Oxygen saturation regularity analysis in the diagnosis of obstructive sleep apnea

Félix del Campo a,*, Roberto Hornerob, Carlos Zamarrón c, Daniel E. Abasolob, Daniel Álvarez b

a Hospital del Río Hortega, Servicio de Neumología, C/Cardenal Torquemada s/n, 47010 Valladolid, Spain
b ETSI-Telecomunicación University of Valladolid, Camino del Cementerio s/n, 47011 Valladolid, Spain
c Hospital Clínico Universitario, Servicio de Neumología, C/Travesia de La Choupana s/n, 15706 Santiago de Compostela, Spain

Received 1 June 2005; received in revised form 10 October 2005; accepted 15 October 2005

KEYWORDS
Approximate entropy; Obstructive sleep apnea; Oximetry

Summary

Objective: The present study assessed the validity of approximate entropy (ApEn) analysis of arterial oxygen saturation (SaO2) data obtained from pulse oximetric recordings as a diagnostic test for obstructive sleep apnea (OSA) in patients clinically suspected of suffering this disease.

Methodology: A sample of 187 referred outpatients, clinically suspected of having OSA, was studied using nocturnal pulse oximetric recording performed simultaneously with complete polysomnography. ApEn analysis was applied to SaO2 data.

Results: Patients with OSA presented significantly higher approximate entropy levels than those without OSA (1.08 ± 0.30 versus 0.47 ± 0.26). Apnea–hypopnea index was correlated significantly with ApEn ($r = 0.607; p < 0.001$). Using receiver operating characteristic curve analysis, we obtained a diagnostic sensitivity of 88.3% and specificity of 82.9%, positive predictive value of 88.3% and a negative predictive value of 82.9%, at a threshold of 0.679. As a diagnostic test, this method presents high sensitivity and specificity compared to traditional methods in the diagnosis of OSA.

Conclusion: We conclude that ApEn analysis of SaO2 data obtained from pulse oximetric recordings could be useful as a diagnostic technique for OSA subjects.

© 2005 Elsevier B.V. All rights reserved.
1. Introduction

Obstructive sleep apnea (OSA) is a common clinical entity characterized by recurrent airflow obstruction as a result of total or partial collapse of the upper airway. Given the high prevalence of OSA, its potential importance as a contributing factor to cardiovascular morbidity [1], and the availability of effective treatment [2,3], numerous efforts have been undertaken to pre-select subjects before performing polysomnography.

OSA is frequently accompanied by cyclical drops in oxygen saturation. Thus, nocturnal oximetry may be useful in detecting OSA. Several studies have analyzed nocturnal oximetry from different perspectives [4]: visual scoring, oxygen desaturation index, cumulative time arterial oxygen saturation (SaO2) spent below 90%, \( \Delta \) index and spectral analysis. However, the reported diagnostic capability of oximetry varies widely, with sensitivity and specificity ranging from 31% to 98% and 41% to 100%, respectively [5].

Recently, new methods based on chaos theory, such as approximate entropy (ApEn), have been widely applied in clinical cardiovascular studies [6,7] but their diagnostic accuracy in OSA has yet to be studied. ApEn is a family of parameters and statistics introduced as a quantification of data regularity [8—11]. It assigns a non-negative number to a time series, such as SaO2, where larger entropy values correspond to greater apparent randomness or irregularity, whereas smaller values correspond to more data regularity.

Motivated by the above idea, in the present study, ApEn was applied to nocturnal SaO2 as a possible diagnostic tool for OSA syndrome.

2. Patients and methods

2.1. Subjects

We studied 187 subjects (147 men and 40 females) who were referred for clinical suspicion of OSA. The patients were consecutively recruited from the outpatient clinic. Subjects ranged in age from 21 to 81 years, with an average age of 57 years. The mean body mass index (BMI) was 29.5 kg/m². The study included 42 (22.5%) patients with chronic obstructive pulmonary disease (COPD). The Review Board on Human Studies at our institution approved the protocol, and each patient gave his or her informed consent to participate in the study.

2.2. Polysomnography

All sleep studies were carried out in our Sleep Unit, usually from midnight to 8 a.m. Patients were prospectively evaluated after a single-night pulse oximetry recording obtained by nocturnal pulse oximetry in conjunction with a simultaneous polysomnographic study. This technique consisted of continuous monitoring using a polygraph (Ultrasom Network, Nicolet, Madison, WI, USA) and included electroencephalogram, electro-oculogram, chin electromyogram, air flow (three-port thermistor), electrocardiogram and measurement of chest wall movement.

The polysomnographic register was analyzed in periods of 30 s and during stages 1—4 and REM according to the Rechtschaffen and Kales method [12]. Rechtschaffen and Kales staging is the standard methodology used to identify the stages of normal human sleep. Apnea was defined as the absence of airflow for more than 10 s [13], and hypopnea as the reduction of respiratory flow /% accompanied by a 3% or more decrease in the saturation of hemoglobin [14]. The average of apnea—hypopnea index (AHI) was calculated in hourly samples of sleep. In this study an AHI of 10 or more was considered as diagnostic of OSA. If the subject had less than 3 h of total sleep, the sleep study was repeated [15].

SaO2 recordings were done with a Criticare 504 oximeter (CSI, Wankeska, WI) with a finger—probe and sampled at a frequency of 0.2 Hz (one sample every 5 s). The SaO2 signal was analyzed and the following indices were calculated: minimal SaO2, mean SaO2, time with SaO2 < 90% (CT90) and oxygen desaturation index per hour of recording (ODI). The computer calculated the number of SaO2 dips of 4% h⁻¹ or more (ODI4), 3% h⁻¹ or more (ODI3) and 2% h⁻¹ or more (ODI2) from baseline. Baseline was set initially as the mean level in the first 3 min of recording [16]. Typical recordings of SaO2 patterns comparing non-OSA and OSA patients are shown in Fig. 1.

2.3. Approximate entropy

ApEn was introduced as a quantification of regularity in sequences and time series data, initially motivated by applications to relatively short, noisy data sets. ApEn is basically a "regularity" statistic and should not be taken as a direct index of complexity [8,9]. Mathematically, it is part of a general development of approximating Markov Chains to a process [17]. Furthermore, it provides a finite sequence formulation of randomness, via proximity to maximal irregularity [18,19]. ApEn assigns a non-nega-
ApEn is scale invariant and model independent, evaluates both dominant and subordinant patterns corresponding to more irregularity in the data. It is applicable to systems with at least 50 data points and is finite level noise, is also robust to meaningful information which the aforementioned measures exhibit minimal distinctions [11]. It is nearly unaffected by low level noise, is also robust to meaningful information with a reasonable number of data points and is finite for both stochastic and deterministic processes [21].

It has two user-specified parameters: a run length \(m\) and a tolerance window \(r\). Briefly, ApEn measures the logarithmic likelihood that runs of patterns that are close (within \(r\)) for \(m\) contiguous observations remain close (within the same tolerance width \(r\)) on subsequent incremental comparisons. It is important to consider ApEn\((m, r)\) — or ApEn\((m, r, N)\), where \(m\) is the dimension of the signal will be expanded, \(r\) the threshold and \(N\) is the number of points of the time series — as a family of parameters. Comparisons between time series segments can only be made with the same values of \(m\) and \(r\) [11].

There are two ways to look at ApEn. From one point of view, it is a statistical characteristic (average of logarithm of a conditional probability), which makes it applicable to both deterministic and stochastic processes. From the other point of view, it reflects the rate of new pattern generation and is thus related to the concept of entropy [22].

Formally, given \(N\) data points from a time series \(\{x(n)\} = x(1), x(2), \ldots, x(N)\), to compute the ApEn, one should follow these steps [11]:

1. Form \(m\)-vectors \(X(1), \ldots, X(N - m + 1)\) defined by \(X(i) = \{x(i), x(i + 1), \ldots, x(i + m - 1)\}\), \(i = 1 - N - m + 1\). These vectors represent \(m\) consecutive \(x\) values, commencing with \(i\)th point.
2. Define the distance between \(X(i)\) and \(X(j)\), \(d[X(i), X(j)]\), as the maximum absolute difference between their respective scalar components:
   \[
   d[X(i), X(j)] = \max_{k=1,2,\ldots,m} |x(i + k - 1) - x(j + k - 1)|
   \]
3. For a given \(X(i)\), count the number of \(j\) \((j = 1 - N - m + 1, j \neq i)\) so that \(d[X(i), X(j)] \leq r\), denoted as \(N^m(i)\). Then, for \(i = 1 - N - m + 1\),
   \[
   C^m_r(i) = N^m(i) / (N - m + 1)
   \]
   The \(C^m_r(i)\) values measure within a tolerance \(r\) the regularity, or frequency, of patterns similar to a given one of window length \(m\).
4. Compute the natural logarithm of each \(C^m_r(i)\), and average it over \(i\),
   \[
   \phi^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln C^m_r(i)
   \]
   \(\phi^m(r)\) portrays the average frequency that all the \(m\)-point patterns in the sequence remain close to each other.
5. Increase the dimension to \(m + 1\). Repeat steps (1)–(4) and find \(C^{m+1}_r(i)\) and \(\phi^{m+1}(r)\).
6. Theoretically, the approximate entropy is defined as:
   \[
   \text{ApEn}(m, r) = \lim_{N \to \infty} [\phi^m(r) - \phi^{m+1}(r)]
   \]
   In practice, the number of data points \(N\) is finite. We implement this formula by defining the statistic [23]:
   \[
   \text{ApEn}(m, r, N) = \phi^m(r) - \phi^{m+1}(r)
   \]
No guidelines exist for optimizing the \(m\) and \(r\) values. In principle, the accuracy and confidence of the entropy estimate improve as the number of matches of length \(m\) and \(m + 1\) increase. The number of matches can be increased by choosing small \(m\) (short templates) and large \(r\) (wide tolerance). However, there are penalties for criteria that are too relaxed [23]. For smaller \(r\) values, one usually
achieves poor conditional probability estimates, while for larger $r$ values, too much detailed system information is lost. Pincus has suggested parameter values of $m = 1$ and 2 and $r = 0.1, 0.15, 0.2$ and 0.25 times the standard deviation (S.D.) of the original data sequence $\{x(n)\}$ [11]. Normalizing $r$ in this manner gives ApEn a translation and scale invariance, in that it remains unchanged under uniform process magnification, reduction or constant shift to higher or lower values [6]. Moreover, several studies [19,22,24] have demonstrated that these input parameters produce good statistical reproducibility for ApEn for time series of length $n \geq 60$, as considered herein.

SaO$_2$ records corresponding to the 187 subjects under study have a mean data length of 5895 ± 442 samples (mean ± S.D.), which corresponds to 8.19 ± 0.62 h (mean ± S.D.) of register. Every SaO$_2$ signal was first divided into epochs of 200 samples. We determined this epoch length according to the apnea duration (between 25 s and 2 min) in such a way that various disordered respiratory events were included in every epoch. The method was then applied over every epoch, providing various ApEn values for each signal, which in the end were averaged to obtain a single result per subject. The analysis of each signal takes about 1 min, depending on the exact sample length and the computer used. For this study, ApEn was estimated with the widely established parameter values of $m = 1$ and $r = 0.2$ times the standard deviation of the original data sequence.

### 2.4. Statistical analysis

The normal distribution of the variables was verified using the Shaphiro–Wilks’ W test and homogeneity of variances using the Levene’s test. The ApEn results for both groups were compared using one-way analysis of variance. Box plots have also been used to analyze differences in the distribution of the results. Moreover, we selected a threshold to improve the sensitivity/specificity pair according to the receiver operating characteristic (ROC) curves in order to make comparisons of the diagnostic accuracy of the ApEn values [25]. Correlations were investigated using Pearson correlation test. We considered $p < 0.05$ to be statistically significant.

### 3. Results

A diagnosis of OSA was confirmed in 111 (59.3%) out of 187 subjects included in the study. Table 1 presents the anthropometric, AHI and ApEn data. There were no significant differences between the OSA and non-OSA groups in age, sleep duration recordings and percentage of patients with COPD. However, the OSA group presented a significantly greater number of male (84.7% versus 69.7%) and obese patients (53.3% versus 29.5%). Furthermore, OSA patients presented significantly higher ApEn level than those without OSA: 1.089 ± 0.308 (95% CI, 1.02—1.14) versus 0.479 ± 0.261 (95% CI, 0.419—0.538).

A total of 42 patients in our study were found to have COPD. Among these patients with COPD, 22 (52.4%) were diagnosed with OSA. COPD patients with OSA showed significantly higher ApEn level than COPD patients without OSA (1.050 ± 0.33 versus 0.615 ± 0.34).

Fig. 2 shows the correlations between ApEn and the AHI. ApEn was significantly correlated with both the ODI4 ($r = 0.499; p < 0.001$) and the AHI ($r = 0.606; p < 0.0001$). In addition, the BMI ($r = 0.206; p < 0.007$), SaO$_2$ media ($r = 0.352; p < 0.001$) and minimal SaO$_2$ ($r = 0.532; p < 0.001$) were significantly correlated with ApEn. However, age was not correlated with ApEn. Oximetry numerical indices (CT90, ODI4, ODI3 and ODI2) were correlated to AHI, although the relationship of ApEn to AHI was similar (Table 2).

The ROC curve of the diagnostic accuracy of the ApEn values are shown in Fig. 3. Different cut-off levels for ApEn of arterial SaO$_2$ were applied. The

### Table 1 Clinical characteristics and approximate entropy data of the population

<table>
<thead>
<tr>
<th></th>
<th>Non-OSA (n = 76)</th>
<th>OSA (n = 111)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>57.5 ± 12.8 (60)</td>
<td>58.3 ± 12.8 (59.5)</td>
<td>NS</td>
</tr>
<tr>
<td>BMI (kg/m$^2$)</td>
<td>28.4 ± 6.01 (26.6)</td>
<td>30.4 ± 4.19 (30.6)</td>
<td>0.03</td>
</tr>
<tr>
<td>AHI</td>
<td>2.03 ± 2.19 (1)</td>
<td>40 ± 19 (38)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Minimal SaO$_2$</td>
<td>83.4 ± 9.5 (88)</td>
<td>69.7 ± 12.6 (72)</td>
<td>0.009</td>
</tr>
<tr>
<td>Recording time (h)</td>
<td>8.2 ± 0.3 (8)</td>
<td>8.1 ± 0.7 (7.9)</td>
<td>NS</td>
</tr>
<tr>
<td>COPD (%)</td>
<td>19.8</td>
<td>26.3</td>
<td>NS</td>
</tr>
<tr>
<td>Males (%)</td>
<td>69.7</td>
<td>84.7</td>
<td>0.014</td>
</tr>
<tr>
<td>ApEn</td>
<td>0.479 ± 0.261 (0.41)</td>
<td>1.089 ± 0.308 (1.09)</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Values are mean ± S.D.; figures in parenthesis are median.
Best results were obtained using a threshold of 0.679. At this level, sensitivity was 88.3%, specificity 82.9%, positive predictive value (PPV) of 88.3% and a negative predictive value (NPV) of 82.9%, with an area under ROC curve of 0.921.

Using this ApEn value, the number of misclassifications was low, with a total number of 13 false-positive cases and 13 false-negatives (11.7%). False-positive subjects were found to have higher BMI (33.3 ± 6.2 versus 27.7 ± 5.7; p < 0.05), higher AHI (4.03 ± 2.2 versus 1.65 ± 2.1; p < 0.001) and higher COPD prevalence (50% versus 21.9%; p < 0.05). Thirteen OSA patients were false-negative, 23% had COPD, 31% had an AHI between 10 and 15, and 50% more than 60 years of age. Mean AHI of the 13 OSA patients misclassified by ApEn using the cited cut-off criterion was 33.2 ± 22, but one of these patients had an AHI = 76. The mean ApEn value for these patients was 0.521 ± 0.125.

Overall results, likelihood ratios and the areas under the ROC curve for the oximetry numerical indices and ApEn are displayed in Table 3 and Fig. 4, using an AHI cut-off value of 10. As can be seen, only ApEn achieves an acceptable value. If the COPD are not taken into account, the diagnostic accuracy for OSA of ApEn improves and reaches a sensitivity of 88.8%, specificity of 89.3%, PPV of 92.4%, NPV of 84.3% and the area under the ROC curve was 0.955 (95% CI, 0.907—0.982).

### 4. Discussion

The main finding of this study is that patients with OSA had significantly higher approximate entropy of SaO2 than non-OSA patients, indicating that OSA patients reflect greater SaO2 irregularity. Various parameters of the severity of OSA, including ODI4, ODI3 and CT90, were significantly correlated with ApEn.

Based on ROC curves, the ApEn of SaO2 presents high sensitivity and specificity for diagnosis of OSA with best results at a cut-off level of 0.679, using
conventional polysomnography as a gold standard for diagnosis of OSA. As a diagnostic test, this method presents good sensitivity and specificity compared to traditional methods (different desaturation indices and the total time spent with SaO2 below 90%) in the diagnosis of OSA, as assessed by ROC analysis. We also found that excluding patients with COPD from the analysis improves the diagnostic value of ApEn.

In a previous study, Sériès et al. found an OSA sensitivity of 98% for the visual inspection of the temporal oximetry recording, but specificity was only 48% [26]. Levy et al., using a mathematical index to detect changes in SaO2, found a sensitivity of 90% and a specificity of 75% [27]. In a more recent study that used a multiple score index on SaO2 recordings, Olson et al. [28] obtained a sensitivity of 88%, but a low specificity (40%). However, a prospective study, which excluded COPD patients, showed that analysis of temporal oximetry recordings in suspected OSA has a sensitivity of 80% and specificity of 89% [29].

Another study by our own group [30,31] demonstrated that spectral analysis and peak detection in the period 30–70 s of SaO2 or heart rate signal obtained from nocturnal pulse oximetry presented a sensitivity of 90% (84–94) and a specificity of 82% (74–88) for OSA diagnosis which resembles the findings for ApEn.

ApEn is a mathematical tool which was introduced to quantify regularity in sequences and time series data. Series of sequential data arise throughout epidemiology in multiple contexts. Enhanced capabilities to quantify differences among such series would be extremely valuable, since these time series reflect essential biological information. ApEn has been used to quantify the differences in apparent regularity between the heart rate interval time series of aborted sudden infant death syndrome and healthy infants [7], to extract features from electroencephalogram and respiratory recordings of a patient during Cheyne–Stokes respiration [32] and to study the connection between panic disorder and respiration dynamics [33]. Moreover, it has been used to investigate changes in respiratory movement during stages of sleep and to associate such alterations with brain function [34]. However, to the best of our knowledge there has been no previous report assessing the accuracy of ApEn of pulse oximetry in the diagnosis of OSA.

It is known that altered respiratory patterns are not exclusive to OSA. COPD patients with OSA showed significantly higher ApEn level than COPD patients without OSA. Furthermore, patients with COPD may present periodic nocturnal oxygen desaturation perhaps explaining some of the false-positive results which have decreased our specificity results. If COPD patients are taken out of consideration, the specificity of ApEn for OSA diagnosis improves significantly. With respect to the false negatives, we have to take into account that some patients present respiratory events that are not significant enough to be detected [27]. Furthermore, some patients with high BMI present a similar problem. False positives were patients with obesity or COPD.

The physiological explanation of why OSA syndrome results in an increase of ApEn could be the following. In contrast to the efficient control during wakefulness, sleeping states are predisposed to increased levels of disordered and unstable breathing. These instabilities are often higher in patients with sleep apnea syndrome. The underlying cause of this is that the resistance to airflow through the upper airway increases significantly in many adults during sleep; moreover, there are also ventilation changes during various sleep stages.

In our study, the recurrence of apnea events in patients with OSA led to a significant increase in ApEn values of SaO2 signals. ApEn \( (m = 1, r = 0.25 \text{D.}) \) reflects the rate of new pattern generation when the dimension \( m \) decreases from 2 to 1. A larger value of ApEn means that the chance of new pattern generation is greater, so the sequence is

### Table 3 Usefulness of conventional oximetric methods in the diagnosis of OSA: ApEn of SaO2 recording, the different ODI and the total time for which SaO2 is below 90%

<table>
<thead>
<tr>
<th></th>
<th>ApEn</th>
<th>SaO2% &gt;90%</th>
<th>ODI1</th>
<th>ODI2</th>
<th>ODI3</th>
<th>ODI4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>88.3</td>
<td>73</td>
<td>58.7</td>
<td>65.1</td>
<td>66.7</td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td>82.9</td>
<td>75.5</td>
<td>94.3</td>
<td>88.7</td>
<td>84.9</td>
<td></td>
</tr>
<tr>
<td>PPV</td>
<td>88.3</td>
<td>78</td>
<td>92.5</td>
<td>87.2</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>NPV</td>
<td>82.9</td>
<td>70.2</td>
<td>65.8</td>
<td>68.1</td>
<td>68.2</td>
<td></td>
</tr>
<tr>
<td>LR+</td>
<td>5.16</td>
<td>2.98</td>
<td>10.38</td>
<td>5.75</td>
<td>4.42</td>
<td></td>
</tr>
<tr>
<td>LR−</td>
<td>0.14</td>
<td>0.36</td>
<td>0.44</td>
<td>0.39</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td>0.921 (0.873–0.956)</td>
<td>0.774 (0.68–0.84)</td>
<td>0.785 (0.699–0.856)</td>
<td>0.761 (0.673–0.835)</td>
<td>0.743 (0.653–0.819)</td>
<td></td>
</tr>
</tbody>
</table>

SaO2% >90%, percentage of time spent with SaO2 below 90% representing 1% of the total recording; PPV, positive predictive value; NPV, negative predictive value; LR+, likelihood ratio for a positive test; LR−, likelihood ratio for a negative test.
more irregular and vice versa. In OSA, oxygen desaturations associated with apnea events cause fluctuations in SaO2 signal leading to higher ApEn values.

Despite our results, there are always some technical or physiological limitations associated with oximetry for recognizing OSA. For example, poor contact between the probe and the finger due to body movements and bad regional circulation occasionally produce signals resembling multiple falls in oxygen saturation. Mindful of this, we examined our recordings before the analysis to see if they evidenced technical problems. For example, we eliminated from our analysis all data that registered drops to zero. From our experience, we knew these falls were probably due to finger-probe disconnections. Another limitation of oximetry is that it fails to identify non-apneic nocturnal hypoventilation. Also, another limitation of the current study is that we did not use nasal cannula and, thus, were not able to identify subtle breathing events.

In conclusion, we have shown that ApEn of SaO2, a measure sensitive to signal regularity, could be useful as a first approach to the analysis of nocturnal oximetry. We believe that ApEn of nocturnal pulse oximetry data may develop into a new index for the evaluation oximetry during sleep and could be incorporated into the oximeter, making it easy to analyze and to use.

References


