes (regular movements), while movements with high directional variability yield lower values of this magnitude. Several values of ρ are considered in our study. Specifically, we have evaluated 100 different values for ρ within its dynamic range.

3) Statistical Analysis, Feature Selection and Classifier Assessment: The statistical analysis consists of a separability analysis of the two involved groups. This analysis is performed by means of either the Student's t test (with penalty when the F test of Snedecor-Fisher gives that variances of data are not equal) if data are Gaussian, or the U test of Mann-Whitney otherwise.

For the evaluation of the classifier performance, we have used ROC (*Receiver Operating Characteristic*) curves [13] with the Fisher's linear discriminant and the *Leave-One-Out* strategy to avoid dependence between the training and the test datasets [14]. Sensitivity, specificity and accuracy —numerically defined as (sensitivity)(prevalence)+(specificity)(1-prevalence)— have also been obtained for the optimal decision threshold. The best classifier has been defined as the one with the highest area under the [ROC] curve (AUC); if, to our working precision, equality exists in two AUCs, the best classifier



(a) (b)

Fig. 1. Results of the separability analysis summarized as a sorted p-values plot. (a) $\mathbf{x}[n]$, (b) $\mathbf{x}_{sph}[n]$.

will be that with higher accuracy. The best features will be those leading to the best classifiers.

In order to improve the performance of the classifier, we have combined several features [15]. Having in mind the "dimensionality curse", Rauys et *al.* proposed that the length of the feature vector should be, at the most, one order of magnitude less than the number of samples per class [16]. In this work we have 31 cases and 32 controls, therefore the length L of the feature vectors must comply with $L \leq 3$. Since the *brute-force* evaluation of every combination of all the features used does not seem sensible, we have used the following multi-stage selection strategy instead: for each interval (24 h registry, activity or rest epochs), we have selected the most discriminant features corresponding to each of the five decimation strategies (15, 30 seconds, 1, 5 and 15 minutes). This has been performed both for the CTM values computed from the acquired $\mathbf{x}[n]$ signal and the transformed $\mathbf{x}_{sph}[n]$. Starting from this initial set containing $2\times3\times5 = 30$ features, a sequential forward selection (SFS) process [14] is carried out to yield 2D and 3D feature vectors as outputs.

III. Results

Results of the separability analysis for all features are summarized in figure 1. In this figure, the p-values obtained from hypothesis testing have been arranged in ascending order. This sorting allows to identify how many of the extracted features lead to significative differences for each of the analyzed signal intervals (24 hours, activity and rest). For instance, if significative differences are considered for $p < 0.05$, 300 out of 500 CTM values[1] lead to significative differences when the whole registry of $\mathbf{x}[n]$ is analyzed. As for the case of $\mathbf{x}_{sph}[n]$ we find that also 300 CTM values are significative but with lower p-values (note that the most significative feature yields $p < 10^{-7}$ in this case, while for $\mathbf{x}[n]$ we have $p < 4\cdot10^{-5}$). If we consider the isolated activity and rest intervals, we have that the number of features leading to significative results is substantially lower for both $\mathbf{x}[n]$ and $\mathbf{x}_{sph}[n]$, also yielding higher p-values.

In table I we summarize the classification performance of the most informative features extracted from both the original $\mathbf{x}[n]$ and the transformed $\mathbf{x}_{sph}[n]$ vector signals. Since a high number of informative features has been obtained, only the best performing for each analyzed time interval (24 h, activity and rest) have been presented. Index values are presented as Mean $\pm$ Std (if Gaussianity) or as [Median ; IQR] (otherwise). p-values have been respectively obtained from the Student's t or the Mann-Whitney's U Tests. The best performing individual features have been obtained from the analysis of the whole registry (24 h). When either the 24 h or the activity intervals are analyzed, lower CTM values in the case group reveal more complex activity patterns (trajectories) in those children diagnosed with ADHD. As for the rest interval, the obtained CTM values are lower for the control group. This result however, should be carefully interpreted. Note that for the rest intervals, both CTM values are very close to 1, which reveal very soft trajectories in both groups. Differences can be explained by considering transitions from periods with no activity to specific isolated movements. This transitions can lead to abrupt changes in the trajectory which directly translate to CTM.

The classification performance of the best 2D and 3D classifiers constructed after feature selection is also presented in table I. A substantial improvement in both AUC and accuracy has been achieved by combining different features. Combining the best performing individual feature (F_4, Acc.= 0.8095, AUC= 0.8614), with F_3 leads to values of Acc.= 0.8413 and AUC=0.9027. The addition of F_2 leads to the highest diagnostic capability: Acc.= 0.8571, AUC= 0.9335. The separate analysis of the activity and rest intervals turns to be very useful at this point. It is clear that the best individual features are those obtained from the analysis of the whole registry. However, the best multidimensional classifiers are achieved after including features extracted from the isolated activity and rest intervals which individually presented

[1]Note that we have evaluated 100 ρ values for each of the five decimation strategies.

TABLE I

SEPARABILITY ANALYSIS AND CLASSIFICATION PERFORMANCE OF THE MOST INFORMATIVE FEATURES EXTRACTED FROM BOTH THE ORIGINAL $\mathbf{x}[n]$ AND THE TRANSFORMED $\mathbf{x}_{sph}[n]$ VECTOR SIGNALS CONSIDERING THE SEPARATE ANALYSIS OF THE 24 HOURS, ACTIVITY AND REST INTERVALS. RESULTS FOR 2D AND 3D CLASSIFIERS ARE ALSO PRESENTED.

Individual features											
Signal	Interval	*EL*	Feat.	ρ	Case group	Control group	*p*-value	Sens.	Spec.	Acc	AUC
$\mathbf{x}[n]$	24 hours	30 s	F_1	1.0857	0.6201 ± 0.0204	0.6485 ± 0.0238	3.96E-06	0.9032	0.625	0.7619	0.8211
	only activity	1 min	F_2	1.4571	0.5331 ± 0.0257	0.5597 ± 0.0255	1.14E-04	0.4839	**0.9375**	0.7143	0.7424
	only rest	30 s	F_3	2.5713	0.9984 ± 0.0016	0.9963 ± 0.0025	2.72E-04	0.9355	0.4375	0.6825	0.7248
$\mathbf{x}_{sph}[n]$	24 hours	30 s	F_4	1.5428	0.5513 ± 0.0211	0.5892 ± 0.0274	6.92E-08	0.7742	0.8438	0.8095	0.8614
	only activity	15 s	F_5	0.3714	0.2418 ± 0.0168	0.2598 ± 0.0231	8.12E-04	0.8387	0.5	0.6667	0.7117
	only rest	5 min	F_6	2.7999	0.9974 ± 0.0079	0.9873 ± 0.0139	8.29E-04	0.871	0.5938	0.7302	0.7308

Multidimensional classifiers					
Dimension	Feat.	Sens.	Spec.	Acc	AUC
2D	$F_4 + F_3$	0.7742	0.9062	0.8413	0.9027
3D	$F_4 + F_3 + F_2$	0.8710	0.8438	0.8571	0.9335
2D	Best combination of single- and multichannel features	0.9032	0.8438	0.8730	0.9219
3D	Best combination of single- and multichannel features	**0.9355**	0.9062	**0.9206**	**0.9516**

Keys: *EL*: Epoch Length; Feat.: Key for feature identification; ρ: value of the parameter for CTM calculation; Sens.: Sensitivity; Spec: Specificity; Acc.: Accuracy; AUC: Area Under the [ROC] Curve.

poorer performance.

Finally, the last two rows in table I summarize the classification performance of the best 2D of 3D classifiers obtained by combining single- and multichannel complexity-based features. Feature selection in these cases was performed by SFS from an initial set composed of the 30 initial features employed in our study and the initial set used in [7]. A performance improvement has been achieved if we compare with the results obtained from our single-channel proposal in [7] (Acc.=0.9048, AUC=0.9496 in the best case) and the multichannel analysis here proposed.

IV. CONCLUSIONS

We have introduced a novel method based on multichannel complexity assessment of 24 h actimetry registries which constitutes a low-cost objective diagnosis tool for ADHD in children. Experimental results have shown that the nonlinear analysis of the directional variability in the activity channels can provide meaningful indices for ADHD diagnostic and follow up. By incorporating features accounting for the directional complexity of separate activity and rest intervals, the diagnostic capability of 24 h-based indices can be substantially improved. The proposed method not only constitutes a useful tool for diagnostic aid, but it also paves the way to complexity-based interpretations of the pathology, providing additional information to the diagnostic procedure.

REFERENCES

[1] L. Adler, T. Spencer, M. Stein, and J. Newcorn, "Best Practices in Adult ADHD: Epidemioólogy, Impairments and Differential Diagnosis," *CNS Spectr.*, vol. 13, pp. 4–5, 2008.

[2] B. Casey, J. Epstein, J. Buhle, C. Liston, M. Davidson, S. Tonev, J. Spicer, S. Niogi, A. Millner, A. Reiss, A. Garrett, S. Hinshaw, L. Greenhill, K. Shafritz, A. Vitolo, L. Kotler, M. Jarrett, and G. Glover, "Frontostriatal Connectivity and Its Role in Cognitive Control in Parent-Child Dyads with ADHD," *Am. J. Psychiatry*, vol. 164, pp. 1729–1736, 2007.

[3] A. Dane, R. Schachar, and R. Tannok, "Does Actigraphy Differentiate ADHD Subtypes in a Clinical Research Setting?" *J. Am. Acad. Child. Adolesc. Psychiatry*, vol. 39, pp. 752–760, 2000.

[4] A. Sadeh and C. Acebo, "The Role of Actigraphy in Sleep Medicine," *Sleep Med. Rev.*, vol. 6, pp. 113–124, 2002.

[5] C. Acebo, A. Sadeh, R. Seifer, O. Tzischinsky, A. Wolfson, A. Hafer, and M. Carskadon, "Estimating Sleep Patterns with Activity Monitoring in Clindren and Adolescents: How Many Nights Are Necessary for Reliable Measures?" *Sleep*, vol. 22, pp. 95–103, 1999.

[6] D. Martin, P. Casaseca, S. Alberola, J. Lopez, F. Ruiz, J. Andres, J. Garmendia, and J. Ardura, "Automatic diagnosis of adhd based on nonlinear analysis of actimetry registries," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, may 2011, pp. 685 –688.

[7] D. Martín-Martínez, P. Casaseca-de-la-Higuera, S. Alberola-López, J. Andrés-de-Llano, J. A. López-Villalobos, J. Ardura-Fernández, and C. Alberola-López, "Nonlinear analysis of actigraphic signals for the assessment of the attention-deficit/hyperactivity disorder (ADHD)," *Medical Engineering Physics*, pp. –, 2012, (in press). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S135045331100350X

[8] A. American Psychiatric, *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR*, 4th ed. American Psychiatric Association, 2000.

[9] A. Sadeh, C. Acebo, R. Seifer, S. Aytur, and M. Carskadon, "Activity-Based Assessment of Sleep-Wake Patterns During the 1 st Year of Life," *Infant Behav. Dev.*, vol. 18, pp. 329–337, 1995.

[10] G. Jean-Louis, D. Kripke, W. Mason, J. Elliot, and S. Youngstedt, "Sleep Estimation from Wrist Movement Quantified by Different Actigraphic Modalities," *J. Neurosci. Methods*, vol. 105, pp. 185–191, 2001.

[11] J. Jeong, J. C. Gore, and B. S. Peterson, "A method for determinism in short time series, and its application to stationary EEG," *IEEE Trans. Biomed. Eng.*, vol. 49, pp. 1374–1379, 2002.

[12] M. Cohen, D. Hudson, and P. Deedwania, "Applying Continuous Chaotic Modeling to Cardiac Signal Analysis," *IEEE Eng. Med. Biol. Mag.*, vol. 15, pp. 97 –102, 1996.

[13] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. San Diego, CA, USA: Academic Press, 1999.

[14] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical Pattern Recognition: A Review," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 4–37, 2000.

[15] P. Casaseca-de-la-Higuera, M. Martín-Fernández, and C. Alberola-López, "Weaning from mechanical ventilation: A retrospective analysis leading to a multimodal perspective," *IEEE T. Bio-Med. Eng.*, vol. 53, pp. 1330–1345, 2006.

[16] S. Raudis and A. Jain, "Small Samplw Size Effects in Statistical Pattern Recognition: Recomendations for Practitioners," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 252–264, 1991.