33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Independent Component Analysis as a Preprocessing Step for Data Compression of Neonatal EEG

Bogdan Mijović, *Student Member, IEEE,* Vladimir Matić, Maarten De Vos, and Sabine Van Huffel, *Fellow, IEEE*

Abstract— **We propose a novel approach for compressive sampling of the neonatal electro-encefalogram (EEG) data. The method assumes that the set of EEG data is generated by linearly mixing a fewer number of source signals. Another assumption is that the sources are nearly-sparse in Gabor dictionary. The presented method, instead of compressing original EEG channels, first performs a data-reduction, and then compresses the obtained sources. With this approach we showed that the gain in reconstruction speed is 33%-50%, whereas the compression rate is enhanced by 33%.**

I. INTRODUCTION

At the neonatal intensive care units, continuous electroencephalographic (EEG) recordings are regularly performed for the assessment of hypoxic brain injuries of newborns. Nowadays, there is a tendency for the development of wireless EEG devices, that would decrease the amount of movement artifacts and provide a comfortable surrounding for the babies. One of the major issues is the large quantity of data that has to be transmitted over the wireless link. It is common that approximately 20 EEG channels are sampled at a sampling frequency (fs) of $256Hz$, thus producing around 5000 samples per second. This significantly affects the battery life, as the recordings should be continuous for a period of 48 up to 72 hours. Therefore, there is a growing interest for the data compression methods that can efficiently compress the EEG data into a few number of samples. This would allow for the fast wireless transmission of the collected EEG data in a clinical setting.

Compressive sensing (CS) provides a new emerging framework for signal compression, which has acquired a lot of attention in recent years within the signal processing society [1], [2], [3], [4]. This theory shows that each signal, which has a sparse representation in a basis (or a certain dictionary), can be recovered from a small number of measurements from the original basis, which is comparable with the so called "sparsity" of the signal. In this formulation, sparsity denotes the number of atoms (information), that is needed to represent the signal in the basis in which it is sparse.

The CS framework for compressing the adult EEG signals has already been proposed, and its near-sparsity in a Gabor dictionary has already been shown [5]. In that work, it was also shown that for adult EEG joint sparsity (proposed in [6]) can be used, although with a slightly worse reconstruction performance than a regular (channel-by-channel) compression. This is due to the fact that joint sparsity assumes all the channels to be composed of the same atoms, only different weighting coefficients are allowed. The "channels" in [5] were not different EEG channels, but the compression was performed for different trials of the same task on the same channel. However, in this work the compression of the whole EEG signal (18 channels) is performed. Additionally, the EEG recordings of neonates have less common information since the brain connectivity functions have not yet been fully developed. Therefore, joint sparsity is not applicable, i.e. it would give large errors during the reconstruction stage.

A new approach for compressing the multichannel data is reported for compression of Hyper Spectral Images (HSI) [7]. In that work, instead of compressing all the images (channels), the compression and reconstruction of the source signals is proposed. It is assumed that the signals are dependent across the channels, and that only few number of sources are generating multichannel observation based on a linear mixture model. The mixing matrix is then used during the reconstruction stage for recovering the original images. It was shown that sources are more sparse than the original image, and a big improvement in the compression rate is obtained.

Following the same logic, the EEG signal on scalp can also be represented like a mixture of the underlying sources, located deep inside the brain. However, in EEG recordings, the mixture matrix of possible sources in the brain is unfortunately not a priori known. In this paper, we propose the following compression algorithm for the EEG data:

- First apply a data reduction algorithm (Principle Component Analysis (PCA) or Singular Value Decomposition (SVD)). In this way we obtain a reduced matrix of sources which retains most of the data variance with a lower number of channels.
- After PCA or SVD, one of the Independent Component Analysis (ICA) algorithms may be optionally used to make the sources more structured, and therefore more easily compressible.

RResearch supported by: Research Council KUL: GOA MaNet, PFV/10/002 (OPTEC); Flemish Government: FWO: G.0427.10N (Integrated EEG-fMRI); IBBT Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO)

B. Mijović, V. Matić, M. De Vos and Sabine Van Huffel are with the Department of Electrical Engineering (ESAT-SCD), Katholieke Universiteit Leuven, Belgium, and IBBT-K.U.Leuven Future Health Department, Leuven, Belgium bogdan.mijovic@esat.kuleuven.be

M. De Vos received Alexander Von Humboldt scholarship with the Neuroscience Lab, Department of Psychology, Oldenburg University, Germany

- Eventually, the sources are compressed, and together with the mixing matrix, they are sent over the link, and reconstructed on the receiver's site using one of the available CS reconstruction algorithms.

In this paper we show the performance of the algorithm formed as explained before. The ICA algorithm used in this work was Second-Order Blind Identification (SOBI), due to its computational efficiency. Iterative Hard Thresholding (IHT) algorithm [8] was used in the reconstruction stage. The results are compared with the performance of a regular (channel-by-channel) compression, and the conclusions are drawn and discussed in the later sections.

II. METHODS

A. Compressive Sensing

In the CS theory, a signal x of length N is called K-sparse in dictionary Ψ if it can be represented by K atoms (elements) from that dictionary, i.e. $x = \sum_{i=1}^{K} a_i \psi_i$, where a_i are the coefficients associated with the atom ψ_i, and $K << N$. The CS theory further states that it is possible to construct a MxN matrix Φ (called *measurement matrix*), where M is of the order of K ($M << N$), which will allow for the reconstruction of the signal x from the measurements $y = \Phi x$. This is possible only if the measurement matrix satisfies the so called *Restricted Isometry Property* (RIP) [9]. In short, if an $M \times N$ matrix Φ satisfies the K-RIP, it ensures that all the submatrices of Φ of size $M \times K$ are close to an isometry, and therefore distance (and information) preserving. While checking whether a measurement matrix Φ satisfies RIP is an *NP*-Complete problem in general [10], random matrices, whose entries are independent and identically distributed (i.i.d.) Gaussian, Rademacher (± 1) or more generally subgaussian, showed to work with high probability [11]. Moreover, these matrices also have the so-called *universality* property, that is for any choice of orthonormal basis Ψ, $\Phi\Psi$ also satisfies RIP with high probability [12], and therefore matrix Φ is a good choice for a measurement matrix.

$$argmin_a ||a||_1, \text{ such that } y = \Phi \Psi a \quad (1)$$

The CS reconstruction problem can be formulated as follows: Find coefficients a, s.t. $y = \Phi\Psi a$, where y, Φ and Ψ are known. This problem can be reformulated as a linear programing (LP) problem (like in equation 1), and can be solved by l_1 convex minimization. It has been shown in [13] that the necessary number of measurements is $M > 2K\log(N/M)(1 + o(1))$. However, due to large calculation costs of LP, we chose one of the greedy approaches, thereby minimizing the l_0 instead of l_1 norm. Available algorithms of this sort are Orthogonal Matching Pursuit (OMP), Iterative Hard Thresholding (IHT), COmpressed SAmpling Matching Pursuit (CoSaMP)... In this work IHT algorithm is used due to its simplicity and speed (IHT is shown to perform faster than the other available algorithms [14], [8]). Although greedy algorithms are generally proven to yield the exact solution if it exists with fewer computations at the expense of slightly more measurements, it has been shown that IHT possess some very interesting properties, among which near-optimal error guaranties, robustness to observation noise, low computational complexity and uniform performance guarantees (depend only on properties of the sampling operator and signal sparsity)[8].

B. Dictionary

It has been previously shown that the EEG data are sparse in an over-complete Gabor dictionary [5]. Therefore, the dictionary used in this work was Gabor dictionary, and it has been created with an atom length of 1024 samples (4 seconds of the EEG signal), which yielded 40.960 atoms in total. This dictionary covered the frequency band up to $128Hz$ ($fs = 256Hz$).

C. Data

In this study, we used 50 blocks of 18 channels, 4 seconds long EEG data recorded on hypoxic neonates. The sampling frequency was $256Hz$. The data were high-pass filtered at $1Hz$ to remove the DC component, and a notch filter at $50Hz$ for removing the power-line interference was applied. No low-pass filtering has been performed. The data were compressed in two different ways. First, the data were compressed on a channel-by-channel basis for 4 different numbers of measurement M, namely 100, 133, 150 and 200 per channel.

The second approach was to first perform the data reduction step, solving the ICA problem $X = AS$. In this equation A is the mixing matrix, and S are the computed sources. Then the data are sampled with the sensing matrix, and the measurements y are acquired $y = \Phi S$, such that the total number of measurements was the same for both approaches. In our approach when ICA was used for preprocessing, also mixing matrix A, has to be included in the total number of measurements. The number of derived components in the reduction stage was 9 for $M = 100$ and $M = 150$ and 12 for $M = 133$ and $M = 200$ (see Table I for details). The sources are reconstructed with the IHT algorithm, and the estimated sources coefficients a are first derived by solving $y = \Phi\Psi a$, and then the estimated sources $\hat{s}$ are computed $\hat{s} = \Psi a$. Finally, the estimated recovered signals are calculated by $\hat{X} = A\hat{S}$. The performances of the reconstructed signals $\hat{x}$ for two approaches were assessed based on the speed of recovery and the normalized root mean square error (NMSE) of the reconstructed signals.

III. RESULTS

It is apparent that in our setup, one first has to perform the ICA algorithm (the data reduction stage) before compression. Therefore, we need to check how much time is required to perform the ICA and to compress the data. In our experiments, for 1024 samples (4 seconds)18-channel EEG data, the compression time (including the data reduction step) was always less than 0.5 seconds on the personal computer with Intell Core2Duo processor, and 4Gb of RAM. Therefore, we

TABLE I

THE NUMBER OF MEASUREMENTS WITH REGULAR COMPRESSION AND WITH OUR ICA APPROACH. THE TOTAL NUMBER OF MEASUREMENT IS FOR BOTH APPROACHES THE SAME

Total Number of Measurements (M)	Number of ICs	Number of Measurements for IC
18×100	9	182 + 18
18×133	12	182 + 18
18×150	9	282 + 18
18×200	12	282 + 18

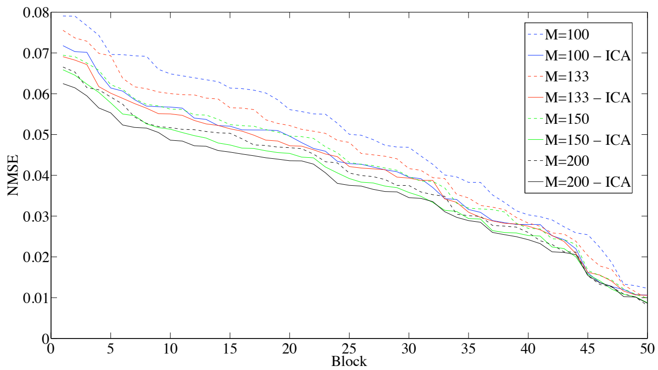

Fig. 1. The NMSE for 50 different 4-second blocks of EEG data. The dashed lines belong to regular compression, whereas the solid lines belong to compression with ICA data reduction step. Different colors denote different settings in terms of M

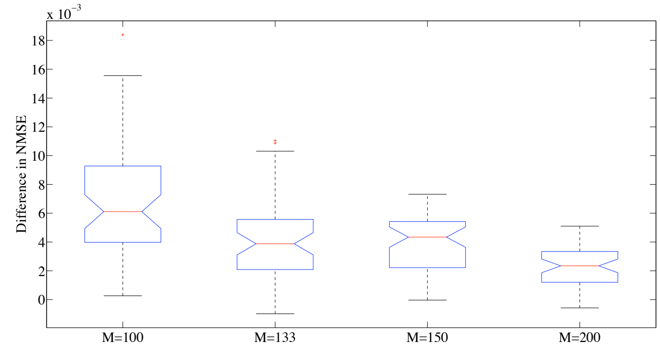

Fig. 2. Boxplot of NMSE differences between regular compression and compression with the ICA reduction step. It is obvious that the difference is always positive, showing that regular compression always has higher NMSE

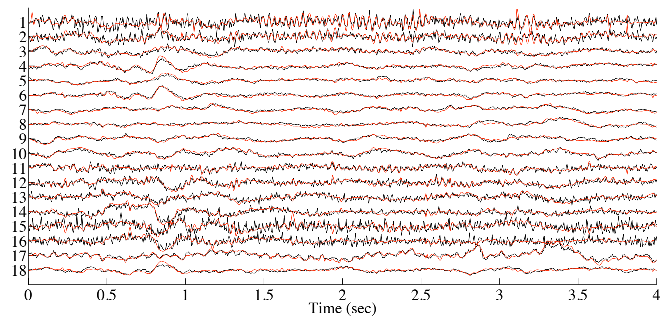

Fig. 3. The worst reconstructed 4 seconds block using our approach with ICA data-reduction step for $M = 200$. Black is the original signal, and the reconstructed signal is shown in red

conclude that speed sets no limits for performing ICA in the compression stage.

Additionally, it is important to know the amount of information lost due to data reduction. In our experiments, the preserved variance was around 90% if the reduction was performed with 9 independent components, and around 95% when 12 components were used.

After we know that ICA can be performed, the benefit of using this kind of preprocessing in the compression stage was checked. Fig. 1 shows the NMSE of the reconstructed data for regular compression (dashed line) and compression when ICA was included (solid line). In this figure, M is the number of measurements per channel. It is apparent that not only reconstruction with ICA always gave better results than a regular compression, but also that even with less measurements ($M = 100$ instead of $M = 133$ or $M = 150$), better reconstruction was achieved. Additionally, the ICA preprocessed data for $M = 150$ were better reconstructed than regularly (channel-by-channel) compressed data with $M = 200$.

Fig. 2 shows the boxplot of differences in the reconstruction errors between regular compression, and compression with ICA preprocessing for different values of M. It is apparent that only in a few cases regular compression yielded better results, and therefore our approach clearly outperforms the regular one.

In Fig. 3 we show the part of the EEG signal where our algorithm performed the worst for $M = 200$ measurement per channel. This part of the EEG data gave also the worst results with the regular compression. It is obvious that the structured part of the data is nicely modeled, whereas our algorithm failed to fully recover some of the high-frequency artifacts. However, the overall reconstruction is still of high quality.

Finally, we tested what is the speed benefit of our algorithm in the reconstruction stage. We found that on our data-set, average reconstruction per channel (both for original channels and independent sources) was around 100 seconds. However, since the number of sources is lower than the number of channels (in our setting 50% for $M = 100$ and $M = 150$, and 25% for $M = 133$ and $M = 200$), we conclude that the speed benefit is 33% to 50% for this experiment.

IV. DISCUSSION

In this paper we propose a compressive sampling approach that uses the data reduction step in the preprocessing stage. We used SOBI ICA, although other data reduction methods are also possible. After the data reduction has been performed, obtained sources are compressed, instead of the original signal. In this way, the amount of data that has to be transmitted is reduced. The sources are then reconstructed, and multiplied with the mixing matrix in order to obtain the reconstructed signals.

We showed that the data reduction is fast enough (less than 0.5 second for the 4 seconds data) and that therefore can be incorporated in the compression stage. Concerning the accuracy of data reconstruction, we showed (Figs. 1,2) that the accuracy is almost always higher with the data when ICA has been used, if the same number of measurement has been transmitted. Moreover, it is apparent from Fig. 1, that even when $M = 100$ measurements has been transmitted with ICA preprocessed compression, the accuracy of the reconstruction is at least comparable with the case when $M = 150$ measurements has been sent with the regular approach. Thus, besides the better reconstruction with the same number of transmitted data, one may also obtain better reconstruction with even 33% less measurements.

The speed of the reconstruction depends also on the number of channels to be recovered. In our approach, instead of compressing, and afterwards reconstructing all of the 18 channels one-by-one, we send only the estimated independent sources. In this experiment, the number of sources was 9 or 12, meaning that only 9 to 12 signals have to be reconstructed. Since the speed of channels or sources reconstruction is approximately the same, we have 33% to 50% gain in reconstruction speed.

Fig. 3 shows the worst reconstruction of our data among the 50 blocks that were compressed and reconstructed. It can be seen that the structured EEG parts were correctly recovered, whereas sometimes the algorithm failed to properly model the noise (muscle artifact), present in the channels 15 and 16. However, even the reconstruction of the noise from channels 1 and 2 is fairly nice.

For the reconstruction stage, the iterative hard thresholding (IHT) algorithm was used, due to it's simplicity, accuracy and speed [8]. However, some other, faster algorithms, like Accelerated IHT (AIHT) [15] may be used as well. In our work, only the difference in reconstruction speed between our algorithm and regular compression was important, not the absolute speed itself.

The compression of *adult* EEG signals has already been demonstrated previously [5]. It has been shown there that joint sparsity can be used for compression in order to accelerate the reconstruction process at the slight cost of the reconstruction accuracy. However, the data used in that study are trials of the same task of one adult EEG channel data, with maximum frequency of only $50Hz$. In that case, the assumption of joint sparsity that all the channel consist of the same atoms, with different coefficients [6] sounds reasonable, and it can be expected that all the reconstruction may be fairly good.

The data used in this setup are multichannel neonatal EEG data. Brain connectivity in neonates is not yet developed, and therefore the information embedded in different channels may be highly different. Therefore, it cannot be expected that the joint sparsity assumption holds. However, with implementing ICA and compressing the independent sources instead of the original channels, we have gained both in reconstruction accuracy, and speed (see Fig. 1). It is also apparent that the difference in reconstruction accuracy drops with increasing the number of measurements (Fig. 2), what was expected. Also, the quality of reconstruction itself in this case increases.

Taking this into account, one can also expect that the compression of the adult EEG data can be recovered with even greater compression rate, and with higher accuracy, since this data is highly redundant. The huge number of adult EEG channels (typically 64 to 128) can be significantly reduced using one of the data reduction techniques without substantial loss of relevant information. The focus of this paper, however, were only the neonatal EEG recordings.

Another possible improvement of the algorithm would be to explore the properties of the mixing matrix A, and to check if this matrix can be adaptively updated, or optimized to a constant (like it is the case in [7]) instead of performing ICA for each block of data separately. This is a first step with promising results towards obtaining higher compression ratio for the neonatal and adult EEG signals, since they are not fully sparse in any orthonormal basis.

V. Conclusions

In this paper we proposed a new method for compressive sampling of the neonatal EEG data. We showed that our new method clearly outperforms the regular compression algorithms, with the speed gain in the reconstruction stage enhanced by 33% to 50%. We believe that the same method can also be successfully applied to adult EEG data.

References

[1] D. L. Donoho, "Compressed Sensing" *IEEE Trans. Inf. Theory*, vol. 52, no. 4, 2006, pp 1289-1306

[2] E.J. Candes, "Compressive Sampling" *in Proc. Int. Congr. Math*, Madrid, Spain, vol. 3, 2006, pp 1433-1452

[3] E.J. Candes, "An Introduction to Compressive Sampling" *IEEE Signal Processing Mag.*, vol. 25, no. 2, 2008, pp 21-30.

[4] R. G. Baranuik "Compressive Sensing", *IEEE Signal Processing Mag.*, vol. 24, no. 4, 2007, pp 118-120.

[5] S. Aviyente, "Compressed Sensing Framework for EEG Compression", *IEEESP 14th Workshop on Stat. Sig. Proc.*, 2007, pp.181-184

[6] M.F. Duarte, S. Sarvotham, D. Baron, M.B. Wakin, and R.G. Baraniuk "Distributed compressed sensing of jointly sparse signals", *in Asilomar Conf. Signals, Sys., Comput*, 2007, pp 1537-1541.

[7] B. K. Natarajan, "Distributed Compressed Sensing of Hyperspectral Images via Blind Source Separation", *in Asilomar Conf. Signals, Sys., Comput*, 2010, pp 227-234.

[8] T. Blumensath and M. E. Davies, "Iterative Hard Thresholding for Compressed Sensing", *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, 2009, pp. 256-274

[9] E.J. Candes, "The restricted isometry property and its implications for compressed sensing" *Comptes Rendus Mathematique*, vol. 346, no. 9, 2008, pp 589-592.

[10] B. K. Natarajan, "Sparse Approximate Solutions to linear systems", *SIAM J. Computat.*, vol. 24, no. 2, 1995, pp 227-234.

[11] D. L. Donoho and J. Tanner, " Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing" *Phil. Trans. R. Soc. A*, vol. 367, no. 1906, 2009, pp 4273-4293

[12] R. G. Baranuik, V. Cevher, M. F. Duarte and C. Hegde, "Model-Based Compressive Sensing", *IEEE Trans. Inf. Theory*, vol. 56, no. 4, 2010, pp 1982-2001.

[13] D. L. Donoho and J.Tanner, "Precise Undersampling Theorems", *Proceedings of the IEEE*, vol. 98, no. 6, 2010, pp. 913-924

[14] J.Tanner, "Phase Transitions for Greedy Sparse Approximation Algorithms", *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, 2011, pp. 188-203

[15] T. Blumensath, "Accelerated Iterative Hard Thresholding", *preprint*, 2011.