

MDL-BASED JOINT DENOISING AND COMPRESSION OF INTRACORTICAL SIGNALS

Elias S.G. Carotti^{*†}

Winnie Jensen^{*}

Juan Carlos De Martin^{*†}

Dario Farina[†]

^{*†} Dipartimento di Automatica ed Informatica
Politecnico di Torino
Corso Duca degli Abruzzi, 24
I-10129 Torino, Italy
[carotti|demartin]@polito.it

^{*} Center for Sensory-Motor Interaction
Dep. of Health Science and Tech.
Aalborg University
Aalborg, Denmark
wj@hst.aau.dk

[†] Dep. of Neurorehabilitation Engineering
Bernstein Center for Computational Neuroscience
Georg-August University
Von-Siebold-Str. 4, - 37075 Gottingen, Germany
dario.farina@bccn.uni-goettingen.de

ABSTRACT

Intra-cortical signals are usually affected by high levels of noise (0 dB SNR is not uncommon) either due to the recording equipment or to magnetical and electrical couplings between surrounding sources and the recording system. Besides from hindering effective exploitation of the information content in the signals, noise also influences the bandwidth needed to transmit them, which is a problem especially when a large number of channels are to be recorded. In this paper we propose a novel technique for joint denoising and compression of intra-cortical signals based on the Minimum Description Length principle (MDL). This method was tested on simulated signals and the results showed that the proposed technique achieves improvements in SNR (up to .6 dB over MNML for very noisy signals) and compression ratios greater than alternative denoising/compression methods.

Index Terms— Signal denoising, biomedical signal processing, data compression

1. INTRODUCTION

Brain-computer interfacing (BCI) has the potential of assisting in neurological rehabilitation for patients suffering from e.g. amyotrophic lateral sclerosis, spinal cord or brain injury or brain stem stroke, which all determine severe motor deficits. A BCI system is a method for these patients to provide a link to the outside world and thereby allow them to regain control of the external environment [1].

Intra-cortical recordings of single-unit, multi-unit or local field potentials have shown that the primary motor cortex (M1) encodes information about limb position, limb velocity, muscle activity, and movement preparation (e.g., see [2].)

Over the last decade, animal experiments have shown that information from the motor cortex can be extracted to reliably control a one dimensional lever arm in rats [3] or closed-loop control of a 3-D robotic arm in monkeys (see e.g., [4, 5]). Also, human subjects have learned to control a computer cursor after implantation of microelectrodes.

Despite the major advances in the field, a number of technical challenges must be addressed to bring intra-cortical BCI systems

from the research environment into long-term usage in clinics. One important challenge is that a fully implantable system should be able to incorporate a high bandwidth (intra-cortical signals are typically sampled with a frequency in the range of 8-10 kHz to provide a sufficient temporal resolution for identifying single spikes in the signals), high-channel count (≥ 100 channels) wireless communication pathway between the body and the output, which puts large requirements on the transmission bandwidth. The transmission bandwidth should thus be reduced by compressing the intra-cortical signals before wireless transmission or before/after processing of the signals.

The amplitude of recorded extra-cellular, single unit action potentials depends primarily on the distance between the cell and the electrode active site, whereas the quality of the signal primarily depends on the amount of noise in the recordings. To optimize the recorded intra-cortical signal quality for spike sorting and classification, denoising should be performed.

This paper addresses these two issues (compression and denoising) by proposing a method for joint compression and denoising of intracortical signals. The method is tested on simulated signals to assess its performance with respect to standard state-of-the art denoising techniques.

Paper organization is as follows: first, the proposed joint-denoising and compression algorithm is introduced in section 2, then the experimental setup consisting of simulated signals is discussed in section 3, then results are presented in section 4, and finally, conclusions are drawn in section 5.

2. MDL-BASED JOINT DENOISING AND COMPRESSION

In this paper we propose a novel technique for joint denoising and compression of intracortical signals, based on the Discrete Wavelet Packet Transform (DWPT), optimal mother wavelet selection and the MDL criterion. Fig. 2 presents the block diagram of the proposed technique.

First, an orthogonal transform is applied to the signal, then denoising based on various Minimum Description Length (MDL) criteria is performed, and finally, compression of the denoised transformed coefficients is performed. Each stage, if performed independently, usually implies various design choices, such as the selection

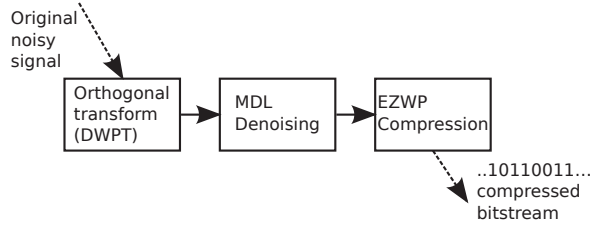


Fig. 1. Block diagram of the proposed technique for joint denoising and compression of intracortical signals.

of the optimal tree in the wavelet packet. Instead, we propose the optimization of all the parameters and thresholds based on the same MDL criterion (modified for each purpose), so that a unique criterion is used for transforming, denoising and compressing. Therefore, we adapted the basic MDL criterion to the various selection procedures.

Several denoising techniques rely on some form of discrete wavelet transform, where the signal is represented by means of the inner products with basis functions which are temporal shifts and dilatation of a function called mother wavelet. Often, the mother wavelet is chosen within a library of well known wavelets. However, it is expected that different mother wavelets provide different performance depending on the signal characteristics. The lattice parametrization described by Vaidyanathan [6] offers the opportunity to design orthogonal wavelet filters via unconstrained parameters, and had already been applied to compression of biomedical signals in [7], where, however denoising was not considered.

In addition to optimal selection of the mother wavelet, we also propose the use of a Discrete Wavelet Packet Transform (DWPT) which provides a better adaptation than a dyadic transform (DWT) to the specific signal characteristics. The first step of the proposed method is thus based on optimization of a DWPT and of the thresholds for coefficient selection based on criteria discussed in the following.

2.1. The MDL criterion

The underlying assumption of the MDL criterion is that the signal y^n is modeled as the linear combination of the transform basis vectors plus Gaussian i.i.d. white noise ϵ^n :

$$y^n = W\beta^n + \epsilon^n, \epsilon \in N(0, \sigma_N^2),$$

where σ_N^2 is the noise variance. Given the observed signal, and an orthogonal transform, the aim is to find the proper coefficient vector $\tilde{\beta}^n$, such that the difference $\tilde{y}^n - y^n$ between the reconstructed signal \tilde{y}^n and the (not observable) original signal contains most of the noise.

For this purpose, most wavelet-based denoising algorithms either perform hard-thresholding or soft-thresholding of the transform coefficients. The former case can be seen as an implicit classification of each coefficient into two classes, one for pure noise, which is discarded, and the other of information plus noise, which is retained. Soft-thresholding implies that a constant value is subtracted from each coefficient. Different methods have been proposed for thresholding with the aim of denoising [8, 9, 10, 11]. Recently, the MDL criterion has been successfully applied to the denoising problem [12]. The MDL criterion consists in comparing different models in a model class and choosing the model that yields the shortest overall description of the data along with the description of the model. The length of the description corresponds, for probabilistic models,

to the negative logarithm of the probability. The key concept is that a model which best compresses the data, i.e., yields the shortest overall codelength, is the one that learns most of the data. The shortest codelength, given a class of models (a set of distributions), is called the stochastic complexity, which is the term of comparison for different model classes; in modern MDL, it is defined by means of the Normalized Maximum-Likelihood (NML) [13]. The NML universal model for a given model class, i.e., a parametric distribution with parameters $\hat{\theta}(y^n)$, is given by:

$$f_{\text{nml}}(y^n) = \frac{f(y^n; \hat{\theta}(y^n))}{\int_A f(z^n; \hat{\theta}(y^n)) dz^n},$$

where the range of integration A can be either the set of all possible sequences of length n or a subset of them, and $\hat{\theta}(y^n)$ is the maximum likelihood estimate of the parameters. The NML is used to evaluate the stochastic complexity for the different model classes and to select the one achieving its minimum cost.

2.2. Denoising criterion

Denoising by means of hard-thresholding is equivalent to a classification task on the transformed coefficients, where k coefficients are retained and the remaining $n-k$ are considered as pure noise. Different models have been proposed in literature, but we adopted the per band version of the MDL cost, as described in [12], where the signal coefficients for each band are modeled with a different per band Gaussian distribution, while all the remaining noise coefficients are assigned a single Gaussian, under the assumption that noise influences each band in the same way. The stochastic complexity then becomes:

$$\sum_{b=0}^B \left(\frac{k_b}{2} \ln \frac{S_{\gamma_b}(y^n)}{k_b} + \frac{1}{2} \ln k_b \right) + \sum_{b=1}^B \ln \binom{n_b}{k_b}, \quad (1)$$

where B is the number of bands, $S_{\gamma_b}(y^n)$ is the sum of the squared coefficients in the band b , γ_b is the set of coefficients either considered to be noise or signal, and k_b is their number.

It was shown [13] that for orthonormal regression matrices the index set that minimizes the criterion is given either by the k largest coefficients in absolute value or the k smallest ones, which implies that only n evaluation of the criterion needs to be performed. Often more than one model yield good performance and it is therefore better to properly weight the contributions of different models; this amounts to considering a *mixture* over all the models indexed by γ :

$$f_{\text{mix}}(y^n) = \sum_{\gamma} f_{\text{mix}}(y^n; \gamma) \pi(\gamma). \quad (2)$$

Soft thresholding is performed by retaining all the coefficients weighted by their contribution to the mixture prediction [12].

2.3. Denoising and Compression with Optimal Wavelet Packets

The DWPT performs an adaptive decomposition of the frequency axis. The specific decomposition (pruned wavelet packet tree) may be selected according to an optimization criterion. Thus, with the adoption of the DWPT it is necessary not only to choose the proper subset of coefficients to minimize the MDL cost, but also to identify the mother wavelet and the wavelet tree more suited for denoising. Unfortunately, due to the great number of possible decomposition

trees it is not feasible to extensively enumerate all of them and explicitly evaluate the model cost for each one, unless the number of decomposition levels is very small. The latter condition is practically never met.

However, if the cost function is additive, i.e., can be evaluated independently for each subband, it is possible to avoid complete enumeration of the trees and still achieve the globally optimal cost. Therefore, several cost functions have been proposed in the literature to prune the wavelet packets tree, depending on the specific application. Among them, a popular choice is Coifmans Entropy[8] $M(\bar{x}) = -\sum_j |x_j|^2 \log |x_j|^2$, where the cost function is given by the sum over all coefficients of minus the logarithm of each squared coefficient. However, since we use the MDL cost function to select the coefficients to retain, it is expected that better performance can be achieved if the tree pruning rule is explicitly designed using the same cost function. Unfortunately, the MDL cost defined in Eq. (1) is not additive, because of the noise term. We therefore propose a modification of the MDL cost function for tree selection that makes the cost function additive.

If we relax the constraint that the noise power is the same on each band, and instead we assume the noise power to vary across different bands, so that the MDL cost is minimized on a per-band basis, then this new cost function becomes additive because the global noise term decomposes into a sum of per band terms, each accounting for the noise coefficients in each band. Albeit suboptimal, it makes the problem computationally tractable, because the global cost can be recursively decomposed into a sum of partial costs for each node of the wavelet packets decomposition tree. For each basis, the cost in terms of bit needed to transmit the tree and the transform coefficients should be taken into consideration and added to the total cost. However, the cost needed to transmit the basis can be further divided into the cost needed to transmit the tree and the basis coefficients. While the latter can be approximately considered constant across different bases, the cost of the tree depends on its depth and requires 2 bits for each node to determine if the node is a terminal node (leaf) or an intermediate node. Once a tree has been selected, denoising can be performed either by means of the hard or soft thresholding, as in Eq. (1) and Eq. (2).

The denoised coefficients are then compressed for transmission. In principle, if hard thresholding is performed (i.e., using Eq. (1)), the MDL immediately provides a way to perform compression, i.e., one could simply code the choice of the mother wavelet and the tree, and the position of the preserved coefficients along with the quantized coefficients (the latter may still imply the need to describe the bit allocation across the subbands and the quantizers if they are not known at the decoder). However, in the case of soft mixture-based denoising Eq. (2), most of the coefficients are preserved to a certain degree, so a regular wavelet-based compression technique, such as the EZW[9] is more convenient. EZW has already been adapted to biomedical signal compression with wavelet packets in previous work[7], where, however, denoising (and its influence on the choice of the wavelet packets tree) had not been considered.

3. SIMULATIONS

For testing the performance of the proposed method, we simulated 10 realizations of intracortical signals for three levels of activity[14]. The levels of activity were obtained by randomly placing a total of 60, 120, or 250 spikes per second (on average). Two scenarios were considered, where additive gaussian noise was added to the signals corresponding to an SNR (see Eq. (3), below) of 0 dB and 10 dB. The performance was measured in terms of the signal to noise ratio

Technique	60 spikes/s (dB)	120 spikes/s (dB)	250 spikes/s (dB)
No denoising	-0.04	-0.04	-0.04
VisuSHRINK	7.99	5.56	3.59
BayesSHRINK	8.51	7.89	7.22
SNML	9.66	7.67	5.93
MNML	10.27	8.42	6.72
Proposed technique	10.81	9.01	7.22

Table 1. Average denoising performance for the tested techniques over three sets (60,120,250 spikes/s, 24 KHz) each consisting of 10 simulated signals with additive gaussian white noise (0 dB SNR).

Technique	60 spikes/s (dB)	120 spikes/s (dB)	250 spikes/s (dB)
No denoising	9.96	9.96	9.96
VisuSHRINK	16.21	13.85	11.93
BayesSHRINK	14.85	14.53	14.27
SNML	17.95	15.97	14.41
MNML	18.55	16.57	15.06
Proposed technique	18.93	16.97	15.25

Table 2. Average denoising performance for the tested techniques over three sets (60,120,250 spikes/s, 24 KHz) each consisting of 10 simulated signals with additive gaussian white noise (10 dB SNR).

(SNR) with respect to the original signal:

$$\text{SNR} = 10 \log_{10} \left(\frac{\sum_{i=1}^N x_i^2}{\sum_{i=1}^N (\hat{x}_i - x_i)^2} \right) \text{dB} \quad (3)$$

where x_i and \hat{x}_i are the i -th sample of the original and the reconstructed signals, respectively, while N is their total length. The following methods were compared: the VisuShrink[10], the BayesShrink[11], the two best performing MDL based denoising algorithms from [12], and the technique proposed in this paper. For the latter, both Daubechies and optimal mother wavelets were tested, either performing hard or soft denoising.

4. RESULTS

Table 1 presents results for the first simulation scenario, the 0dB SNR signals. Blind denoising was then performed and the performance was measured against the pristine clean simulated signals. In all cases, the proposed wavelet packets denoising with mother wavelet optimization substantially increased the SNR with respect to the original signal and performed better than the other methods tested. Table 2 presents similar results for the second simulation scenario, with less severe noise, where the noisy signal SNR was 10 dB.

Fig.2 depicts the average performance in terms of SNR with respect to the original noiseless signal for the tested wavelet packets based denoising techniques followed by EZWP compression over 0 dB SNR noisy signals, while Fig. 3 shows the results on the same signals under less severe noise (10 dB SNR). Both figures depict the performance averaged over 10 realizations of simulated signals with additive Gaussian noise for different activity levels, i.e., 60, 120, and 250 spikes per second on average.

In all the tested cases, the compression factors higher than 4% became virtually lossless and provide almost no benefit in terms of SNR. It is also noteworthy to say that, in general, soft thresholding outperformed hard thresholding, however, for smaller compression

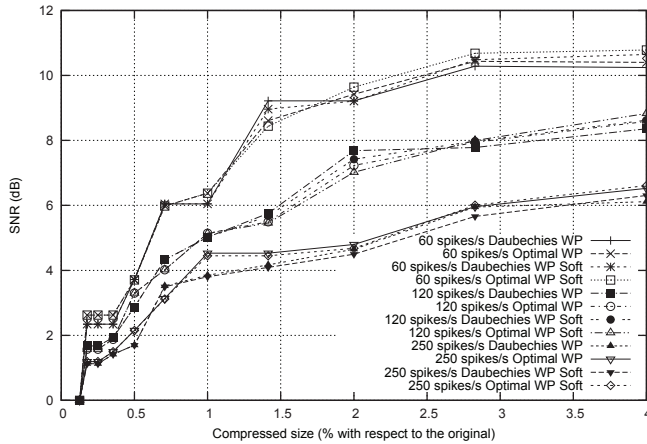


Fig. 2. Average performance (over 10 realizations) in terms of SNR with respect to the original uncompressed signals. Each algorithm was tested on 10 realizations of a low, medium, and high activity level neural signal: 60, 120, and 250 spikes/s respectively, 24 KHz. SNR prior to denoising: 0 dB.

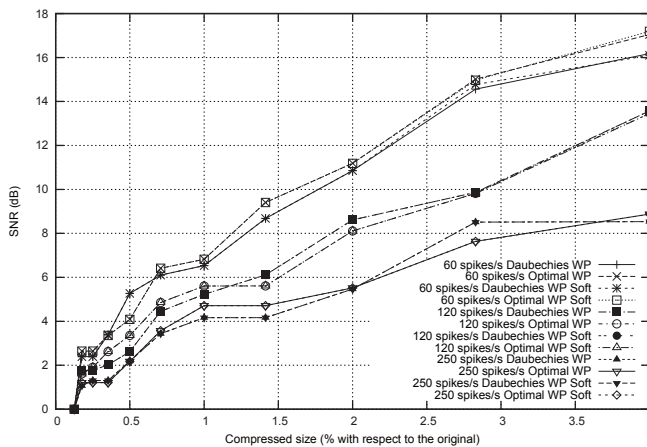


Fig. 3. Average performance (over 10 realizations) in terms of SNR with respect to the original uncompressed signals. Each algorithm was tested on 10 realizations of a low, medium, and high activity level neural signal: 60, 120, and 250 spikes/s respectively, 24 KHz. SNR prior to denoising: 10 dB.

factors it may happen that hard thresholding performs best, probably due to the higher number of insignificant coefficients introduced.

5. CONCLUSION

We have proposed a joint denoising and compression method based on the MDL criterion. Results on synthetic signals showed that the best combination in terms of quality of the reconstruction after joint denoising and compression is consistently achieved for all the activity levels by means of mother wavelet optimization and soft denoising.

Future work includes assessing the performance of the proposed technique on experimental signals, where, however, the SNR of the pristine noisy signal is unknown and has to be estimated somehow.

The proposed method allows high compression ratios with joint denoising which is a necessary step in fully implanted BCI systems.

6. REFERENCES

- [1] N.G. Hatsopoulos and J.P. Donoghue, "The science of neural interface systems," *Annual Review of Neuroscience*, vol. 32, pp. 249–266, 2009.
- [2] J.P. Donoghue, J.N. Sanes, N.G. Hatsopoulos, and G. Gaál, "Neural discharge and local field potential oscillations in primate motor cortex during voluntary movements," *Journal of Neurophysiology*, vol. 79, no. 1, pp. 159–173, Jan. 1998.
- [3] J.K. Chapin, K.A. Moxon, R.S. Markowitz, and M.A.L. Nicolelis, "Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex," *Nature neuroscience*, vol. 2, no. 7, pp. 664–670, July 1999.
- [4] J. Wessberg, C.R. Stambaugh, J.D. Kralik, P.D. Beck, M. Laubach, J.K. Chapin, J. Kim, S.J. Biggs, M.A. Srinivasan, and M.A.L. Nicolelis, "Real-time prediction of hand trajectory by ensembles of cortical neurons in primates," *Nature*, vol. 408, no. 6810, pp. 361–365, Nov. 2000.
- [5] D.M. Taylor, S.I.H. Tillery, and A.B. Schwartz, "Direct cortical control of 3d neuroprosthetic devices," *Science*, vol. 296, no. 5574, pp. 1829–1832, June 2002.
- [6] P.P. Vaidyanathan, *Multirate systems and filter banks*, Prentice Hall PTR, 1993.
- [7] L. Brechet, M.-F. Lucas, C. Doncarli, and D. Farina, "Compression of biomedical signals with mother wavelet optimization and best-basis wavelet packet selection," *Biomedical Engineering, IEEE Transactions on*, vol. 54, no. 12, pp. 2186–2192, Dec. 2007.
- [8] R.R. Coifman and M.V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 713–718, Mar. 1992.
- [9] J.M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.
- [10] Donoho, David L. and Johnstone, Jain M., "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, Dec. 1995.
- [11] S.G. Chang, Bin Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, Sept. 2000.
- [12] T. Roos, P. Myllymaki, and J. Rissanen, "MDL denoising revisited," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3347–3360, Sept. 2009.
- [13] J. Rissanen, "MDL denoising," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2537–2543, 2000.
- [14] R.Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering," *Neural computation*, vol. 16, no. 8, pp. 1661–1687, Aug. 2004.