Techniques for noise removal from EEG, EOG and air flow signals in sleep patients

Matthew J. Berryman[a], Sheila Messer[a,b], Andrew Allison[a], and Derek Abbott[a]

[a]Centre for Biomedical Engineering and
School of Electrical and Electronic Engineering,
The University of Adelaide, SA 5005, Australia.
[b]Texas Instruments Incorporated,
141 Stony Circle, Suite 130, MS 4110
Santa Rosa, California 95401, USA

ABSTRACT

Noise is present in the wide variety of signals obtained from sleep patients. This noise comes from a number of sources, from presence of extraneous signals to adjustments in signal amplification and shot noise in the circuits used for data collection. The noise needs to be removed in order to maximize the information gained about the patient using both manual and automatic analysis of the signals. Here we evaluate a number of new techniques for removal of that noise, and the associated problem of separating the original signal sources.

Keywords: EEG, EOG, air flow, sleep, blind source signal separation

1. INTRODUCTION

Electroencephalograph (EEG) and electrooculograph (EOG) measurement techniques provide valuable information on sleep disorders.[1,2] Recent studies looking at memory and learning during sleep have used these techniques as predictors of waking performance.[1,3,4] Comparison of thoracic and abdominal movements associated with breathing can reveal important information about breathing disorders and events such as apneas and hypopneas during sleep.[5–7]

The process by which the EEG and EOG signals are recorded is described by Telpan,[8] but a brief summary is given here. The EEG and EOG signals are recorded by placing electrodes on the patient's scalp. These detect electric potentials generated by the flow of ions in neural cells that set up electric dipoles between the body of the neuron (soma) and the neural branches (apical dendrites). For the data we used, these signals were amplified, then digitized at 250 Hz for the EEG and 50 Hz for the EOG signal. We estimate the mutual information between the second EEG channel, from the left anterior position E1 to just below the opposite ear, with the left EOG channel, from just to the side of the left eye to the position just above the nose between the eyes. The signals are broken down into (typically) 30 second long epochs, these are then classified by a human operator into various stages of sleep and wakefulness.

The main problems with analyzing the EEG and EOG signal are:

- Notch filtering of the the 50 Hz interference ripple from the signals also removes useful information.

- Skin conductances can vary over time in different ways in different locations (however the gel used helps prevent this problem).

- Due to conductances across the skin, the signal received by an electrode is a mixture of the true signals one is trying to measure.

Send correspondence to Derek Abbott
E-mail: dabbott@eleceng.adelaide.edu.au, Telephone: +61 8 8303 5748

The most significant problem is the mixing of signals, this can be reduced using blind signal separation techniques using higher order statistics.[9, 10] The noise can then be removed using wavelet transforms.[11–13] We also consider these techniques for the thoracic and abdominal movements, for which similar problems may arise.[7]

Time and power spectra plots for one of the EOG and EEG sets of data is shown in Figure 1. Those for one of the thoracic and abdominal sets of data are shown in Figure 2.
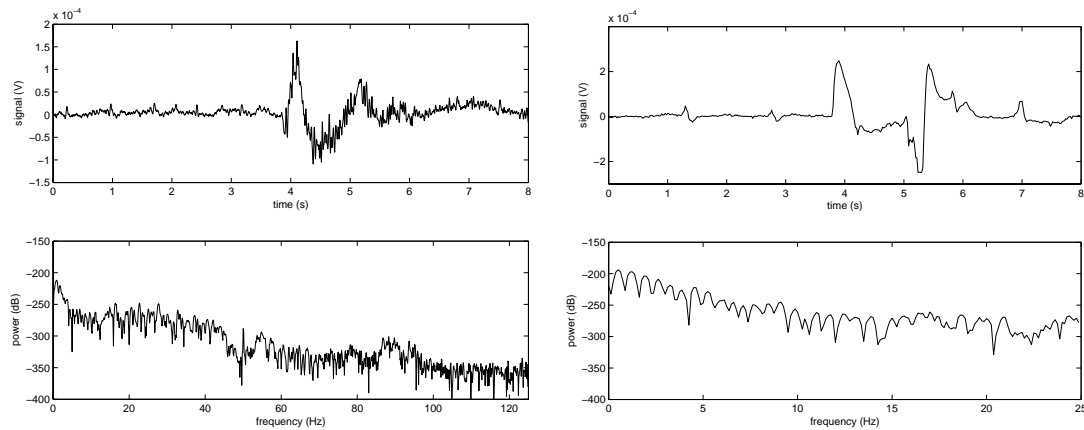
(a) The time and power spectrum plots for the first eight seconds of the EEG data. Note there are many higher frequency signals superimposed on lower frequency signals, giving the appearance of noise, however this is important signal information that needs to be preserved. Note that some of the EOG signal is present on this signal.
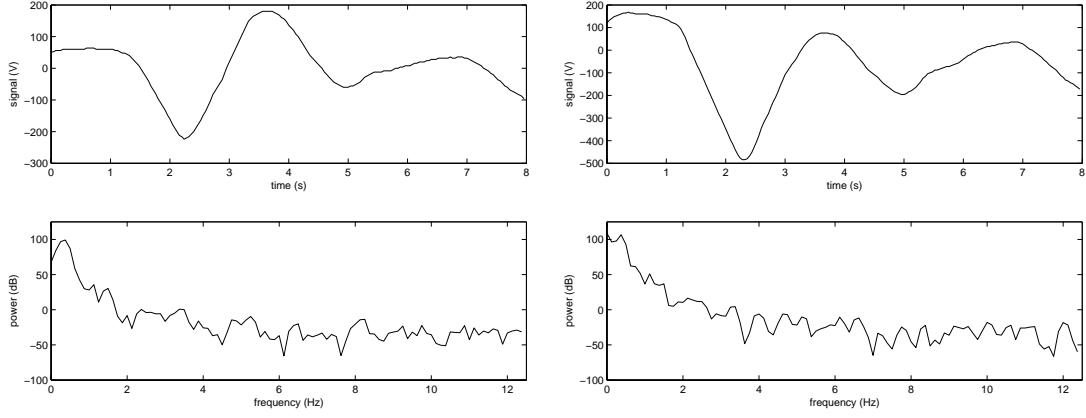
(b) The time and power spectrum plots for the first eight seconds of the EOG data. Note the lack of high frequency signals present in the EEG signal, since we are concerned here with low frequency muscle movement signals. The sampling rate used was correspondingly lower (50 Hz as opposed to 250 Hz for the EEG).

Figure 1. The time and power spectra plots for the first eight seconds of the EEG and EOG data. Note the spectral differences between the two, with the EEG having many higher frequency components.

2. METHODS

There are several problems to solve in eliminating noise from the signals. For the problem of the EOG and EEG signals, we must first make ensure the EOG and EEG signals have the same number of data points. To do this we use a Gaussian smoothing procedure, that is detailed in subsection 2.1. Other related smoothing procedures could be used, here we assume the distribution of the signal data is Gaussian, this is reasonably true for the data we use. We then have the problem of separating the sources from the observed signals, which contain a mixture of both. Three algorithms for this, detailed in subsections 2.4 to 2.6. The noise is removed from both the sources using wavelet transforms as elaborated on in subsection 2.7. A flowchart of the process is shown in Figure 3.

To evaluate the performance of the blind signal separation used to separate the source data, and to evaluate the noise removal we use an efficient algorithm, given in subsection 2.8 to estimate the mutual information between two signals. We expect this measure to decrease when comparing the original signals with the separated signals, assuming greater independence between the separated signals, and to remain the same when comparing the noisy signals with those where the noise has been removed, assuming the noise is uncorrelated

(a) Time and power spectrum plots for the first eight seconds of the thoracic breathing data.

(b) Time and power spectrum plots for the first eight seconds of the abdominal breathing data.

Figure 2. Time and power spectra plots for the first eight seconds of the thoracic and abdominal breathing data. The sampling rate was 25 Hz, allowing the capture of relatively slow breathing signals. Note the thoracic signal has a slight phase lead over the abdominal breathing signal.

between the two signals. This is largely true for the signals of interest, although there may be some information at certain frequencies that is correlated due to extraneous electromagnetic signals being received by the leads, as they act as antennas. This is kept to a minimum through appropriate grounding.

2.1. Gaussian smoothing

The EEG signal has a sampling rate five times higher than the EOG (250 Hz to 50 Hz). These are recorded simultaneously, so every fifth time point in the EEG corresponds to a time point of the EOG signal. We use Gaussian smoothing to reduce the number of data points in the EEG by a factor of five:

$$s\left(i\right) = \frac{\displaystyle\sum_{j=i-m}^{i+m} x_j w\left(x_j, x_i\right)}{\displaystyle\sum_{j=i-m}^{i+m} w\left(x_j, x_i\right)}, \tag{1}$$

where the weights $w\left(x_j, x_i\right)$ are

$$w\left(x_j, x_i\right) = e^{-(x_i-x_j)^2/\left(2\hat{\sigma}^2\right)}, \tag{2}$$

i is the discrete time point we are calculating the smoothed average for, and $\hat{\sigma}^2$ is the estimate of variance for the entire set of samples in the EEG signal. Now that the two sets of data, which can be written as $\mathbf{x}\left(t\right) = \left[x_1\left(t\right), x_2\left(t\right)\right]^T$, we can apply blind signal separation, using the following model of the data.

2.2. Data model for blind signal separation

We write the original, m-dimensional source data as $\mathbf{s}\left(t\right) = \left[s_1\left(t\right), s_2\left(t,\right), \ldots, s_m\left(t\right)\right]^T$. It is assumed that the sources are independent. We then consider an unknown linear model $\mathbf{A}_{n\times m}$ generating the observed signals, written as an n-dimensional vector $\mathbf{x}\left(t\right) = \left[x_1\left(t\right), x_2\left(t\right) \ldots, x_n\left(t\right)\right]^T$ by

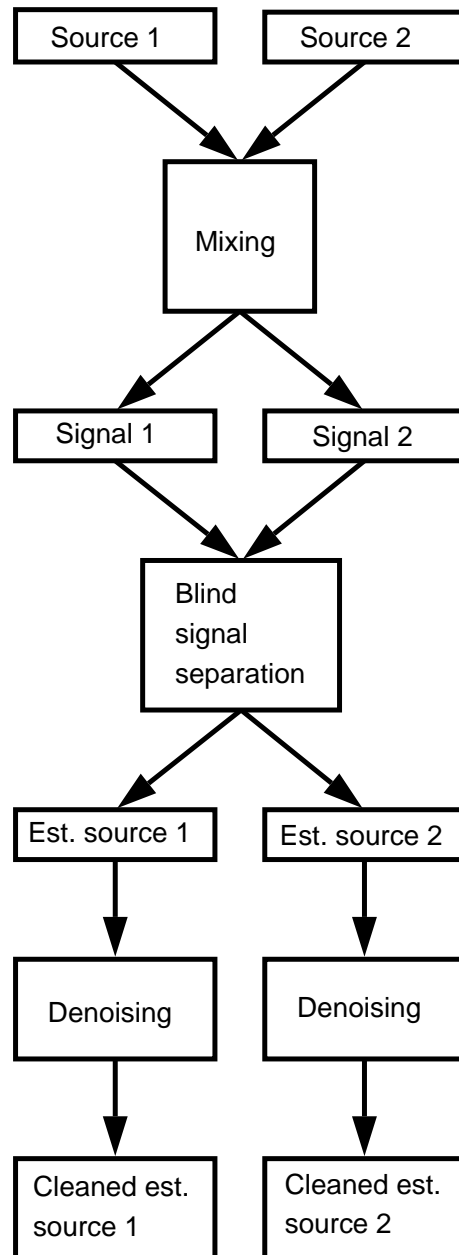$$\mathbf{x}\left(t\right) = \mathbf{A}\mathbf{s}\left(t\right), \tag{3}$$

Figure 3. Flowchart showing the general steps involved in going from the original sources to the cleaned, estimated sources. The optional Gaussian smoothing step is not shown.

where $\mathbf{A}$ is referred to as the mixing matrix. Being able to swap columns of $\mathbf{A}$, and scaling a source by a scaling change in a row of $\mathbf{A}$ means there is an ambiguity in both the permutation (of labeling) of the sources and the scaling of the sources respectively. With this model of the data, we can now apply blind signal separation techniques.

2.3. Blind signal separation

There are two key blind signal separation approaches that are combined to form the joint cumulant and correlation (JCC) algorithm in subsection 2.6. They are the second order blind identification (SOBI) algorithm, discussed in subsection 2.4, and the joint approximate decomposition of eigenmatrices (JADE), discussed in subsection 2.5. Both approaches have a common first step, in which the data is whitened using a *sphering* matrix $\mathbf{W}$, which transforms the mixing matrix $\mathbf{A}$ into a unitary matrix $\mathbf{U}$, which is a matrix for which $\mathbf{U}\mathbf{U}^T = \mathbf{I}$.[9, 10] The next step of estimating $\mathbf{A}$ is dependent on the choice of algorithm and detailed below.

2.4. SOBI algorithm

Given a hypothesis of sources with different spectra and the linear model of Equation 3, we can calculate time-delayed, cross-correlation matrices,

$$\mathbf{R}\left(\tau\right) = \mathrm{E}\left[\mathbf{x}\left(t\right)\mathbf{x}\left(t-\tau\right)^T\right]$$
$$\mathbf{A}\mathbf{R}_s\left(\tau\right)\mathbf{A}^H. \tag{4}$$

for $\tau \neq 0$ and

$$\mathbf{R}_s = \begin{pmatrix} \mathrm{E}\left[s_1\left(t\right)s_1\left(t-\tau\right)\right] & 0 & 0 & \cdots & 0 \\ 0 & \mathrm{E}\left[s_2\left(t\right)s_2\left(t-\tau\right)\right] & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & & \vdots \\ 0 & \cdots & & 0 & \mathrm{E}\left[s_m\left(t\right)\left(t-\tau\right)\right] \end{pmatrix}, \tag{5}$$

for $E[\cdot]$ the expectation operator. The correlation matrices can then be whitened,

$$\underline{\mathbf{R}} = \mathbf{W}\mathbf{R}\left(\tau\right) = \mathbf{U}\mathbf{R}_s\left(\tau\right)\mathbf{U}^T, \tag{6}$$

$\forall t \neq 0$. The joint diagonalization of the set of p whitened correlation matrices $\{\underline{\mathbf{R}}\left(\tau_i\right) | i = 1, \ldots, p\}$.[9] The matrix $\mathbf{U}$ can only be uniquely determined iff for any (i, j), there exists at least one lag τ_k such that $\mathrm{E}\left[s_i\left(t\right)s_i\left(t-\tau\right)\right] \neq \mathrm{E}\left[s_j\left(t\right)s_j\left(t-\tau\right)\right]$.[9] The mixing matrix is then estimated by $\hat{\mathbf{A}} = \mathbf{W}\mathbf{U}$. An alternative to the SOBI algorithm is the JADE algorithm.

2.5. JADE algorithm

Here we assume the linear model of Equation 3 and assume independence of sources. To each n-dimensional vector x is associated a quadicovariance matrix $\mathbf{Q} : \mathbf{M} \to \mathbf{N}$ defined by $\mathbf{N} = \mathbf{Q}\mathbf{M}$ such that

$$N_{i,j} = \sum_{(k,l)} \mathrm{Cum}\left(x_i, x_j, x_k, x_l\right) M_{k,l}, \tag{7}$$

where $\mathrm{Cum}\left(\cdot\right)$ is defined as

$$\mathrm{Cum}\left(x_i, x_j, x_k, x_l\right) = \mathrm{E}\left[\bar{x}_i\bar{x}_j\bar{x}_k\right] - \mathrm{E}\left[\bar{x}_i\bar{x}_j\right]\mathrm{E}\left[\bar{x}_k\bar{x}_l\right] - \mathrm{E}\left[\bar{x}_i\bar{x}_k\right]\mathrm{E}\left[\bar{x}_j\bar{x}_l\right] - \mathrm{E}\left[\bar{x}_i\bar{x}_l\right]\mathrm{E}\left[\bar{x}_j\bar{x}_k\right], \tag{8}$$

and where $\bar{x}_i = x_i - \mathrm{E}\left[x_i\right]$, etcetera.[14] As the set of $n \times n$ matrices is an n^2-dimensional linear space, it can be shown that there exist n^2 real eigenvalues λ_r and n^2 orthonormal eigenmatrices $\mathbf{M}_r$ satisfying $\mathbf{Q}\mathbf{M}_r = \lambda_r\mathbf{M}_r$.[9] It can be proved that only n of the eigenvalues are non-zero,[15] and joint diagonalizaton of the n corresponding eigenmatrices, labeled $\underline{\mathbf{M}}_r$, gives the unitary matrix U.[9] As with the SOBI algorithm, the mixing matrix is estimated by $\hat{\mathbf{A}} = \mathbf{W}\mathbf{U}$. Combining the two algorithms gives us the JCC algorithm.

2.6. JCC algorithm

In JCC, we use both the correlation information provided by the SOBI algorithm of subsection 2.4, $\underline{\mathbf{R}}(\tau_i)$, and the cumulant quadricovariance eigenmatrices, $\underline{\mathbf{M}}_r$, provided by the JADE algorithm. Joint diagonalization gives the unitary matrix $\mathbf{U}$, which again acts to give an estimator $\underline{\mathbf{A}} = \mathbf{WU}$. Using $\hat{\mathbf{A}}$ we separate the signals into estimates of the original source data. We then consider removing the noise from this data, using wavelet techniques.

2.7. Wavelet noise removal

The mathematical description of the continuous wavelet transform (CWT) of $f \in \mathbf{L^2}(\mathbb{R})$ is described by Mallat[16] as

$$(Wf)(u,s) = \int_{-\infty}^{+\infty} f(t)\,\psi_{u,s}^*(t)\,dt, \tag{9}$$

where

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}, \tag{10}$$

is a family of orthogonal wavelets, $|\psi_{u,s}| = 1$, $\langle \psi_{u,s}, \psi_{u',s'} \rangle = 0$ for $(u,s) \neq (u',s')$, and

$$\int_{-\infty}^{+\infty} \psi_{u,s}(t)\,dt = 0. \tag{11}$$

The scale of the wavelet may conceptually be considered the inverse of the frequency.

The CWT reveals much detail about a signal, however due to the continuous nature it cannot be computed for real signals on a digital computer. Therefore, the discrete wavelet transform (DWT) is normally used. The DWT calculates the wavelet coefficients at discrete intervals of time and scale instead of at all scales. With the DWT, a fast version of the algorithm is possible, analogous to the fast Fourier transform. This version of the algorithm makes use of the fact that if scales and positions are chosen based on powers of two (dyadic scales and positions) the analysis is very efficient. In 1988, Mallat developed an efficient way to implement this algorithm, which is known as a two-channel sub-band coder.[17] For a single level of decomposition, this algorithm passes the signal through two complementary (high-pass and low-pass) filters resulting in approximations which are high-scale, low-frequency components of the signal, and details, which are low-scale, high-frequency components of the signal. This results in twice as many data-points so the data is down-sampled. For further levels of decomposition, successive approximations may be iteratively broken down into details and approximations as shown in Figure 4. Coefficients below a certain level are regarded as noise and thresholded out. Thresholding may be soft or hard. Hard thresholding is defined as

$$\begin{aligned} y &= x && \text{for } |x| > \theta \\ y &= 0 && \text{for } |x| \leq \theta \end{aligned} \tag{12}$$

and soft thresholding as

$$\begin{aligned} y &= \text{sign}(x)(|x| - \theta) && \text{for } |x| > \theta \\ y &= 0 && \text{for } |x| \leq \theta \end{aligned} \tag{13}$$

where x is the original signal, y is the thresholded signal, and θ is the threshold. Hard thresholding tends to create discontinuities at $x = \pm\theta$ because any values of the signal less than the threshold are immediately set to zero. With soft thresholding, the thresholded values are shrunk towards zero without creating the discontinuities. The signal is then reconstructed without significant loss of information. Then the signal may be reconstructed by up-sampling, passing the approximations and details through the appropriate reconstruction filters and combining the results. Based on SNR measures of wavelet performance, we used Daubechies wavelets of order 5, with soft thresholding and a decomposition level of 5; although this is not the best for noise removal, we are more interested in preservation of information when going from the estimated sources to the denoised estimated sources.

To evaluate the performance of the above techniques we introduce a measure of mutual information.
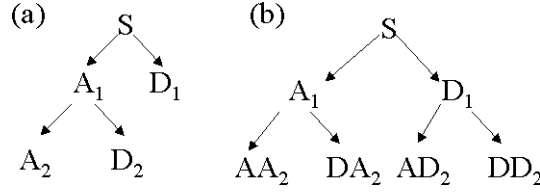
Figure 4. This figure illustrates how (a) the discrete wavelet transform decomposes a signal into details and approximations iteratively decomposing the approximations; (b) wavelet packets iteratively decompose the approximations and details.

2.8. MI estimation algorithm

The following is an outline of the method we use in calculating the mutual information between the EEG and EOG signals, as given in Kraskov *et al.*.[18]

Mutual information for two signals X and Y is defined in Equation 14

$$I(X,Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x,y) \log \frac{\mu(x,y)}{\mu_x(x)\mu_y(y)} dx dy, \tag{14}$$

where μ, μ_x and μ_y are probability measures. We then take the set of points $z_i = (x_i, y_i)$ for the EEG x_i and EOG y_i, $i = 1, \ldots N$. Then we find the kth closest neighbor of each z_i according to the metric

$$|z - z'| = \max\{|x - x'|, |y - y'|\}. \tag{15}$$

The kth nearest neighbor is then projected onto the x and y axes giving the distances $\epsilon_x(i)/2$ and $\epsilon_y(i)/2$ respectively. The mutual information is estimated by:

$$\hat{I}_k(X,Y) \approx \psi(k) - 1/k - \langle \psi(n_x) + \psi(n_y) \rangle + \psi(N), \tag{16}$$

where $\psi(\cdot)$ is the digamma function given by

$$\psi(z) = \frac{d}{dz} \ln \Gamma(z), \tag{17}$$

and

$$\langle \ldots \rangle = \frac{1}{N} \sum_{i=1}^{N} \mathrm{E}[\ldots(i)]. \tag{18}$$

3. RESULTS

3.1. Blind signal separation

We compared the blind signal separation for three algorithms, across four sets of data, two of thoracic and abdomen (TA) breathing data, and two of EEG and EOG (EE) data. The differences in mutual information between the signal data and the estimated source data are shown in Table 1. The higher the mutual information, the better the algorithm is for separating the original sources, given the assumptions of that algorithm.

3.2. Wavelet denoising

For each of the generated estimates of the sources (three blind signal separation algorithms applied to four pairs of signals) we estimated the mutual information between the estimates of the sources, and the denoised estimates of the sources. These are given in Table 2. We observe no difference in mutual information between

Table 1. The difference in mutual information (in nats/sample) between the two estimated sources and the two signals, $\hat{I}_k$ (est. sources) $- \hat{I}_k$ (signals), for the two sets of thoracic-abdominal (TA) data and the two sets of EEG and EOG (EE) data. Nats are units of information, when a natural logarithm is used. This is computed for all three (SOBI, JADE, and JCC) algorithms

	TA1	TA2	EE1	EE2
SOBI	1.0319	0.4539	0.4522	0.6101
JADE	< 0.0001	< 0.0001	< 0.0001	< 0.0001
JCC	0.3640	0.4292	0.4032	0.5342

Table 2. The difference in mutual information (in nats/sample) between the JADE estimated sources and the wavelet denoised estimated sources is computed for the two sets of thoracic-abdominal (TA) data and the two sets of EEG and EOG (EE) data, $\hat{I}_k$ (est. denoised sources) $- \hat{I}_k$ (est. sources). Nats are units of information, when a natural logarithm is used.

TA1	TA2	EE1	EE2
< 0.0001	< 0.0001	< 0.0001	0.0001

4. DISCUSSION & CONCLUSIONS

For our particular set of thoracic and abdominal breathing data, the SOBI algorithm works well, with an increase in the mutual information, probably because the sources have reasonably distinct spectra. Since the JCC combines information from both the SOBI and JADE algorithms by way of joint diagonalization, it introduces the problems associated with using the JADE algorithm for this data, namely that the sources are not independent. The two sources have a high level of dependence, being almost synchronous during regular breathing, tending to differ only for compliant chests in young children or when a breathing obstruction occurs.[5, 7] Similarly, for the EEG and EOG data, although these are more independent, the SOBI algorithm performs best at separating the original sources from the observed signals.

The wavelet denoising performs well, in that it preserves (as far as we can determine) the information present in the signals. Further work will consider wavelet packet and matching pursuit denoising algorithms,[17, 19, 20] and how these effect mutual information between two different channels. We will also consider the effect of swapping the denoising and blind signal separation techniques, in theory this should have little to no difference on the results.

5. ACKNOWLEDGEMENTS

REFERENCES

1. K. Kaemingk, A. Pasvogel, J. Goodwin, M. S.A., F. Martinez, P. Enright, G. Rosen, W. Morgan, R. Fregosi, and S. Quan, "Learning in children and sleep disordered breathing: findings of the Tucson children's assessment of sleep apnea (TuCASA) prospective cohort study," *Journal of the International Neuropsychological Society* **9**, pp. 1016–1026, 2003.
2. E. Sforza, S. Grandin, C. Jouny, T. Rochat, and V. Ibanez, "Is waking electroencephalographic activity a predictor of daytime sleepiness in sleep-related breathing disorders?," *European Respiratory Journal* **19**, pp. 645–652, 2002.
3. A. Muzur, E. Pace-Schott, and J. Hobson, "The prefrontal cortex in sleep," *Trends in Cognitive Sciences* **6**, pp. 475–481, 2002.
4. P. Anderer, G. Gruber, G. Klösch, W. Klimesch, B. Saletu, and J. Zeitlhofer, "Sleep and memory consolidation: the role of electrophysiological neuroimaging," *Somnologie* **6**, pp. 54–62, 2002.

5. K. Brown, R. Platt, and J. Bates, "Automated analysis of paradoxical ribcage motion during sleep in infants," *Pediatric Pulmonology* **33**, pp. 33–46, 2002.

6. G. Prisk, J. Hammer, and C. Newth, "Techniques for measurement of thoracoabdominal asynchrony," *Pediatric Pulmonology* **34**, pp. 462–472, 2002.

7. V. Menon and D. Agrawal, "Physics of quiet and suffocative breathing," *American Journal of Physics* **71**, pp. 474–478, 2003.

8. M. Teplan, "Fundamentals of eeg measurement," *Measurement Science Review* **2**, pp. 1–11, 2002.

9. I. Gorodnitsky and A. Belouchrani, "Joint cumulant and correlation based signal separation with application to EEG data analysis," in *ICA 2001*, 2001.

10. A. Belouchrani, K. Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans. on Sig. Proc.* **45**, pp. 434–444, Feb. 1997.

11. M. Matalgah and J. Knopp, "Improved noise stability multiresolution filters based on a class of wavelets," in *Proceedings of the 37th Midwest Symposium on Circuits and Systems: Vol 2.*, pp. 817–820, IEEE, Aug. 1994.

12. O. Bertrand, J. Bohorquez, and J. Pernier, "Time-frequency digital filtering based on an invertable wavelet transform: an application to evoked potentials," *IEEE Transactions on Biomedical Engineering* **41**, pp. 77–88, Jan. 1994.

13. L. M. Lim, M. Akay, and J. A. Daubenspeck, "Identifying respiratory-related evoked potentials," *IEEE Engineering in Medicine and Biology* **14**, pp. 174–178, Mar.Apr. 1995.

14. J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation* **11**, pp. 157–192, 1999.

15. J.-F. Cardoso and A. Souloumiac, "An efficient technique for blind separation of complex source," in *Proc. IEEE Sig. Proc. Workshop on Higher-Order Stat.*, pp. 275–279, 1993.

16. S. G. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1999.

17. S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, pp. 674–693, July 1989.

18. A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *ArXiv cond-mat/0305641* , May 2003.

19. S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing* **41**, pp. 3397–3415, Dec. 1993.

20. S. Krishnan and R. M. Rangayyan, "Automatic denoising of knee-joint vibration signals using adaptive time-frequency representations," *Medical and Biological Engineering and Computation* **38**, pp. 2–8, Jan. 2000.