

## Most discriminative atom selection for apnea-hypopnea events detection

R.E. Rolon<sup>1</sup>, L.E. Di Persia<sup>1,2</sup>, H.L. Rufiner<sup>1,2,3</sup> and R.D. Spies<sup>2,4</sup>

<sup>1</sup> Centro de I+D en Señales, Sistemas e Inteligencia Computacional (sinc(i)), Fac. de Ingeniería y Cs. Hídricas, Univ. Nacional del Litoral, Santa Fe, Argentina

<sup>2</sup> Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

<sup>3</sup> Laboratorio de Cibernética, Fac. de Ingeniería, Univ. Nacional de Entre Ríos, Entre Ríos, Argentina

<sup>4</sup> Instituto de Matemática Aplicada del Litoral (IMAL), Santa Fe, Argentina

**Abstract**— The sleep apnea-hypopnea syndrome is characterized by repetitive episodes of upper airway obstruction that occur while sleeping, usually associated with a reduction in blood oxygen saturation (SaO<sub>2</sub>). This work presents a novel most discriminative atom selection method to predict the occurrence of apnea-hypopnea (AH) events. First two types of dictionaries (one using class information and the other without it) are estimated, then a greedy pursuit algorithm is used in order to obtain the activation coefficients. The SHHS polysomnography database which includes nearly 1000 polysomnograms, is used for training and testing. A subset of the most discriminative coefficients is then selected for each dictionary, training a pattern recognition neural network to detect the AH events. Finally these events from a test set of 64 studies with different grades of illness are detected. Correlation coefficients of 0.90 and 0.74 are obtained for the dictionaries trained with and without class information, respectively.

**Keywords**— Sleep apnea-hypopnea syndrome, Sparse representation, Dictionary, Neural network.

### I. INTRODUCTION

The American Academy of Sleep Medicine distinguishes more than 80 different sleep disorders [1]. One of those pathologies is the Obstructive Sleep Apnea-Hypopnea (OSAH) syndrome. This syndrome is characterized by repetitive episodes of airway narrowing or collapse during sleep. The current gold standard diagnostic test for OSAH is an overnight polysomnography in a sleep laboratory which is costly both in terms of time and money, and the accessibility in some areas is very limited. Due to its ease of access and wide availability, pulse oximetry has become a very attractive option for detecting OSAH, since a decrease in blood oxygen saturation (SaO<sub>2</sub>) indicates respiratory problems.

In the last fifteen years, many different approaches to traditional signal processing problems were taken. Some of these new formulations gave rise to techniques based on non-linear systems and higher-order statistics, including Independent Component Analysis (ICA) [2] and methods to obtain a Sparse Representation (SR) [3] of a signal. They provide new ways of phrasing the problems of signal modeling and representation. One underlying idea is that of representing the involved signals using only a few significant characteristics,

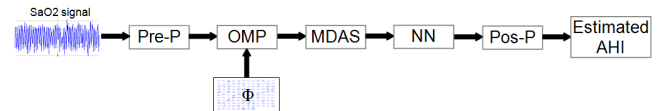


Fig. 1: General diagram proposed for detecting AH events using a dictionary  $\Phi$ .

e.g. as an SR with just a few basic waveforms. In a previous work sparse representation of a SaO<sub>2</sub> signal was obtained by means of an optimal dictionary used in order to detect OSAH [4]. In this work a novel most discriminative atom selection (MDAS) method to detect apnea-hypopnea (AH) events by using a subset of the most discriminative atoms of a dictionary is developed.

In this work we start by comparing the performances of an overcomplete assembled dictionary (OAD), trained using class information and a complete dictionary (CD), trained without class information. Those dictionaries were used as generators of an SR of the SaO<sub>2</sub> signal, preserving as much as possible the morphology of the signal. After that, the most discriminative activation coefficients are selected and used as input of a pattern recognition neural network (NN) in order to detect the respiratory events from the same dataset used in the dictionary learning process. Fig. 1 shows a block diagram of the system proposed for detecting AH events by using a dictionary. In this figure, Pre-P denotes a pre-processing algorithm for denoising and windowing the signal, OMP is a greedy pursuit algorithm for estimating the activation coefficients of the dictionary of each signal frame, MDAS denotes a feature selection method for estimating the most discriminative atoms and the best number of inputs and hidden layers of the NN, which has been previously trained to detect the AH events by the activation coefficients, and the last block is a post processing (Post-P) algorithm for eliminating spurious detections [4]. The next section describes methods used to obtain the dictionary and the coefficients of an SR of a signal. Preliminary classification and correlation results are also presented.

### II. MATERIALS AND METHODS

#### A. Sparse representation of signals

By a dictionary we shall mean a matrix  $\Phi \in \mathbb{R}^{N \times M}$  (with  $M \geq N$ ) whose columns  $\vec{\phi}_j$  are waveforms called atoms. A

representation of a signal  $\vec{s} \in \mathbb{R}^N$  in terms of the dictionary  $\Phi$  is an expression of the form:

$$\vec{s} = \sum_{j=1}^M \vec{\phi}_j a_j = \Phi \vec{a}, \quad (1)$$

where  $\vec{a} = (a_j) \in \mathbb{R}^M$ .

The term “basis” instead of “dictionary” is sometimes used. However, since the atoms are not required to be linearly independent, and quite often more atoms than the space dimension are used, the latter is preferred. When there are more waveforms in the dictionary than samples  $\vec{s}$ , when  $M > N$  (referred to as an overcomplete dictionary), or when the waveforms do not form a basis, then there may be non-unique representations of a given signal. In this situation a suitable criterion is required to select only one of them. In this context, sparseness often refers to the criterion of choosing a representation with “as few non-zero coefficients as possible”, although several other criteria can be used. In particular the problem of an SR of  $\vec{s}$  can be stated as follows:

$$\min \|\vec{a}\|_0 \text{ subject to } \Phi \vec{a} = \vec{s}, \quad (2)$$

where  $\|\cdot\|_0$  denotes the zero-norm.

It is important to note that although the mapping  $\vec{s} \rightarrow \vec{a}$  in the representation (1) is obviously linear, if the dictionary  $\Phi$  is overcomplete, the mapping  $\vec{s} \rightarrow \vec{a}_{SR}$ , where  $\vec{a}_{SR}$  denotes the solution of (2), is not necessary linear. Briefly said, under the sparsity condition (2) the mapping signal-to-coefficients may not be linear. Let us consider now the problem of finding SR of a given family of signals with respect to a fixed dictionary  $\Phi$ , where constructing such a dictionary is part of the problem. Clearly one could build up the dictionary using all the signals in the given family. Although this choice of  $\Phi$  will result in optimal sparsity, most likely it will be highly undesired, mainly because of its size and redundancy. It becomes then necessary to find a dictionary that be optimal, in a certain sense, for a given family.

### B. Learning of the dictionary

A slightly more general framework is assumed, where Eq. (1) is modified to include an additive Gaussian noise  $\vec{\epsilon}$  as follows:

$$\vec{s} = \Phi \vec{a} + \vec{\epsilon}. \quad (3)$$

Following usual ICA terminology, Eq. (3) is referred to as the generative model, meaning that one generates the signal  $\vec{s}$  from a set of hidden sources  $a_j$ , arranged as a state vector  $\vec{a}$ , using a dictionary  $\Phi$ . The sources  $a_j$  are initially assumed to be statistically independent with a joint priori density  $\pi_{\text{prior}}(\vec{a}) = \prod_{j=1}^M \pi(a_j)$ . The atoms of  $\Phi$  can be estimated by maximizing the log-likelihood function of the data given the dictionary [5],  $L(\vec{s}, \Phi) \doteq E[\log \pi(\vec{s}|\Phi)]$ , i.e. as follows:

$$\hat{\Phi} = \underset{\Phi}{\operatorname{argmax}} L(\vec{s}, \Phi). \quad (4)$$

The log-likelihood function can be found by marginalizing the product of the conditional distribution of the data given the dictionary and the prior distribution of the coefficients. That is:  $\pi(\vec{s}|\Phi) = \int_{\mathbb{R}^M} \pi(\vec{s}|\Phi, \vec{a}) \pi_{\text{prior}}(\vec{a}) d\vec{a}$ . The maximum in Eq. (4) can also be approximated by using a gradient ascent method with the updating rule  $\Delta\Phi = \eta \Lambda_{\epsilon}((\vec{s} - \Phi \vec{a}_{MAP}) \vec{a}_{MAP}^T - \Phi H^{-1})$ , [6, 7]. Where  $\vec{a}_{MAP}$  denotes the mode of the posterior distribution of the coefficients, i.e.  $\vec{a}_{MAP} = \underset{\vec{a} \in \mathbb{R}^M}{\operatorname{argmax}} \pi_{\text{post}}(\vec{a}|\Phi, \vec{s})$ .

In order to obtain  $\Phi$  and the vector coefficient  $\vec{a}$ , the implementation proposed by Lewicki and Olshausen [5] was used at the dictionary training stage.

### C. Orthogonal Matching Pursuit

The Orthogonal Matching Pursuit (OMP) algorithm proposes a modification of the greedy Matching Pursuit algorithm (MP) of Mallat and Zhang [8], where the full backward orthogonality of the residual convergence is maintained. It is shown that all additional computation required for the OMP algorithm may be performed recursively [9]. The reason for choosing this algorithm is because it provides a good sparse approximated solution faster than most other methods to solve (2).

Mallat and Zhang show a sequence of approximations used for the MP algorithm in order to obtain an SR of the signal  $\vec{s}$ . It is assumed that  $\vec{a}$  has only  $m$  non-zero components, and therefore the signal vector  $\vec{s} = \Phi \vec{a}$  is a linear combination of  $m$  columns from  $\Phi$ . In the language of SR, it is common to say that “ $\vec{s}$  has an  $m$ -term representation over the dictionary  $\Phi$ ”. To identify one such a vector  $\vec{a}$ , it is necessary to distinguish which columns of  $\Phi$  participate in the measured signal  $\vec{s}$ . The idea of the OMP algorithm basically consists in properly selecting columns of  $\Phi$ . At each iteration, the column most correlated with the current residual of  $\vec{s}$  is taken. Then the residual is updated and iterated. In this way, after  $m$  iterations the algorithm will choose a set of  $m$  columns of  $\Phi$  [10].

### D. Most Discriminative Atoms Selection

Given a fixed dictionary, most discriminative atoms selection (MDAS) is a novel method developed in order to improve the NN performance in Fig. 1. The idea behind this method is to select the most discriminative atoms of this dictionary. Then each activation coefficient (obtained applying MDAS method) which corresponds to each discriminative atom is taken into account as input of the NN.

The feature selection is obtained by computing the atom activation frequency given the class here being (with and without AH). The candidates to be considered as input of the NN are then those atoms with higher absolute difference between frequency activation for each of the classes. That is, if some atom is active many times for signals with AH events than for the signals without AH events, it is taken into account.

### E. Neural Network Classifier

This work is focused on the development of a novel method to detect AH events. In this work a multilayer neural network for pattern classification is used, this type of classifier is the most popular NN and is used in many engineering fields.

Two layer feed-forward pattern recognition network, with sigmoid hidden and an output neuron was used. A scaled conjugate gradient algorithm<sup>1</sup> was used to train the NN with the most discriminative activation coefficients as input and the AH events as target. Here an AH event is an  $n$ -dimensional vector of ones and zeros, where a one is associated to an AH event of the SaO<sub>2</sub> signal, and a zero to the lack of it, respectively.

### F. Database

The Sleep Heart Health Study (SHHS) database<sup>2</sup> was used for this work. This database contains exhaustive information about detailed studies which were appropriately designed to investigate the relationships between sleep breathing disorders and cardiovascular diseases. Several biomedical signals are contained in this database and each study has EEG, nasal airflow, respiratory effort and SaO<sub>2</sub> signals. Annotations of sleep stages, arousals and respiratory events (apnea and hypopnea) are also included. The latter are of our interest and are defined as:

- apneas: if the amplitude of the airflow signal decreases below 25% of the “baseline” breathing amplitude (identified during a period of regular breathing with stable oxygen levels) and it remains below that level for more than 10 seconds.
- hypopneas: if the amplitude of the respiratory signal decreases below 70% of the “baseline” breathing amplitude, it remains so for more than 10 seconds for more than 2 breathe periods and the SaO<sub>2</sub> saturation decreases at least by 2% [11].

There exists a time delay from the cessation of the nasal airflow to the onset the oxygen desaturation. Commonly this delay is of about 20 seconds [12]. Also, the desaturation and the flow reduction have different durations. In this work the AH events are detected by using only the SaO<sub>2</sub> signal (without the airflow signal). The detected AH events are used to estimate the Apnea Hypopnea Index (AHI), defined as the average number of AH events per hour.

## III. EXPERIMENTS AND RESULTS

The first step in Fig. 1 is pre-processing (Pre-P). For this stage, the wavelet processing technique proposed in [4] was

used in order to denoise the SaO<sub>2</sub> signal, which was sampled at 1Hz and denoised by zeroing the approximation coefficients, at level 8, of the dyadic discrete wavelet transform with mother wavelet Daubechies 2. The effect of this procedure is a highpass filter which eliminates the baseline wander and low frequency noise. In the sequel, the SaO<sub>2</sub> signal will refer to the denoised one. Dataset partitions of 24% for training and 76% for testing were selected. More precisely, for the dictionary training stage, four groups of five studies with AHI values below 5, between 5 and 10, between 10 and 15 and above 15, were considered. The rest of the dataset comprised 64 studies for testing with different degrees of illness. In all cases the same number of frames with and without AH events were randomly selected.

Frames of 128 samples of the SaO<sub>2</sub> signals were selected in order to train the dictionaries. Then the dictionaries were estimated by using the Noise Overcomplete ICA (NOCICA) method [7]. We propose two alternatives, one using class information and the other without class information. For the first case, two CD are estimated for signals with and without AH respectively. Segments of 128 samples for which presence of an AH event is present during the same are named “class 1”. Segments for which there was not an AH event are labeled “class 2”. Next, by joining atoms from both dictionaries, an OAD is constructed. For the other alternative, a new CD representing both classes of signals is estimated.

For both the OAD and CD cases, it was observed that the atoms with higher activation frequencies were mainly the same for both signals with and without AH. Fig. 2 shows, for each atom of the OAD, the absolute value of the difference between the number of activations of such atom in class 1 and its number of activations in class 2. Thus, for the OAD case, the activation of atoms with higher absolute difference of activation were selected as input for the NN. In this way the atoms with high activation frequency represent common information which is not useful for class discrimination. Instead we propose the MDAS method to use the activation of atoms that are more active only for one of the classes. For each atom of the CD, the activation frequency was obtained and the atoms with higher frequency were selected as input of the NN.

The same signal training set used for the dictionary learning stage was used to train the NN. Different NN structures were used, i.e. different number of inputs and different number of neurons in its hidden layer. To determine the optimal number of atoms to use for both cases, we evaluate the NN performance by varying its number of inputs and hidden layers from 2 to 100 and from 10 to 20, respectively. Finally the optimal NN parameters are selected and fixed. It was observed that the maximum fraction of well classified frames was obtained with 24 and 30 inputs for the OAD and CD cases, respectively. A total of 14 neurons in the NN hidden

<sup>1</sup>scg algorithm: <http://www.mathworks.com/help/nnet/ref/trainscg.html>

<sup>2</sup>database: <http://physionet.org/physiobank/database/shhpsgdb/>

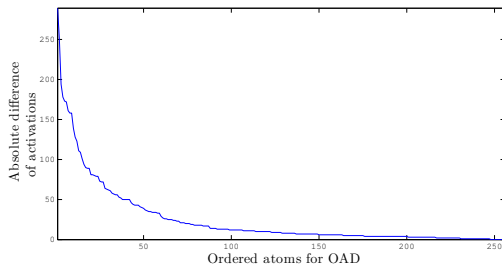


Fig. 2: Absolute difference of the frequency of activation for OAD.

layer was obtained for both cases.

AH events can then be detected from the NN outputs by thresholding them at the value 0.5 (an output value larger than 0.5 is labeled as “1” and as “0” otherwise). Each label “1” is then associated to an AH event. Finally the estimated AHI (AHIest) is defined as the total number of ones divided by the time duration of the study (in seconds). Fig. 3 shows scatter plots and linear regressions between AHIest and AHI for OAD and CD cases.

Sensitivity and specificity measures (SE and SP) are calculated in order to evaluate the performance of the classifier. Table 1 shows the SE and SP values for the AH events detections obtained by applying both MDAS-OAD and MDAS-CD methods, as well as the corresponding AHIest-AHI correlation percentages. Note that SE and SP measures obtained by applying the MDAS-OAD method are reasonable and the corresponding correlation percentage is highly significant.

Table 1: Performance of the proposed method

Procedure	Sensitivity	Specificity	Correlation
MDAS-OAD	74.52%	76.73%	90.04%
MDAS-CD	68.86%	67.69%	74.57%

#### IV. DISCUSSION AND CONCLUSIONS

CD and OAD dictionaries were used to obtain sparse representations of an SaO<sub>2</sub> signal. The OMP algorithm was used to obtain the sparse coefficient vectors and a neural network was constructed to detect the AH events. As observed in Table 1 a considerably high AHIest-AHI correlation was obtained by applying the MDAS-OAD methodology, which constitutes a strong evidence that such a procedure can be successfully used for detecting OSAH.

In the SHHS database used for this work, the AH events are marked at the nasal airflow signal. However, it is well known that the detection of AH events in this way is both very complex and costly. For future work we propose using directly the SaO<sub>2</sub> signal alone by appropriate synchronizing it with the AH events. It is reasonable to suppose that this synchronization will significantly reduce the number of atoms containing discriminative information. Clinical implications

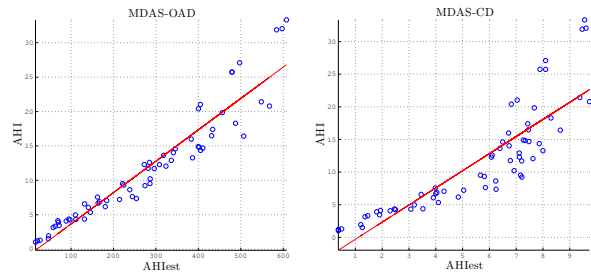


Fig. 3: AHIest-AHI scatter plots and regression lines for the OAD (left) and CD (right).

of the results will also be considered.

#### V. ACKNOWLEDGEMENTS

The authors are grateful to M.D. Luis Larrateguy for many valuable comments and suggestions.

#### REFERENCES

1. Thorpy M. J. *The international classification of sleep disorders, Revised. Diagnostic and coding manual.* American Academy of Sleep Medicine. 3 ed. 2001.
2. Hyvärinen A. Survey on independent component analysis. *Neural Computing Surveys.* 1999;2:94-128.
3. Olshausen B.A., Field D.J. Vision and the coding of natural images. *American Scientist.* 2000;88:238-245.
4. Lestussi F.E., Rufiner H.L., Di Persia L.E., Milone D.H. Sparse coding for apnea-hypopnea detection. in *XIV Reunión de Trabajo en Proceso de la Información y Control (RPIC):1-6* 2011.
5. Lewicki M.S., Olshausen B.A. A probabilistic framework for the adaptation and comparison of image codes. *Optical Society of America.* 1999;16:1587-1601.
6. Abdallah S.A.. *Towards music perception by redundancy reduction and unsupervised learning in probabilistic models.* Ph.D. Thesis, Department of Electronic Engineering, King's College London. 2002.
7. Lewicki M.S., Sejnowski T.J. Learning overcomplete representations. *Advances in Neural Information Processing 10 (Proceedings NIPS'97).* 1998:556-562.
8. Mallat S.G., Zhang Z. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing.* 1993;41:3397-3415.
9. Pati Y.C., Rezaiifar R., Krishnaprasad P.S. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. *IEEE Asilomar Conference on Signals, Systems and Computers.* 1993;1:40-44.
10. Tropp J.A., Gilbert A.C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory.* 2007;53:4655-4666.
11. Arnold J. *Reading center manual of operations: scoring procedures.* Sleep Heart Health Study. 2002.
12. Alice S.L., Thomas K.S. Wong, Gohel M.D.I., Winnie W.M. Yu, Joanne W.Y. Chung, Fan K.L. Using pulse oximetry level to indicate the occurrence of sleep apnoea events. 2006;122:672-675.

Author: Rolon, Román Emanuel  
 Institute: Centro de I+D en Señales, Sistemas e Inteligencia Computacional  
 Street: Ciudad Universitaria – CC 217 Ruta Nac. No 168, km 472.4  
 City: Santa Fe  
 Country: Argentina  
 Email: romanrolon@gmail.com