

Biomedical Physics & Engineering Express



PAPER

Dictionary selection for compressed sensing of EEG signals using sparse binary matrix and spatiotemporal sparse Bayesian learning

OPEN ACCESS

RECEIVED
1 July 2020

REVISED
23 September 2020

ACCEPTED FOR PUBLICATION
14 October 2020

PUBLISHED
29 October 2020

Manika Rani Dey¹, Arsam Shiraz² , Saeed Sharif¹, Jaswinder Lota¹  and Andreas Demosthenous² 

¹ Department of Engineering and Computing, University of East London, E16 2RD, United Kingdom

² Department of Electronic and Electrical Engineering, University College London, WC1E 7JE, United Kingdom

E-mail: a.demosthenous@ucl.ac.uk

Keywords: compressed sensing, dictionary, DWT, EEG, signal reconstruction

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Abstract

Online monitoring of electroencephalogram (EEG) signals is challenging due to the high volume of data and power requirements. Compressed sensing (CS) may be employed to address these issues. Compressed sensing using a sparse binary matrix, owing to its low power features, and reconstruction/decompression using spatiotemporal sparse Bayesian learning have been shown to constitute a robust framework for fast, energy efficient and accurate multichannel bio-signal monitoring. EEG signal, however, does not show a strong temporal correlation. Therefore, the use of sparsifying dictionaries has been proposed to exploit the sparsity in a transformed domain instead. Assuming sparsification adds values, a challenge, therefore, in employing this CS framework for the EEG signal, is to identify the suitable dictionary. Using real multichannel EEG data from 15 subjects, in this paper, we systematically evaluate the performance of the framework when using various wavelet bases while considering their key attributes namely number of vanishing moments and coherence with sensing matrix. We identified Beylkin as the wavelet dictionary leading to the best performance. Using the same dataset, we then compared the performance of Beylkin with the discrete cosine basis, often used in the literature, and the alternative of not using a sparsifying dictionary. We further demonstrate that using dictionaries (Beylkin and Discrete Cosine Transform (DCT)) may improve performance tangibly only for a high compression ratio (CR) of 80% and with smaller block sizes, as compared to using no dictionaries.

1. Introduction

The dynamic nature of biomedical signals such as electroencephalographic (EEG) and electrocorticographic (ECoG) traces results in a wide variation in normal and pathologic features in different individuals. The use of manually extracted features for prediction of pathological events is impractical with a large volume of data, even for a small number of electrodes, leading to large processing delays. Thus, automated feature extraction and signal processing methods are necessary for real time and clinically useful implementation in such applications. Real-time processing can be facilitated using cloud computing, Internet of Things (IoT) and deep learning, to effectively monitor and predict seizures using EEG signal [1], which requires high data volume transmission of the acquired bio-signals. In addition, remote

online monitoring and diagnosis using EEG signals can reduce the frequency of patient visits to hospitals [2–5].

Energy consumption and high volume of data are major constraints in transmission of EEG signal due to limited battery life and processing capability of sensor nodes. Recent efforts aiming to increase battery life focus on reducing the power of transmission and data rate with compressed sensing (CS) [6, 7]. As CS can lead to significant computational savings for on-chip implementation with relatively low sampling rates, recently it has been viewed with considerable interest as a viable technique for the transmission of large data volumes and high data rate signals [5]. In CS data is projected into a compressed format non-adaptively upon acquisition using a sensing matrix, which differs from conventional compression techniques where data is acquired then compressed and indices are stored.

A requirement of conventional CS is that the signal must be sparse in the domain where it is compressed [7]. EEG signal, however, is not sparse in time or the frequency domains [5]. A challenge, therefore, in employing conventional CS for the EEG signal, is to identify the domain known as the dictionary in which the EEG signal is sufficiently sparse. This leads to another sufficient requirement for CS which is the incoherence between the dictionary and the sensing basis matrix i.e., the level of dissimilarity between the two. For an accurate reconstruction of the original signal, the dictionary and sensing matrix must be highly incoherent. For EEG signals, the accuracy of reconstruction of the signal with CS depends on a suitable dictionary that is maximally incoherent with the sensing basis [8]. Various dictionaries have been developed and investigated to enable sparse representation. These include Gabor transforms (GT), discrete wavelet transforms (DWT), spline and discrete cosine transforms (DCT) [5, 9]. Results in these techniques indicate accurate reconstructions with less error, however, the specific features that make these appropriate or suitable dictionaries, have not been investigated or explained. Selecting a specific DWT for a given application to ensure an accurate reconstruction of the compressed signal is challenging. In most applications (other than EEG with CS), a key feature employed in selection of a DWT is the number of vanishing moments, which determines its ability to represent complex signals efficiently or more sparsely. According to the Strang-Fix condition (as a special case) the approximation order of a DWT increases with the number of vanishing moments up to the smoothness index (Hölder regularity) of the approximated signal [10]. That is, the sparseness of the wavelet-transformed signal is in general higher for longer wavelets. An equal number of vanishing moments for the DWT can also be viewed as all doing 'similar amounts of work' [11].

For reconstruction, block sparse Bayesian learning (BSBL) may be employed to exploit the block sparsity of bio-signals. Current motivations in employing CS include low hardware complexity with optimization algorithms, and novel BSBL approaches to reduce latency.

The authors in [5] propose a novel method to use the BSBL framework to compress/reconstruct non-sparse raw FECG recordings. Experimental results show that the framework can reconstruct the raw recordings with higher quality as compared to other BSBL and CS DWT based methods. The authors in [8] depart from previous CS-based approaches and formulate signal recovery from under-sampled measurements. In [9] the authors compare and detail performance of various dictionaries for CS in EEG and ECG signals in order to come up with an optimal dictionary and its suitability for deployment in embedded hardware. However, the authors do not reflect on prior analysis of dictionary properties such as

incoherency and vanishing moments for the choice of the dictionaries. A novel BSBL approach is given in [12] and the DCT is employed for increasing sparsity with the results presented for both ECG and EEG signals but does not relate to the choice of selecting the DCT [12]. In [13] an explanation in terms of incoherency is given for the choice of dictionary followed by an optimization algorithm leading to the optimal selection of the dictionary, based on a pre-selected class of dictionaries. The work detailed in [14] is on hardware implementation, no novel properties of the dictionaries are discussed. Other variations of BSBL include the spatiotemporal sparse Bayesian learning (STSBL) that exploits signal correlation [15]. The work in [15] offers a novel computational improvement over the BSBL methods and is not aimed at highlighting the attributes of DWTs for an optimal dictionary choice. The approach in [16] compares the accuracy of reconstruction for various dictionaries. It does not mention the choice or selection of wavelet in terms of the properties of incoherence and vanishing moments.

In this paper, we primarily aim to evaluate the usefulness of using a sparsifying dictionary with a sparse binary matrix (SBM) used as a sensing matrix for CS of multichannel EEG while STSBL is used for reconstruction\decompression. In doing so we arrive at the following novel contributions not reported in earlier literature:

We first investigate various DWT bases while considering their key attributes of incoherence with SBM, an important feature in basic CS methods, together with vanishing moments of DWT dictionaries, a defining feature of wavelet functions. Our results indicate that both features should be considered at the same time when selecting the dictionary.

We, also, provide clear evidence that Beylkin (highly incoherent with SBM and with relatively high number of vanishing moments) leads to the best performance amongst DWT dictionaries evaluated in this paper.

We then compare the performance of the framework while implementing Beylkin as the sparsifying matrix with the case of using DCT and using no dictionary at all for various compression ratios (CR) and block sizes. It is shown that in terms of reconstruction time and accuracy using sparsifying dictionary provides added value in this framework, but only for specific levels of compression and under specific settings.

The paper will be useful for finalizing a framework for online EEG monitoring systems with CS that includes dictionary selection, CRs, block sizes and reconstruction time.

A brief introduction to the theory of CS and STSBL algorithm is in section 2. Materials and methods are presented in section 3, followed by the associated results in section 4, discussions in section 5 and concluding remarks in section 6.

2. Theory

2.1. CS, BSBL and STSBL

In this section, after briefly explaining the basic CS theory irrespective of block sparsity, key formulations regarding BSBL and its modifications leading to STSBL for CS will be discussed.

In CS, a signal of length N , denoted by $\mathbf{x} \in \mathbb{R}^{N \times 1}$, is linearly compressed by a sensing matrix denoted by $\Phi \in \mathbb{R}^{M \times N}$, to yield y (noting $M < N$, hence the word compressed, where N could be the number of samples corresponding to Nyquist rate), which is the measured signal and is given by:

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{v}, \quad (1)$$

where \mathbf{v} is vector representing compression error or CS system noise. \mathbf{x} also may contain noise and may be represented as $\mathbf{x} = \mathbf{u} + \mathbf{n}$, where \mathbf{u} is the clean signal and \mathbf{n} is the signal noise and, subsequently, it is trivial to show:

$$\mathbf{y} = \Phi \mathbf{u} + \mathbf{w}, \quad (2)$$

where $\mathbf{w} = \Phi \mathbf{n}$ [17].

Under certain conditions, described later in this section, this ill-conditioned problem may be solved and signal \mathbf{x} may be reconstructed. A key concept in CS is the sparsity of \mathbf{x} , defined as having only a few non-zero elements. Even if \mathbf{x} is not sparse, one may represent it in a suitable domain in which it exhibits sparsity. This domain may be represented by a dictionary matrix, denoted by $\Psi \in \mathbb{R}^{N \times N}$. Thus, \mathbf{x} can be represented as:

$$\mathbf{x} = \Psi \mathbf{z} \quad (3)$$

where \mathbf{z} contains the coefficients of \mathbf{x} in Ψ domain. Assume \mathbf{x} is K -sparse in this domain (i.e., \mathbf{z} has only $K < N$ non-zero elements; in practice \mathbf{z} may contain K relatively large elements whilst the rest may be ignored, in which case the signal is *compressible* in this domain). Ignoring \mathbf{v} , from (1) and (3) we have:

$$\mathbf{y} = \Phi \Psi \mathbf{z} = \Theta \mathbf{z}. \quad (4)$$

Therefore, for reconstructing the original signal, CS algorithms need to reconstruct \mathbf{z} first using \mathbf{y} and Θ ; subsequently, the original signal \mathbf{x} can be reconstructed at the receiver end.

For successful reconstruction, Θ should follow a condition referred to as restricted isometry property (RIP). RIP may be achieved with high probability if the sensing matrix is random [7]. A condition related to RIP is the incoherence that denotes rows of Φ , $\{\phi_k\}$, and columns of Ψ , $\{\psi_j\}$ should not be correlated. It is noted that M should be sufficiently large. Coherence (μ) is quantified as shown in (5).

$$\mu(\Phi, \Psi) = \sqrt{N} \max_{1 \leq k, j \leq N} |\phi_k, \psi_j|. \quad (5)$$

A smaller μ indicates a lower level of similarity between the elements of the two bases, i.e., Φ and Ψ are highly incoherent. The value of μ is between 1 and \sqrt{N} [7]. The reconstruction performance of CS depends on the level of incoherence between Φ and Ψ [8].

The choice of Φ is directed towards minimal power usage in the hardware in this application and SBM is often used since it consumes very low power [5, 18]. This is because SBM has very few of its entries as ones and most entries are zeros [5]. This reduces the complexity and power requirements as it simplifies the hardware implementation, which is crucial for the design of low-power and efficient transmitters.

The original N datapoints may then be reconstructed from M measurements in CS framework using methods such as basis pursuit with L1 norm minimization [7], which relies on sparsity; thus, as EEG is not sparse in time domain or frequency domain [8], it would be essential to find a suitable Ψ for sparsity while ensuring that it is maximally incoherent with the selected Φ [19]. BSBL based methods, which are of interest in this paper, exploit the block sparsity of signal. A block structured signal \mathbf{x} may be represented as in (6) where g blocks are shown.

$$\mathbf{x} = [x_1, \dots, x_d, \dots, x_{d_{g-1}+1}, \dots, x_d]^\top. \quad (6)$$

For a block sparse signal, only $K \ll g$ blocks are non-zero. If the signal is not block sparse in the original domain, by transforming it into a domain in which it is sparse, block sparsity may ensue. Assuming the EEG signal is transformed using a dictionary in which the signal is sparse or compressible, the coefficient vectors form a concatenation of a number of blocks, only a few of which are non-zero or relatively large and the rest are all zeros or negligible.

The bound optimization method, BSBL-BO, can be employed that assumes the vector it operates on consists of some non-overlapping blocks. The block size can be chosen arbitrarily when using a sparsifying dictionary, and it is not necessary that the block partition of the signal has a clear block structure [5, 20]. Although BSBL-BO is employed successfully for reconstructing single channel EEG signals, for multi-channel signals, signal reconstruction is channel by channel which is time consuming. This increases latency and is not suitable for on-line health monitoring applications. BSBL-BO exploits only the intra-channel correlation of the signal instead of exploiting the inter-channel correlation of the signals from different channels. For exploiting both the intra-channel and inter-channel correlation of the signals, a STSBL method has been proposed in [15]. STSBL reconstructs multichannel EEG signals simultaneously. This exploits temporal correlation in each channel signal and additionally also the spatial correlation among

signals of different channels. Thereby, its computational complexity does not increase with the number of channels [15].

2.2. Wavelet dictionaries

The number of vanishing moments is related to the order, decay rate and smoothness of wavelets. A continuous wavelet (CW), φ , has p vanishing moments when:

$$\int t^k \varphi(t) dt = 0, \text{ for } 0 \leq k \leq p \quad (7)$$

and for the DWT with filter coefficients h

$$\sum_k k^n h(k) = 0, \text{ for } n = 0, 1, \dots, p-1. \quad (8)$$

The number of vanishing moments is the differentiability or a measure of the smoothness of functions. DWT has two functions, called scaling functions and wavelet functions, which are associated with low-pass and highpass filters, respectively. The decomposition of the signal into different frequency bands is obtained by successive highpass and lowpass filtering of the time domain signal. DWT has p vanishing moments if and only if the wavelet function can generate polynomials up to degree $p-1$. The ‘vanishing’ part means that the wavelet coefficients are zero for polynomials of degree at most $p-1$. A higher value of p implies that the wavelet filter is able to filter out high frequency components of the signal accurately from any of the low-frequency or long-term data variations. This accordingly leads to an accurate reconstruction of the signal. CWs and DWTs with a higher value of p can represent more complex functions.

A higher p also increases sparsity of a large class of signals being represented by the DWTs. In most cases, the DWT name is suffixed by its order n . The Daubechies- n and Symlet- n DWTs both have $p = n$ vanishing moments. The number of filter coefficients n_c for the DWTs is $2p$. Their difference lies wherein Symlet filters are as symmetrical as possible as compared to the Daubechies filters which are highly asymmetrical. The Coiflet- n DWT has $p = 2n$ vanishing moments with $n_c = 6n$. The Battle-Lemarie also known as Battle- n DWT generates spline orthogonal wavelet filters, where n is the degree of spline. The Battle- n DWTs have $p = n + 1$. The Battle- n have infinite support but with an exponential decay, and filter coefficients below 10^{-4} are neglected in this paper, giving $n_c = 12$ and 21 for Battle-1 and Battle-3, respectively. The Beylkin is optimised for placement of additional zeros close to half the sampling frequency to for obtaining higher attenuation of high-frequency components for the scaling filter and close to DC for attenuation of the low-frequency components. It has fixed number of filter coefficients $n_c = 18$ and although it has three zeros at $z = -1$ and 1, it has $p \sim 9$. The Vaidyanathan DWT is optimised for speech coding with $n_c = 24$ with additional zeros close to high frequency and DC for the scaling and wavelet filters. It offers accurate

reconstruction of the decomposed signal just as in case of other DWTs including Beylkin but does not satisfy any moment condition. The Haar DWT is the least complex to implement as it has $n_c = 2$, has one zero at $z = -1$ and 1 for the scaling and wavelet function indicating $p = 1$.

3. Materials and methods

3.1. Incoherence of SBM with wavelet dictionaries

As the first step, the number of non-zero entries of the SBM that would lead to a moderate incoherence for all the wavelet dictionaries to be used was identified by calculating the coherence of randomly generated SBM with each dictionary for a varying number of non-zero entries. The fifteen DWT basis considered are Daubechies-3, Daubechies-4, Daubechies-8, Daubechies-10, Symmlet-10, Vaidyanathan, Coiflet-1, Coiflet-2, Coiflet-3, Coiflet-4, Coiflet-5, Harr, Battle-1, Battle-3 and Beylkin of size 256×256 as the Ψ matrix. The result is shown in figure 1. Subsequently, the number of nonzero entries selected was 30.

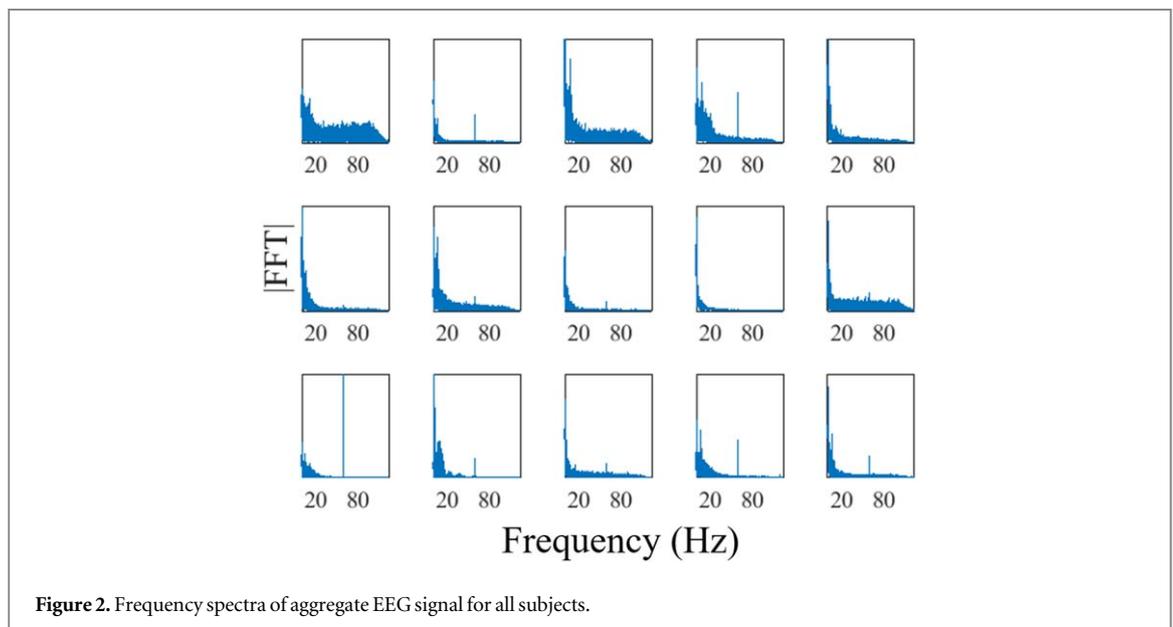
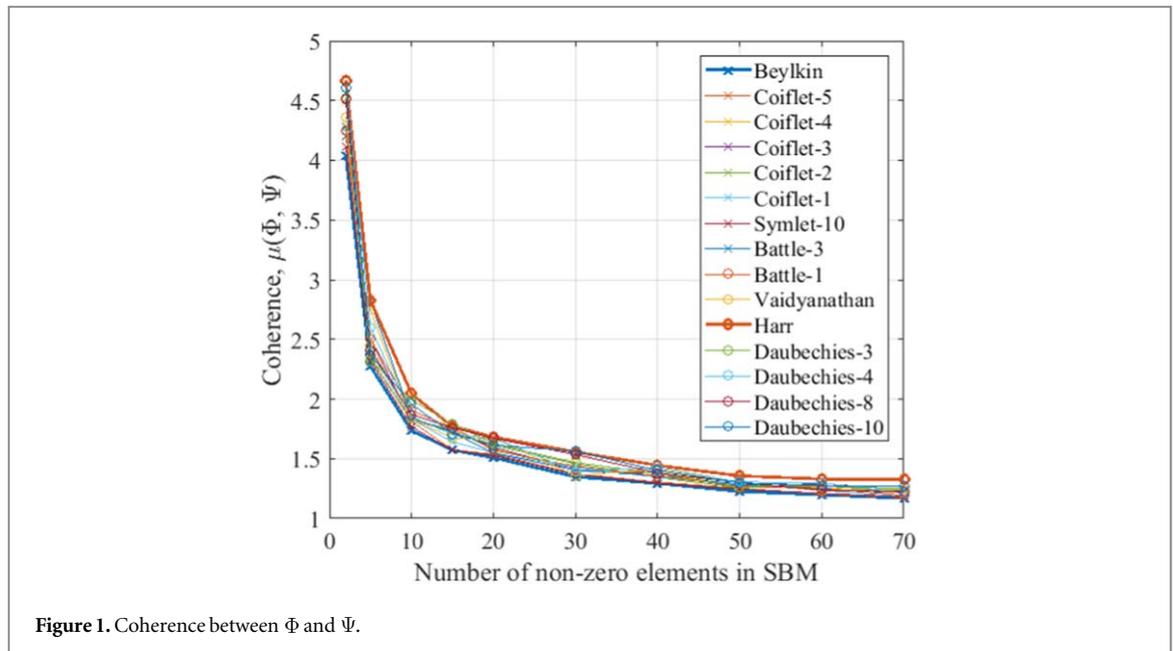
3.2. Reconstruction using wavelet dictionaries

The simulations were undertaken in Matlab[®]2017a on EEG data of 15 subjects involving 10 epileptic and 5 non-epileptic datasets from the Temple University Hospital EEG data corpus [21] with 23 channels containing EEG data selected sampled at 250 samples per second. The signal amplitude typically ranges from about $1 \mu\text{V}$ to $100 \mu\text{V}$ and frequency ranges between 1 Hz – 100 Hz as shown in the fast Fourier spectra of normalised aggregate signal shown in figure 2. To form the spectra shown in figure 2, data points of all 23 channels at a given time were summed to demonstrate the spectra of all channels at the same time. The signals exhibit non-linear, uncorrelated properties and random nature. In processing EEG data in this paper, we considered 256 samples as an epoch. This led to 117 epochs for each subject. The block size used was set to 24 similar to [5].

The reconstruction quality of EEG signals using different DWT dictionary (Daubechies-3, -4, -8, -10, Symlet-10, Vaidyanathan, Coiflet-1, -2, -3, -4, -5, Harr, Battle-1, Battle-3, Beylkin) were compared here using two performance indicators. One is the normalised means square error (NMSE), defined as

$$\|\hat{x} - x\|_2^2 / \|x\|_2^2 \quad (9)$$

where \hat{x} is the estimate of the original signal x . The second is the structural similarity index (SSIM), which measures the similarity between the reconstructed signal and the original signal [6]. Higher value of SSIM indicates better reconstruction. When the reconstructed signal and the original signal are same, $\text{SSIM} = 1$. To compare the performance of the dictionaries in the first instance a 50% CR defined as $\left(\frac{N-M}{N}\right) \times 100$ was used.



The median of NMSE and SSIM for all the epochs associated with a subject was calculated as the measure of center due to the skewed distribution of values across the 117 epochs. The mean and standard deviation of the center were subsequently calculated across the 15 subjects.

3.3. Beylkin, DCT and no dictionary

As will be demonstrated in section 4.1 the best performance is associated to Beylkin dictionary amongst the DWT dictionaries assessed in the paper. The performance was compared with the DCT dictionary as well as the case of using no sparsifying dictionary for different CR values ranging from 50% – 90% and different block sizes (16, 32 and 64) in terms of NMSE, SSIM and reconstruction time. Furthermore, the effect of number of non-zero elements in

SBM on the performance of the framework when using no dictionary was evaluated.

4. Results

4.1. Reconstruction using wavelet dictionaries

Figures 3 and 4 show NMSE and SSIM (bar indicating the mean and error bar showing the standard deviation) of the reconstructed signal (CR = 50%) for all the subjects and for all the 15 DWT dictionaries. Both the NMSE and SSIM indicate a superior performance by Beylkin. Figure 5 shows the scatter plot of coherence versus vanishing moments for all the dictionaries and indicate the correlation these features have (μ and p) with the reconstruction performance (mean of NMSE). The results indicate that those dictionaries that tend to have both high incoherence

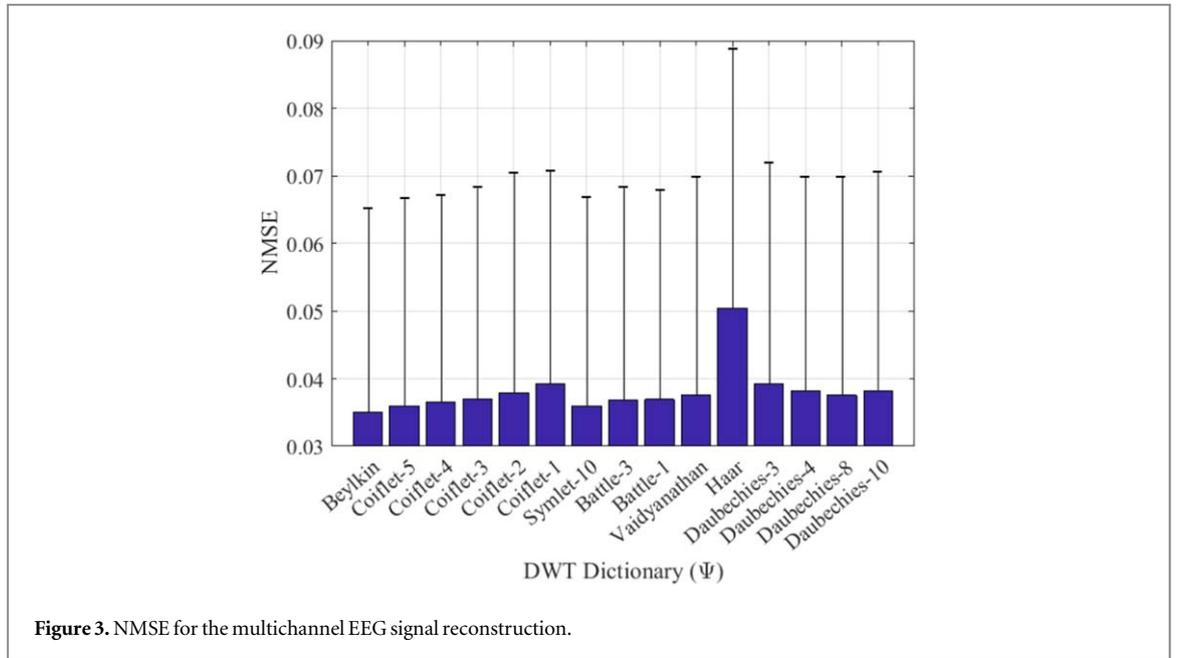


Figure 3. NMSE for the multichannel EEG signal reconstruction.

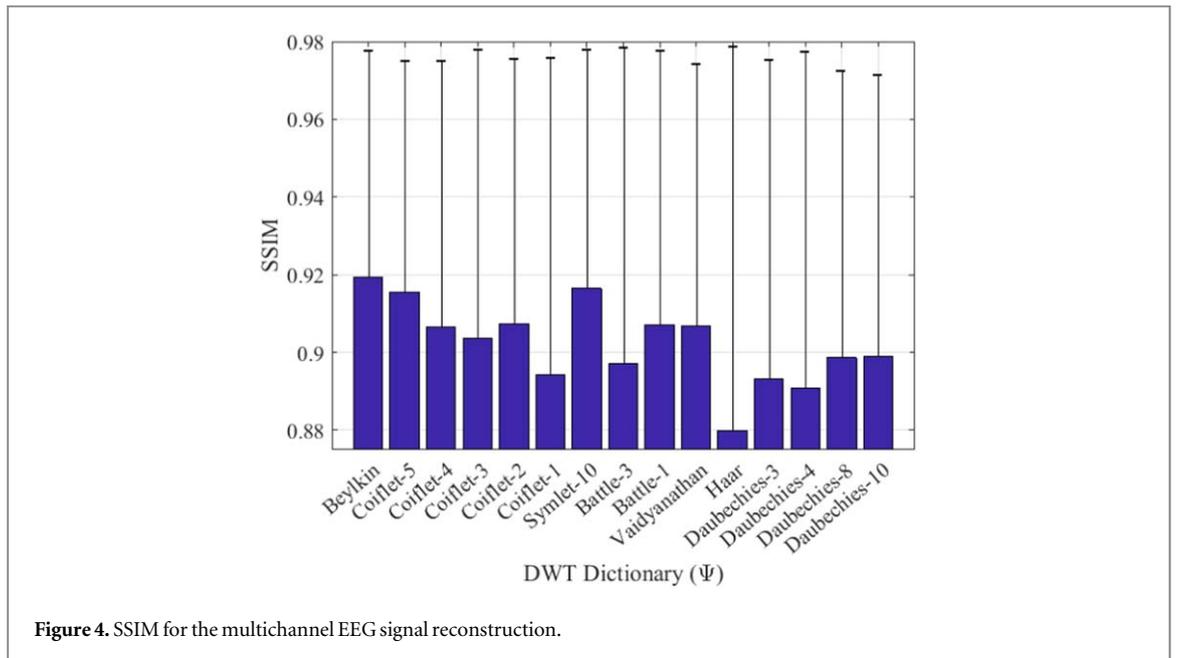


Figure 4. SSIM for the multichannel EEG signal reconstruction.

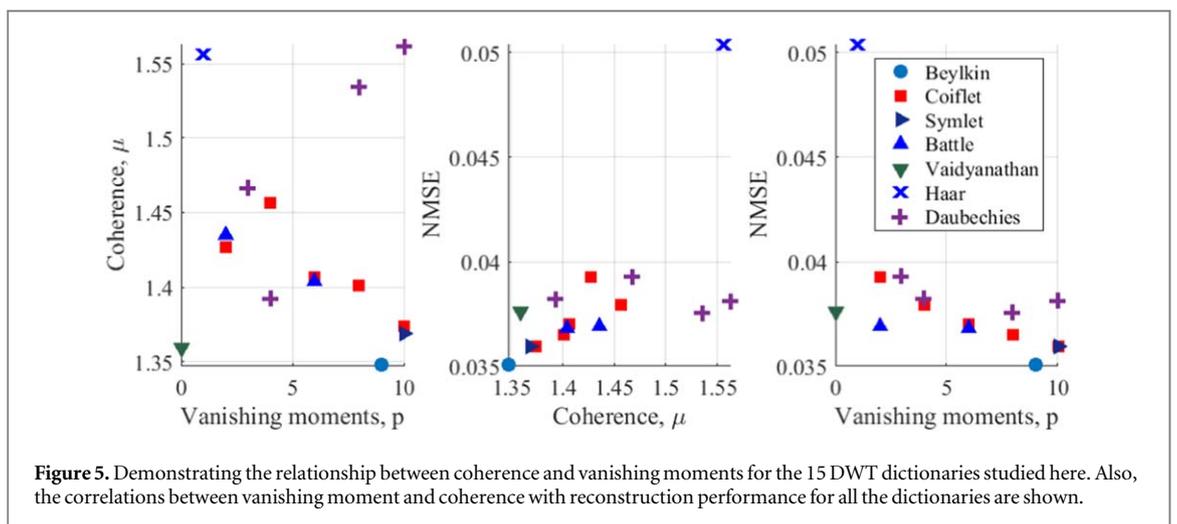
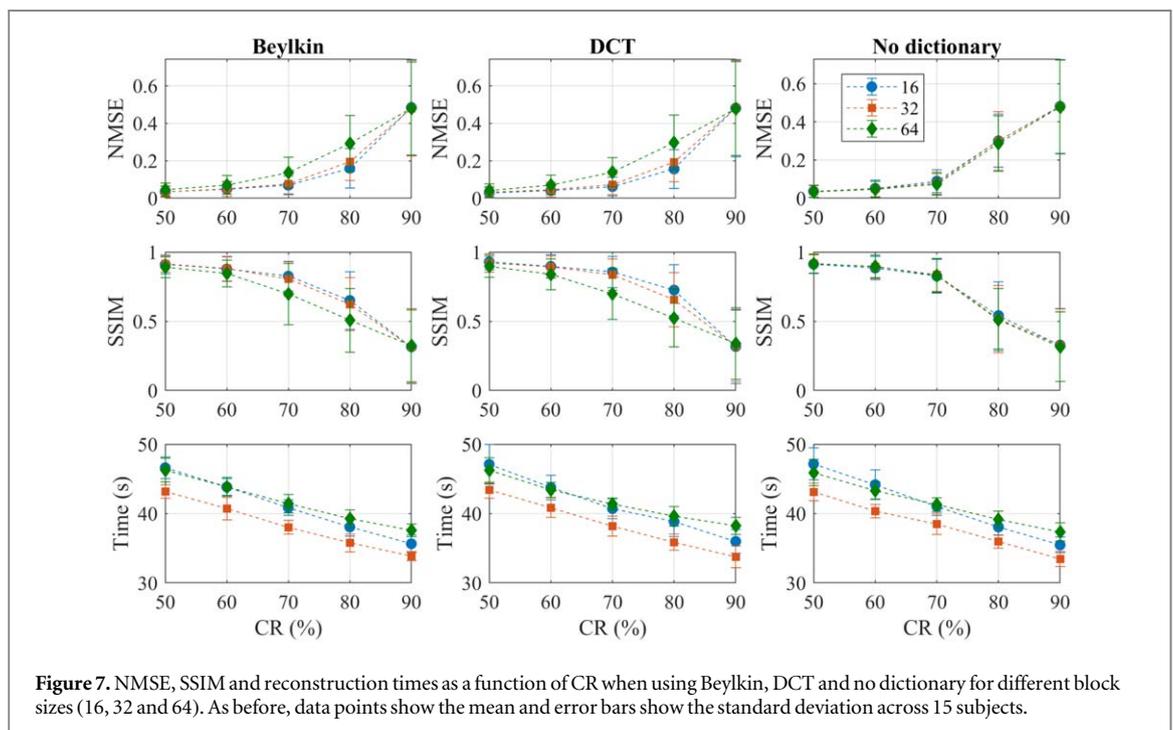
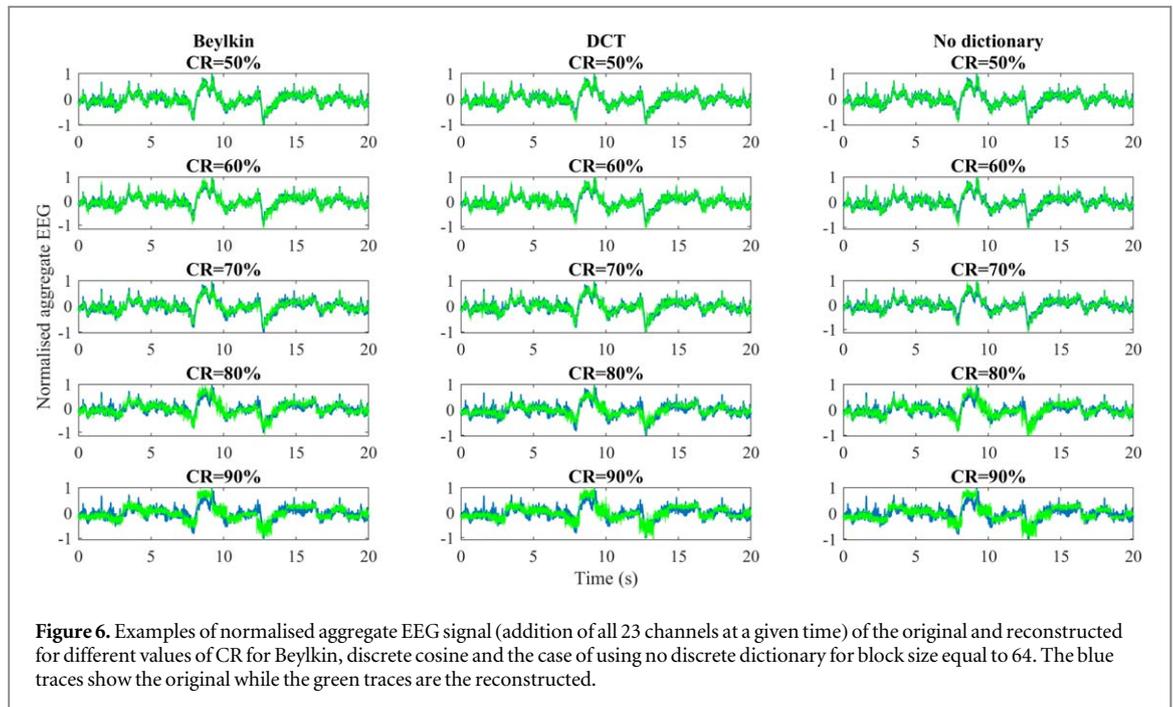


Figure 5. Demonstrating the relationship between coherence and vanishing moments for the 15 DWT dictionaries studied here. Also, the correlations between vanishing moment and coherence with reconstruction performance for all the dictionaries are shown.



and vanishing moments tend to perform better. These show the effect of coherence is more significant when comparing Beylkin with Symlet or Coiflet. That is, Beylkin has higher incoherence with SBM but lower number of vanishing moments compared to these two but the overall performance associated with Beylkin is better.

4.2. Beylkin, DCT and no dictionary

Figure 6 shows an example of aggregate EEG signals (original and reconstructed using different CRs) associated with using Beylkin and DCT as the

ictionaries and using no dictionary at all when the block size was set to 64. It is noted that the reconstruction quality as qualitatively evaluated, based on this figure, appears to be the same for all three cases.

Figure 7 shows NMSE, SSIM and reconstruction time for the three cases of using Beylkin, DCT and no dictionary for various CRs and different block sizes. Larger block size appears to lead to higher errors in reconstruction for Beylkin and DCT for $CR < 90\%$ while when no dictionary is used changing block size does not affect the result. For a block size of 64, the reconstruction time demonstrates a degree of

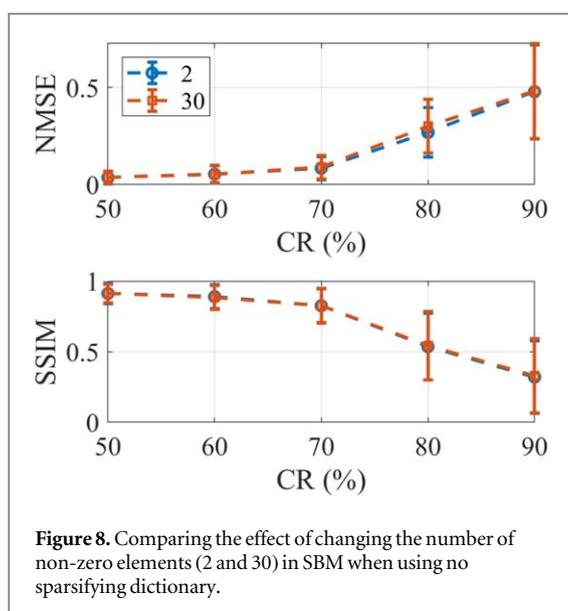


Figure 8. Comparing the effect of changing the number of non-zero elements (2 and 30) in SBM when using no sparsifying dictionary.

nonlinearity with respect to CR. Figure 8 compares the effect of changing the number of non-zero elements (2 and 30) in SBM when using no sparsifying dictionary. It is clear that NMSE and SSIM are not affected by the number of non-zero elements in SBM in this case.

5. Discussion

In recent years, CS has gained considerable attention as a key enabler for transfer of large data rate and volume signals over a sensor network, primarily driven by emerging technologies such as the IoT. The choice of the DWT is normally based on its ability to represent complex signals given by the number of vanishing moments. A higher number of vanishing moments increases sparsity of a large class of signals being represented by the DWTs. However, incoherence with the sensing matrix also needs to be considered as it can affect the quality of reconstructed signal. A high level of incoherence with the sensing matrix is required for accurate reconstruction of the EEG signal with minimal error. The Debauchies DWT is widely employed for most applications as it has a high number of vanishing moments. While Debauchies-10 has an equal number of vanishing moments to Beylkin, Symlet-10 and Coiflet-5, it has one of the lowest incoherence levels with the sensing matrix. The Debauchies-10 DWT produces a lower quality of reconstructed signal with higher errors and lower accuracy in comparison. Although a high number of vanishing moments may indicate an increase in sparsity of a large class of signals, incoherence of the DWT with the sensing matrix is often the only consideration for accurate reconstruction of the EEG signal. To reduce the complexity of implementation among those having similar values of incoherence and vanishing moments, dictionaries with a lower number of filter coefficients can be implemented to minimize

the order of complexity with a view to reducing the power requirements in EEG data transmission.

An interesting demonstration in this paper is that Beylkin and DCT lead to a similar performance quality (DCT only slightly better). Furthermore, using a dictionary only offers tangible improvement for CR = 80% and smaller block sizes. At CR = 90% the error levels and dissimilarity are high to a level that all the plots converge irrespective of block size and whether or not a dictionary is used. Looking at the example data in figure 6, at higher CR levels more high frequency content is lost. Therefore, while NMSE and SSIM gave stringent figures to compare different cases, this comparison cannot necessarily be extended to evaluating clinical outcome. Some applications may only be interested in low frequency events, in which case CR > 80% may lead to an acceptable outcome.

6. Conclusion

In this study we proposed a framework for the selection of a DWT dictionary used in tandem with SBM as the sensing matrix and STSBL method as the reconstruction algorithm. It was demonstrated that in selecting the dictionary its incoherence with the sensing matrix as well as its number of vanishing moments should be considered at the same time. Amongst the DWT dictionaries we studied, Beylkin led to the best performance. This indicates that incoherence presumably has a slightly stronger impact on the outcome based on the methods used in this paper. It was shown in comparing Beylkin, DCT and using no dictionary at all that using a dictionary only leads to improved performance for CR = 80% and for smaller block sizes. Further work could be directed at identifying the exact clinical implications based on specific pathologies. In addition, there have been efforts to develop data-driven schemes for learning the best sparsifying dictionaries as well as using a deep neural network for reconstructing the compressed signals [22–25]. Considering these methods for further benchmarking could be a promising direction for future research.

ORCID iDs

Arsam Shiraz  <https://orcid.org/0000-0003-1732-1401>

Jaswinder Lota  <https://orcid.org/0000-0002-7868-0347>

Andreas Demosthenous  <https://orcid.org/0000-0003-0623-963X>

References

- [1] Hosseini M P *et al* 2017 Optimized deep learning for EEG big data and seizure prediction BCI via internet of things *IEEE Trans. Big Data* **3** 392–404
- [2] Ivanov S *et al* 2012 Virtual groups for patient WBAN monitoring in medical environments *IEEE Trans. Biomed. Eng.* **59** 3238–46

- [3] Lim H B et al 2011 A human body model for efficient numerical characterization of UWB signal propagation in wireless body area networks *IEEE Trans. Biomed. Eng.* **58** 689–97
- [4] Seyed M et al 2013 A survey on intrabody communications for body area network applications *IEEE Trans. Biomed. Eng.* **60** 2067–79
- [5] Zhang Z et al 2013 Compressed sensing of EEG for wireless telemonitoring with low energy consumption and inexpensive hardware *IEEE Trans. Biomed. Eng.* **60** 221–4
- [6] Liu Y et al 2015 Compressed sensing of multichannel EEG signals: the simultaneous cosparsity and low-rank optimization *IEEE Trans. Biomed. Eng.* **62** 2055–61
- [7] Candes E J and Wakin M B 2008 An introduction to compressive sampling *IEEE Signal Process Mag.* **25** 21–30
- [8] Mazumdar A et al 2014 A low-rank matrix recovery approach for energy efficient EEG acquisition for a wireless body area Network *Sensors* **14** 15729–48
- [9] Craven D et al 2015 Compressed sensing for bioelectric signals: a review *IEEE J. Biomed. Health Informat.* **19** 529–40
- [10] Unser M 1996 Vanishing moments and the approximation power of wavelet expansions *Proc. 3rd Int. Conf. on Image Process.* 629–32
- [11] Selesnick I W 2004 The double-density dual-tree DWT *IEEE Trans. Signal. Proc.* **52** 1304–13
- [12] Mahrous H and Ward R 2016 Block sparse compressed sensing of electroencephalogram (EEG) signals by exploiting linear and non-linear dependencies *Sensors* **16** 1–16
- [13] Majumdar A and Ward R K 2015 Energy efficient EEG sensing and transmission for wireless body area networks: a blind compressed sensing approach *Bio. Signal Process. Control* **20** 1–9
- [14] Singh W et al 2017 Energy efficient EEG acquisition and reconstruction for a wireless body area network *J. Integration VLSI* **58** 285–302
- [15] Zhang Z et al 2014 Spatiotemporal sparse bayesian learning with applications to compressed sensing of multichannel physiological signals *IEEE Trans. Neural Syst. Rehabil. Eng.* **22** 1186–97
- [16] Mishra A et al 2012 Comparative analysis of wavelet basis functions for ECG signal compression through compressive sensing *Int. J. of Computer Sci. Telecommun.* **3** 23–31
- [17] Zhang Z et al 2013 Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ECG via block sparse bayesian learning *IEEE Trans. Signal Process.* **60** 300–9
- [18] Mamaghanian H et al 2011 Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes *IEEE Trans. Biomed. Eng.* **58** 2456–66
- [19] Pereira M P et al 2014 On the design of maximally incoherent sensing matrices for compressed sensing using orthogonal bases and its extension for biorthogonal bases case *J. Digit. Signal Process.* **27** 12–22
- [20] Zhang Z and Rao B D 2013 Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation *IEEE Trans. Signal Process.* **61** 2009–15
- [21] Veloso L 2017 Big data resources for EEGs: enabling deep learning research *Proc. IEEE Signal Process. Med. Biol. Sym. (SPMB)* **10.1109/SPMB.2017.8257044**
- [22] Hamner B, Chavarriaga R and Millan J D R 2011 Learning dictionaries of spatial and temporal EEG primitives for brain-computer interfaces *Presented at Workshop on Structured Sparsity: Learning and Inference, Int. Conf. Mach. Learning (ICML 2011)*
- [23] Majumdar A and Ward R 2016 Real-time reconstruction of EEG signals from compressive measurements via deep learning' *Proc. Int. Joint Conf. Neural Netw. (IJCNN)* **2856–63**
- [24] Palangi H, Ward R and Deng L 2016 Exploiting correlations among channels in distributed compressive sensing with convolutional deep stacking networks *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)* **2692–6**
- [25] Tour L et al 2018 Multivariate convolutional sparse coding for electromagnetic brain signals *arXiv preprint arXiv: <https://arxiv.org/abs/1805.09654>*