

A Wavelet Packet-Based Algorithm for the Extraction of Neural Rhythms

OSBERT C. ZALAY,¹ EUNJI E. KANG,^{1,2,3} MARIJA COTIC,^{1,3} PETER L. CARLEN,^{3,4} and BERJ L. BARDAKJIAN^{1,2}

¹Institute of Biomaterials and Biomedical Engineering, University of Toronto, 164 College Street, Rosebrugh Building, Room 407, Toronto, ON M5S 3G9, Canada; ²Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, ON M5S 3G4, Canada; ³Division of Fundamental Neurobiology, Toronto Western Research Institute, 399 Bathurst Street, Toronto, ON M5T 2S8, Canada; and ⁴Department of Physiology, University of Toronto, 1 King's College Circle, Toronto, ON M5S 1A8, Canada

(Received 29 February 2008; accepted 30 December 2008; published online 14 January 2009)

Abstract—Neural rhythms are associated with different brain functions and pathological conditions. These rhythms are often clinically relevant for purposes of diagnosis or treatment, though their complex, time-varying features make them difficult to isolate. The wavelet packet transform has proven itself to be versatile and effective with respect to resolving signal features in both time and frequency. We propose a signal analysis technique, called neural rhythm extraction (NRE) that incorporates wavelet packet analysis along with a threshold-based scheme for separating rhythmic neural features from non-rhythmic ones. We applied NRE to rat *in vitro* intracellular recordings and human scalp electroencephalogram (EEG) signals, and were able to isolate and classify individual neural rhythms in signals containing large amplitude spikes and other artifacts. NRE is capable of discriminating signal features sharing similar time or frequency localization, as well as extracting low-amplitude, low-power rhythms otherwise masked by spectrally dominant signal components. The algorithm allows for independent retention and reconstruction of rhythmic features, which may serve to enhance other analysis techniques such as independent component analysis (ICA), and aid in application-specific tasks such as detection, classification or tracking.

Keywords—Wavelet packet transform, Feature extraction, Brain rhythms, EEG.

INTRODUCTION

Neural signals are temporally and spectrally complex, requiring specialized techniques for their analysis.³⁴ Despite the popularity of Fourier transform-based methods, such methods are ill-suited to deal with nonstationary signals. As an alternative, the wavelet transform has proven itself to be versatile and effective,

with improved ability to resolve signal features in both time and frequency.

In the continuous wavelet transform (CWT), the data signal is convolved with a mother wavelet that is infinitely scalable and translatable.¹⁰ This generates an arbitrary wavelet basis for signal representation. The resulting analysis is computationally intensive with inherent redundancy in the wavelet coefficients, and exact reconstruction of the signal is not possible for non-orthogonal bases. The discrete wavelet transform (DWT) eliminates redundancy by using an orthogonal or biorthogonal wavelet basis, usually with dyadic scaling and translation.³⁰ The basis is used to formulate a finite impulse response (FIR) quadrature mirror filter-bank for signal decomposition and reconstruction.^{10,30} (In the case of biorthogonal wavelets, the decomposition basis is the dual of the reconstruction basis.) The DWT therefore allows for multiresolution analysis, which is superior to conventional filtering methods for isolating rhythms, because multiresolution analysis subdivides and covers the time–frequency plane, whereas single-stage FIR or infinite impulse response (IIR) filters are limited in bandwidth and resolve frequencies but not temporal features. Additionally, IIR filters implemented causally are not generally suitable because of their nonlinear phase response, and their stability must also be ensured.

The discrete wavelet packet transform (DWPT) is a generalization of the DWT, where the scaled frequency sub-bands attributed to the DWT are further parsed, generating a full binary tree decomposition structure.⁸ This offers the greatest depth of analysis. By applying an appropriate cost functional, an optimal basis for signal representation can be selected,⁹ one that customizes the tiling and coverage of the time–frequency plane to suit the intended application.¹⁸ The versatile resolution of the DWPT and efficiency of minimal cost representations has spurred many applications, including data compression,^{31,37} signal and image denoising,^{35,37} and feature detection.^{1,16}

Address correspondence to Berj L. Bardakjian and Osbert C. Zalay, Institute of Biomaterials and Biomedical Engineering, University of Toronto, 164 College Street, Rosebrugh Building, Room 407, Toronto, ON M5S 3G9, Canada. Electronic mail: berj@cbl.utoronto.ca, oz.zalay@utoronto.ca

We propose a method, called neural rhythm extraction (NRE), that makes use of the DWPT in a wavelet packet optimization scheme similar to denoising, but instead of eliminating random fluctuations from the data, it selectively finds and extracts rhythmic features. As proposed, NRE offers dual functionality because the extraction process not only removes rhythms, it also serves to isolate and preserve those same features so that they can be used in other signal-processing-related tasks such as detection, classification or tracking. For example, NRE may be used to decompose neural signals recorded intracellularly or from field potentials, and extract out the component rhythms in order that they may be classified according to their time–frequency characteristics. Additional techniques such as independent component analysis (ICA) can then be applied to further separate rhythmic artifacts (such as cardiac or electrical noise artifacts) from neural source rhythms.

Many cortical and hippocampal networks generate complex rhythms related to their neurological function.^{4–6,27} Because these rhythms are variable in time and frequency, bandpass or notch filtering is often insufficient; instead, the rhythms must be separated from other signal features that share the same time or frequency localization, which is readily accomplished with the DWPT. Large amplitude spikes (LAS) are also common features in neural signals, whether they are generated spontaneously or as a byproduct of stimulation. If LAS are associated with an underlying rhythm and require removal, then the DWPT alone is not capable of eliminating all spike remnants, since spikes with large amplitude in relation to the rest of the signal tend to dominate the coefficients of the wavelet transform across all frequency bands analyzed. Therefore, the signals should first be “despiked” using an appropriate method that attenuates the spectrally dominant spike features, yet leaves the remainder of the signal intact. With NRE, a spike attenuation stage is implemented before wavelet packet analysis.

In this paper, we demonstrate the efficacy of NRE on synthesized test signals, on an intracellular recording from a rat hippocampal slice preparation exhibiting rhythmic LAS, and on human scalp EEGs. In the case where rhythmic artifacts were extracted along with neural signals, independent component analysis was performed to further classify and separate these features.

METHODS

In Vitro Intracellular Recordings

Brain slices were prepared from 15 to 21 day-old Wistar rats. Animals were anesthetized with halothane

and decapitated in accordance with the guidelines of the Animal Care Committee. The brain was quickly removed and placed in ice-cold, continuously oxygenated (95% O₂ and 5% CO₂) artificial cerebrospinal fluid (ACSF) for 2–4 min before the brain slices were prepared. ACSF contained (in mM): 125 NaCl, 26 NaHCO₃, 2.5 KCl, 1.8 CaCl₂, 2 MgCl₂, 1.25 NaH₂PO₄, and 10 glucose. Hippocampal slices were prepared at the room temperature as follows: the brain was hemisected along the midsagittal line and the cerebellum and the forebrain were removed. The dorsal cortex then was cut parallel to the longitudinal axis and the brain was fixed, ventral side up, to an aluminum block with cyanoacrylate glue. The aluminum block was secured at a 12° angle in a Vibratome (Series 1000, Technical Products International, St. Louis, MO) with the caudal end of the brain facing the blade. Slices (500 μm thick) were incubated for at least 1 h before being transferred to an interface-type chamber for electrophysiological recording. Standard ACSF was used for incubation and for perfusion of slices in the recording chamber. Once the recording slice was stabilized, the perfusion solution was changed to low-Mg²⁺ ACSF (125 NaCl, 26 NaHCO₃, 5 KCl, 1.8 CaCl₂, 0.5 MgCl₂, 1.25 NaH₂PO₄, and 10 glucose (in mM)). ACSF for perfusion of slices was continuously aerated with carbogen (95% O₂ and 5% CO₂) and the slice in the recording chamber was also aerated with humidified carbogen. The temperature of the recording chamber was regulated to 33 °C. Intracellular whole-cell patch recordings at a sampling rate of 10 kHz were made from CA3 pyramidal neurons using borosilicate glass tubing filled with a solution containing (in mM): 135 K-Glu, 10 HEPES, 5 KCl, and 2 Mg-ATP (275 ± 5 mOsm, pH 7.2 adjusted with KOH). All chemicals used in the experiments were purchased from Sigma–Aldrich Canada Ltd. (Oakville, ON, Canada).

Electroencephalography

Human EEG recordings were performed using Ag/AgCl electrodes arranged on the scalp in accordance with the international 10–20 system. For the two EEG examples in this paper, the first signal was obtained from the subject’s occipital region with a sampling rate of 960 Hz. The second signal was recorded from a different subject’s forehead at a sampling rate of 2 kHz.

Neural Rhythm Extraction

The neural rhythm extraction (NRE) algorithm consists of spike attenuation followed by wavelet packet filtering (Fig. 1a). An overview of the methodology of

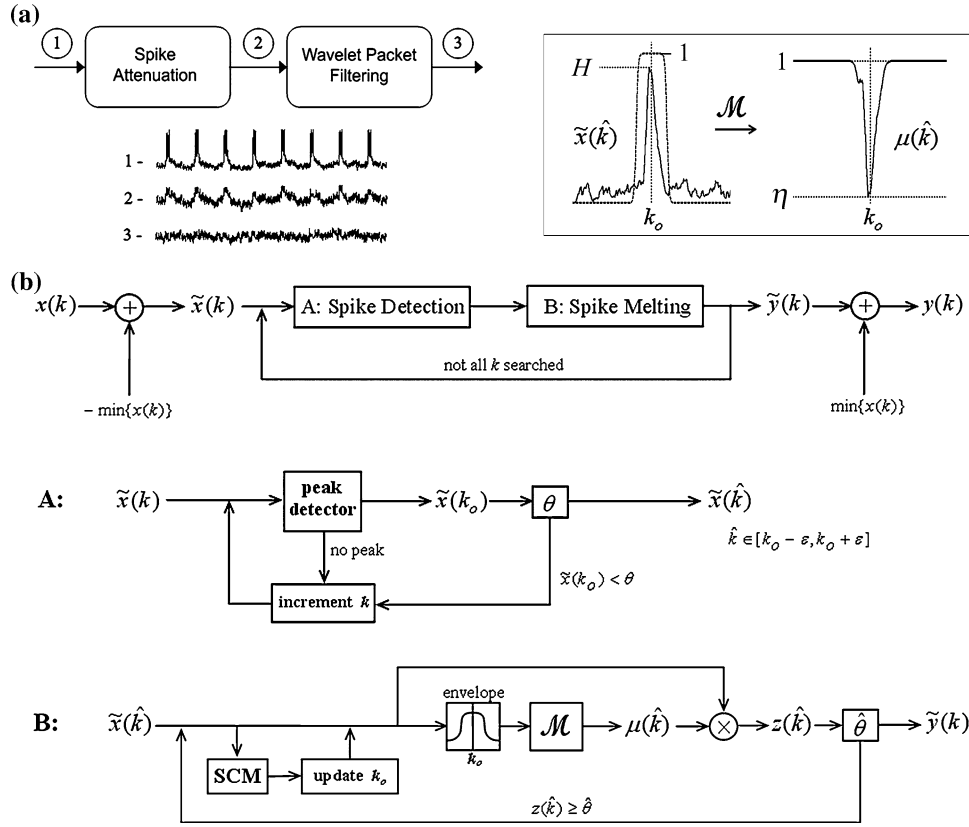


FIGURE 1. (a) Block diagram of neural rhythm extraction (NRE). (b) Automated spike attenuation. A sequence containing large amplitude spikes, $x(k)$, is iteratively processed through a detection loop (A) and a melting loop (B), producing a despiked signal, $y(k)$. If a spike with peak amplitude H is detected at k_o , a sigmoidal envelope is fitted around the spike to isolate it and a scaled inversion transformation carried out to obtain the mirror image of the spike (*inset*), $\mu(\hat{k})$, which is used to melt down the original spike through repeated multiplication, with the most prominent features of the spike being reduced the fastest, and the baseline activity being left virtually untouched. As a spike is melted, its profile changes and its center-of-mass shifts; therefore, k_o , and subsequently $\mu(\hat{k})$, are updated to reflect these changes following each iteration of the melting loop.

wavelet transforms and wavelet packet analysis is provided in the Appendix.

Spike Attenuation

The technique devised to remove large amplitude spikes from a time series relies on amplitude threshold detection and involves using the identified spikes' own rescaled mirror images to iteratively “melt” them down. Therefore, spikes are not filtered out or excised; rather, they are transformed so that their prominent high-frequency components and spectral power are greatly diminished. This approach aims to blend the attenuated spikes in with their surroundings, rather than remove them completely, so that as much information as possible is preserved before wavelet packet filtering is performed. If removal of spikes is not necessary, the detection threshold can be set arbitrarily high so as to pass the entire signal through without alteration.

A schematic of the spike attenuation algorithm is presented in Fig. 1b. There are two blocks—a spike detection stage (A), and a spike melting stage (B). The outline of the algorithm for the two stages proceeds as follows:

1. Subtract the minimum value of the input time series, so that all values are non-negative: $x(k) - \min\{x(k)\} = \tilde{x}(k)$, where $\tilde{x}(k) \geq 0$ for all k .
2. Compute a global threshold, θ , for large amplitude spike/artifact detection. An appropriate threshold is one proportional to the root mean square of the non-negative input, $\theta = C\sqrt{[\bar{x}(k)]^2}$, where $C \geq 1$ and the bar denotes the signal mean (i.e. $\bar{x} = \sum_{k=1}^N x(k)/N$), though a different definition for θ may be utilized. The proportionality constant is ideally greater or equal to one because it is desirable to detect peaks that are beyond the range of the average baseline activity.

Block A:

3. Determine if there exists a local maximum (peak) in $\tilde{x}(k)$ at the current index, $k = k_o$.
4. If $\tilde{x}(k_o)$ is not a local maximum, increment k by 1 and repeat step 3, else go to step 5.
5. Determine whether the local maximum is a large amplitude spike (LAS), subject to global amplitude threshold θ . If $\tilde{x}(k_o) < \theta$, then the local peak is not an LAS, so repeat steps 3–5, else if $\tilde{x}(k_o) \geq \theta$, go to Block B.

Block B:

Let $\hat{k} \in [k_o - \varepsilon, k_o + \varepsilon]$ pertain to an interval subset of k defined in the neighborhood ε of the identified LAS. The choice of ε is made by inspection before application of the spike-melting algorithm, and should be large enough to allow for encompassing of each LAS or LAS cluster (if tightly grouped) without overlapping the next. As a pre-processing step, the minimum of the given time series segment $\tilde{x}(\hat{k})$ may be subtracted out before executing Block B (as in step 1) and added back at the end of the stage (see step 11) once iterative processing is completed.

6. Compute a local threshold for spike melting, where again the definition of the threshold is not strict, although $\hat{\theta} = \hat{C} \sqrt{[\tilde{x}(\hat{k})]^2}$ ($0 < \hat{C} \leq 1$) is an appropriate threshold definition (similar to step 2). Here, the proportionality constant should be around one or less so that the LAS amplitude is attenuated to within a range equivalent to or less than that of the average baseline activity in its neighborhood.
7. Set

$$k_o \sim \frac{\sum_{\hat{k}} (\tilde{x}(\hat{k}) \cdot \hat{k})}{\sum_{\hat{k}} \tilde{x}(\hat{k})}$$

which is the approximate center-of-mass of the LAS.

8. Isolate and create an inverse scaled image of the LAS:

$$\mu(\hat{k}) = 1 - (1 - \eta) \cdot \Omega(\hat{k}) \frac{\tilde{x}(\hat{k})}{\max\{\tilde{x}(\hat{k})\}}$$

with scaling $0 \ll \eta < 1$ (i.e. arbitrarily close to 1), and sigmoidal envelope

$$\Omega(\hat{k}) = \left(\frac{1}{1 + \exp(-(\xi_1/2\varepsilon)(\hat{k} - k_o + \xi_2\varepsilon))} - \frac{1}{1 + \exp(-(\xi_1/2\varepsilon)(\hat{k} - k_o - \xi_2\varepsilon))} \right)$$

where ξ_1 and ξ_2 are parameters adjusting the edge slant and relative width of the envelope, respectively. As scaling tends to unity, the amount of attenuation effected by each iteration tends to zero, so that finer incremental attenuation can be achieved by increasing the precision of η toward unity, provided enough iterations are available to reach the desired attenuation. The properties of the envelope as controlled by ξ_1 and ξ_2 affect the both the isolation of the LAS from its surroundings, and preservation of the baseline activity from the effects of the spike melting transformation. As such, ξ_1 should be large enough to enable sharp delineation of the LAS from its neighborhood, and ξ_2 should be some fraction in relation to the neighborhood radius, ε . For this paper, we fixed $\eta = 0.999$, $\xi_1 = 100$ and $\xi_2 = 0.4$.

9. Obtain spike-attenuated intermediate, $z(\hat{k}) = \mu(\hat{k}) \cdot \tilde{x}(\hat{k})$.
10. Compare $z(\hat{k})$ against the local threshold, $\hat{\theta}$. If $z(\hat{k}) \geq \hat{\theta}$ for any \hat{k} , set $x(\hat{k}) = z(\hat{k})$ and repeat steps 7–10, else output $\tilde{y}(k)$, the transformed version of $\tilde{x}(k)$ with $\tilde{y}(\hat{k}) = z(\hat{k})$, and proceed to step 11.
11. Check whether last index for k has been reached; if not, return to Block A, else output despiked signal, $y(k) = \tilde{y}(k) + \min\{x(k)\}$.

If information about spike amplitude and phase is important, a time series consisting solely of the removed spikes, $y_s(k)$, can be retained separately through simple subtraction, $y_s(k) = x(k) - y(k)$, and added back to the processed time series at a later time.

Wavelet Packet Filtering

For a sampled signal of length $L = 2^N$, analysis of the signal using wavelet packets generates a binary tree with a maximum depth of N levels. This results in a possible $(2^{N+1} - 1)$ total number of coefficient nodes. For many applications, the full binary tree decomposition is not required. Depending on the desired application, an optimal wavelet packet tree can be generated, either through selective pruning of tree nodes or by placing conditions on node splitting. The most commonly used optimization scheme is information cost minimization.⁹ An appropriate cost functional (e.g. Shannon entropy) is used to gauge the information content of coefficients of the signal for a given orthonormal basis. The wavelet packet basis that generates the lowest information cost is selected. This approach is particularly suited to image and data compression applications because of the efficiency in data representation.

To isolate rhythms from a signal, we define a metric based on the variance of peak-to-peak intervals in a given time series, called the Normalized Interpeak Interval Variance:

$$\lambda = \frac{1}{N_p - 2} \sum_{m=1}^{N_p-1} \left(\frac{\delta_m}{\hat{\delta}} - 1 \right)^2 \quad (1)$$

where N_p is the number of identified peaks (local maxima), δ_m is the m th interpeak interval, and $\hat{\delta}$ is the mean interpeak interval. As defined, λ is a measure of the temporal variability of a time series. The more regular the signal, the lower the value of λ . For instance, a pure sinusoid has $\lambda = 0$, since $\delta_m = \hat{\delta}$ for all m . Conversely, a signal with uniformly distributed random interpeak intervals has $\lambda = 1/3$ (in the limit of infinite signal length). Normalization by $\hat{\delta}$ in Eq. (1) ensures the value of λ is independent of the periodicity of a signal.

When implemented in the context of wavelet packet analysis, Eq. (1) provides a means of evaluating which coefficient nodes to prune with respect to the rhythmicity of their reconstructed time series. In this way, decomposed signal components can automatically be extracted should they fall below a set threshold for λ . To avoid errors caused by spurious peaks, a small-window moving average filter (e.g. 3 samples) may be applied to smooth the time series prior to calculation. Also, constraints on which peaks are to be included in the determination of λ may be introduced if necessary. For example, one such constraint might dictate that only peaks whose magnitude exceeds fraction χ of the mean peak size are to be included in the calculation.

To automate the wavelet packet algorithm, we define conditions for node-splitting with respect to both λ and average signal power of a given nodal time series reconstructed from its coefficients. The latter is to ensure the wavelet packet tree decomposition is not excessively thorough, in that as the original signal is decomposed into finer components, node splitting ceases at a particular node if its contribution to the cumulative power of the signal is insignificant.

For an arbitrary time series, $s(k)$, of length L , the expression defining its average power is given by

$$P = \frac{1}{L} \sum_{k=1}^L [s(k)]^2. \quad (2)$$

Let θ_λ be the threshold for λ and $0 \leq \beta_P < 1$ be the power significance level. If we examine the splitting of one (parent) node into two daughter nodes, then the following three conditions must hold for the splitting to be accepted:

$$\lambda_{\text{parent}} > \theta_\lambda \quad (3)$$

$$P_{\text{parent}} > \beta_P P_x \quad (4)$$

$$n_{\text{daughter}} \leq \log_2(L_x) \quad (5)$$

where P_x is the average power of the starting signal, L_x is its length, and $n = 0, 1, 2, \dots$, is the node level index. Practically, β_P controls the size of the resultant wavelet packet tree by controlling the depth of the analysis, up to the maximum level defined by Eq. (5).

Once the splitting process is complete and an optimal tree has been created, a final pruning operation removes the tree leaves (terminal nodes) that are below θ_λ . If desired, a different threshold value for pruning may be selected. Since the wavelet packet analysis tree preserves all information about the original signal, pruned nodes containing rhythmic components need not be discarded. These nodes can be retained and independently reconstructed, thereby allowing for their isolated rhythmic features to be utilized in other applications.

Rhythm Reconstruction and Artifact Removal

Nodes generated by NRE can be ordered according to increasing frequency, and contiguous rhythms can be assembled from pruned nodes clustered together in the frequency domain. For a frequency-ordered wavelet packet tree with level index $n = 0, 1, 2, \dots, N$, frequency index $v = 0, 1, \dots, 2^n - 1$, and time series sampling period Δt , the sub-band localization of coefficient nodes approximately follows

$$f_{n,v} \sim \frac{v}{2^{n+1}\Delta t} \quad (6)$$

where the exact relationship depends on the frequency characteristics of the analyzing wavelet packet functions. To quickly assess the frequency contents and spread attributed to a given node, a fast Fourier transform (FFT) spectral analysis of its reconstructed time series is sufficient. Direct reconstruction of neural source rhythms from pruned nodes is straightforward if there are no competing rhythmic artifacts, but if such artifacts are present, it may be difficult to separate them out by inspection alone, due to their overlapping time-frequency features.

Independent component analysis (ICA) is one technique that can supplement NRE to enable separation of rhythmic artifacts extracted together with genuine source rhythms. To implement ICA, we used the FastICA toolbox for MATLAB (The MathWorks, Natick, MA) developed by Gävert *et al.* at the Laboratory of Information and Computer Science, Helsinki University of Technology. The software is based on the fast fixed-point algorithm of Hyvärinen and Oja.²⁰ The premise of ICA for source separation assumes that

the observed multivariate time series data $\mathbf{x} = [x_1(t), x_2(t), \dots, x_q(t)]^T$ are comprised of mixtures of independent sources, $\mathbf{s} = [s_1(t), s_2(t), \dots, s_m(t)]^T, m \leq q$, which are not directly observable:

$$\mathbf{x} = \mathbf{M}\mathbf{s} \quad (7)$$

where \mathbf{M} is an unknown $q \times m$ mixing matrix. Estimating \mathbf{s} typically requires that Eq. (7) be linearly transformed,

$$\tilde{\mathbf{x}} = \mathbf{D}\mathbf{x} = \mathbf{D}\mathbf{M}\mathbf{s} = \mathbf{A}\mathbf{s} \quad (8)$$

so that the components of $\tilde{\mathbf{x}}$ are reversibly whitened (i.e. de-correlated and normalized to have unit variance), and the dimensionality is reduced by principal component analysis (PCA) so that $\tilde{\mathbf{x}}$ has the same dimensions as \mathbf{s} . It is also prudent to assume the components of \mathbf{s} have unit variance, since it is impossible to estimate their true variances. Consequently, \mathbf{A} is a new $m \times m$ orthogonal matrix that is easier to estimate than \mathbf{M} , and its inverse is simply its transpose, hence

$$\mathbf{s} = \mathbf{A}^T \tilde{\mathbf{x}} = \mathbf{W} \tilde{\mathbf{x}}. \quad (9)$$

The elements of \mathbf{W} , known as weights, are chosen so that the mutual information between the m source signals of \mathbf{s} is minimized, thereby maximizing their independence. This is equivalent to maximizing the negative entropy, $J(\cdot)$, approximated by the expression¹⁹

$$J(\mathbf{w}_j) \sim (E(G(\mathbf{w}_j \tilde{\mathbf{x}})) - E(G(u)))^2 \quad (10)$$

where \mathbf{w}_j is a $1 \times m$ row (weight) vector of \mathbf{W} , $E(\cdot)$ is the expected value, u is a standardized Gaussian random sequence, and $G(\cdot)$ is a nonlinear contrast function suitably chosen to discriminate between component $\mathbf{w}_j \tilde{\mathbf{x}} = s_j(t)$ and u . To generalize Eq. (10) for multiple weight vectors, the optimization problem becomes the following:

$$\mathbf{W} = [\hat{\mathbf{w}}_1 \quad \hat{\mathbf{w}}_2 \quad \dots \quad \hat{\mathbf{w}}_m]^T \text{ such that} \quad (11)$$

$$\sum_{j=1}^m J(\hat{\mathbf{w}}_j) = \max \left\{ \sum_{j=1}^m J(\mathbf{w}_j) \right\}$$

where appropriate constraints are prescribed to preserve the statistical independence of the sources, the orthogonality of \mathbf{W} , and to ensure the boundedness of individual weight vectors so that the search space is sufficiently constrained. The robust fixed-point algorithm of Hyvärinen and Oja²⁰ utilizes modified versions of Eqs. (10) and (11) in a non-gradient scheme that iteratively optimizes for the weight vectors, thereby improving convergence relative to gradient-descent methods.²¹

Once \mathbf{s} has been determined by ICA (Eq. 9), sources that are clearly identified as artifacts (such as 60 Hz electrical noise) can be removed by setting

those elements of \mathbf{s} to zero, and the artifact-free observables can then be reconstituted using Eq. (8), upon de-whitening \mathbf{A} .

In practice, ICA requires that the multivariate time series \mathbf{x} be obtained by simultaneously recording from q separate channels. However, with NRE, there is one signal to be analyzed that contains multiple rhythmic sources. Therefore, following NRE wavelet packet filtering that selects for and isolates those rhythmic features, the pruned nodes containing the extracted features are used to construct virtual channels for ICA. This is accomplished by partitioning the nodes (according to sequential or frequency order) and combining their nodal reconstructions into q time series. NRE is beneficial as a preprocessing step to ICA, because it concentrates signal features and reduces noise, thereby facilitating source separation. More virtual channels can be constructed than are needed, and the dimensionality can be reduced through principal component analysis that precedes the optimization stage of ICA. Limitations to ICA include the arbitrary scaling and polarity of the independent components, and the need to assume or estimate the number of sources *a priori*.

Quantifying Performance

In order to evaluate the performance of NRE, we compared the results of NRE against test signals based on their relative similarity and error. The selected measure of similarity was the correlation coefficient, \hat{R} , which is an estimate of signal correlation taken at time lag equal to zero, and normalized by the autocorrelation product³:

$$\hat{R} = \frac{R_{XY}(0)}{\sqrt{R_{XX}(0)R_{YY}(0)}} \quad (12)$$

where

$$R_{UV}(\tau) = \sum_{k=1}^N u(k)v(\tau+k). \quad (13)$$

The correlation coefficient as defined takes on values from -1 to 1 , such that at 1 the signals have perfect correlation, at zero they have no correlation, and at -1 they are perfectly anti-correlated. In the context of this paper, Eqs. (12) and (13) were applied to evaluate the similarity of finite test signals at fixed zero lag, and were not used in the traditional sense of determining the degree of correlation between unknown signals over a range of lags; therefore, the measure could be applied without regard to the stationarity of the signals.

The relative error was designated to be the percentage root mean squared deviation of the signal, $y(k)$, with respect to a reference signal, $x(k)$, and

normalized by the mean peak-to-peak magnitude of the reference signal, \hat{x} :

$$e = \frac{1}{\hat{x}} \sqrt{\frac{1}{N} \sum_{k=1}^N (y(k) - x(k))^2} \times 100\%. \quad (14)$$

RESULTS

Demonstration of Rhythmic and Non-Rhythmic Signal Separation

NRE wavelet packet filtering (WPF) was applied to a test signal created by combining a normally distributed random sequence, X1, with a time-varying sinusoid, X2 (Fig. 2). The shifting frequency–time characteristics of X2 and its spectral overlap with X1 make clean separation of the two signals difficult. Various combinations of the normalized interpeak interval variance threshold (θ_λ) and power significance level (β_P) were explored to achieve reasonable separation. In this example, the wavelet packet filtered signal was compared against X1, while the time series reconstruction of the pruned nodes containing the extracted rhythmic components was compared against X2. Continuous wavelet transforms of the

reconstructed signals visually indicate separation of the signals in the time–frequency plane; the filtered signal (Y1) is similar to X1, while the extracted signal (Y2) is similar to X2. Color maps of the correlation coefficient and percentage root mean squared error illustrate an optimal ridge for signal separation for a pruning threshold between 10^{-3} to 10^{-2} , corresponding to the power significance in the range of an order of magnitude larger. Along the ridge, both the error and the correlation of the extracted signal to X1 are minimized, and the correlation of the extracted signal to X2 is maximized. As the pruning level becomes too high, eventually all components are filtered out (and therefore extracted), including arrhythmic or noisy components, which explains the surge in the correlation to X1 as θ_λ surpasses $\sim 2 \times 10^{-2}$, and why the correlation to X2 drops off. Conversely, as θ_λ becomes too small, the constraint on what is considered rhythmic tightens to the extent that only some or none of the rhythmic components are extracted. Furthermore, there is a shift of the optimal ridge toward lower θ_λ threshold values as β_P decreases. The reason is because β_P controls the depth of the wavelet packet tree decomposition, and as tree depth increases, there is a narrowing of the frequency sub-bands analyzed by wavelet packet nodes at each consecutive level, which means the node contents are more rhythmically pure; hence a lower pruning threshold is needed to compensate.

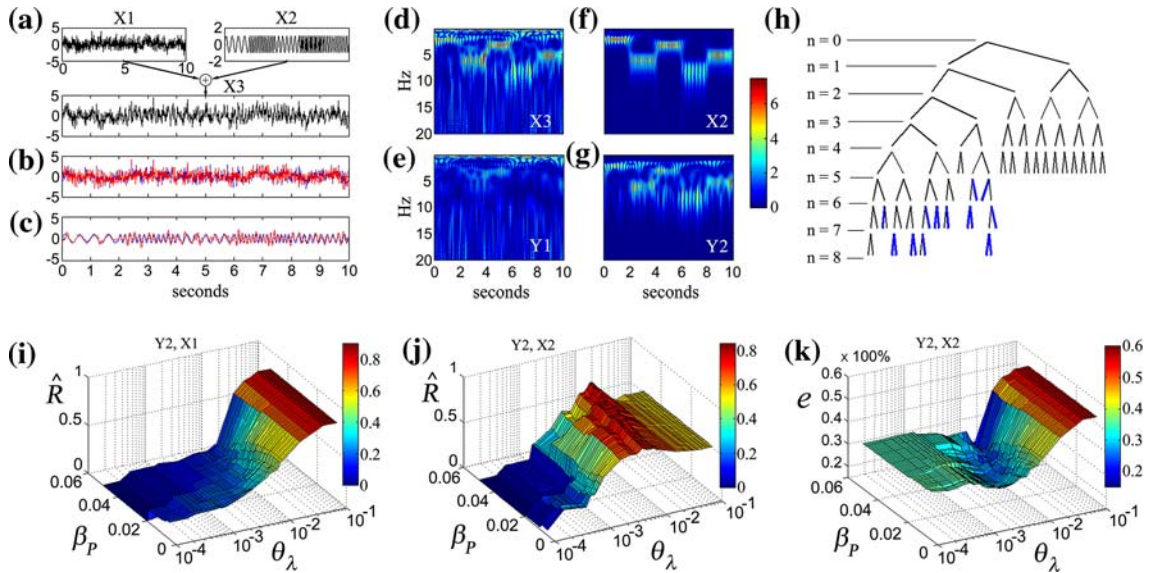


FIGURE 2. Demonstration of NRE wavelet packet filtering and extraction on a test signal. (a) Test signal (X3) created by combining a random sequence (X1) with a time-varying sinusoid (X2). (b) and (c) Wavelet packet filtering. Threshold for analysis: $\theta_\lambda = 0.0005$; threshold for pruning: $\theta_\lambda = 0.005$; power significance: $\beta_P = 0.03$. (b) Non-rhythmic filtered signal (red) superimposed on X1 (blue); percentage root mean squared error, $e = 28.4\%$. (c) Extracted rhythmic signal (red) reconstructed from pruned nodes, superimposed on X2 (blue); $e = 23.5\%$. (d)–(g) Continuous wavelet transforms of (d) X3; (e) the filtered signal, Y1; (f) X2; (g) the extracted signal, Y2. (h) Wavelet packet analysis tree, consisting of eight levels and 101 nodes, of which 51 nodes are leaves. Pruned leaves are indicated by blue edges. (i)–(k) Color maps detailing performance of wavelet packet extraction for various values of θ_λ and β_P (the same θ_λ was set for analysis and pruning): (i) zero-lag correlation of the extracted rhythmic signal vs. X1; (j) zero-lag correlation of the extracted rhythmic signal vs. X2; (k) %RMS error of the extracted signal vs. X2.

using the raw output of wavelet packet decomposition as a pre-cursor to ICA is not practical or reliable, since the percentage variance explained by the 8 or less dominant components in the trials was below 20% of that of the original signal. In the second set (Fig. 3b), with nodes combined into frequency-ordered virtual channels, separation was much improved, as more than 90% of the signal variation could be captured by at least 6–8 components. Correlation coefficient values for set 2 were also much higher, as a result of better separation. For example, in the 8-component trial, one can quantitatively identify components c1, c3, c5, and c8 as belonging principally to X1, whereas, components c6 and c4 belong to X2, and c7 and c2 are still mixtures. This degree of separation approaches but does not surpass that of NRE. Note the magnitude of the correlation is more important than the sign, as a negative sign implies inverted polarity of the component in relation to its originating signal. Convergence was not achieved for the 4-component trial, despite trying various initializations and exchanging contrast functions.

Separation of Rhythmic Source Components

Taking advantage of the combination of ICA with wavelet packet analysis, if ICA is applied to either NRE wavelet packet filtered or extracted signals, it may be possible to further discriminate sub-components within each group. Again, using the test signals of the previous examples, we applied NRE followed by ICA on the pruned nodes, retaining >95% of the variation in 5 components. Figure 4 shows that separation of rhythmic sub-components is feasible, and that each separated component can be correlated to its respective rhythmic segment in X2, to within a sign. Separation of components for the noise-free signal is naturally superior to that of the combined signal, though the correlation with X2 is not perfect because of the inherent spectral discontinuities in X2 at the segment junctions. Also in the noise-free case, segment *b* is split between c4 and c5, as identified visually and by the correlation coefficient values.

Large Amplitude Spikes

In order to effectively remove large amplitude spikes or extract their underlying rhythms from neural recordings, spike attenuation (SA) should be performed in advance of WPF. To demonstrate, a composite signal (X3) made up of a intracellularly recorded baseline rhythmic signal (X1) and a large amplitude bursting signal (X2) simulated using coupled oscillators³⁹ was constructed (Fig. 5a), whereby the X3 is similar in form to interictal bursting activity recorded intracellularly from neurons.^{12,26} As is typically the case with neuronal

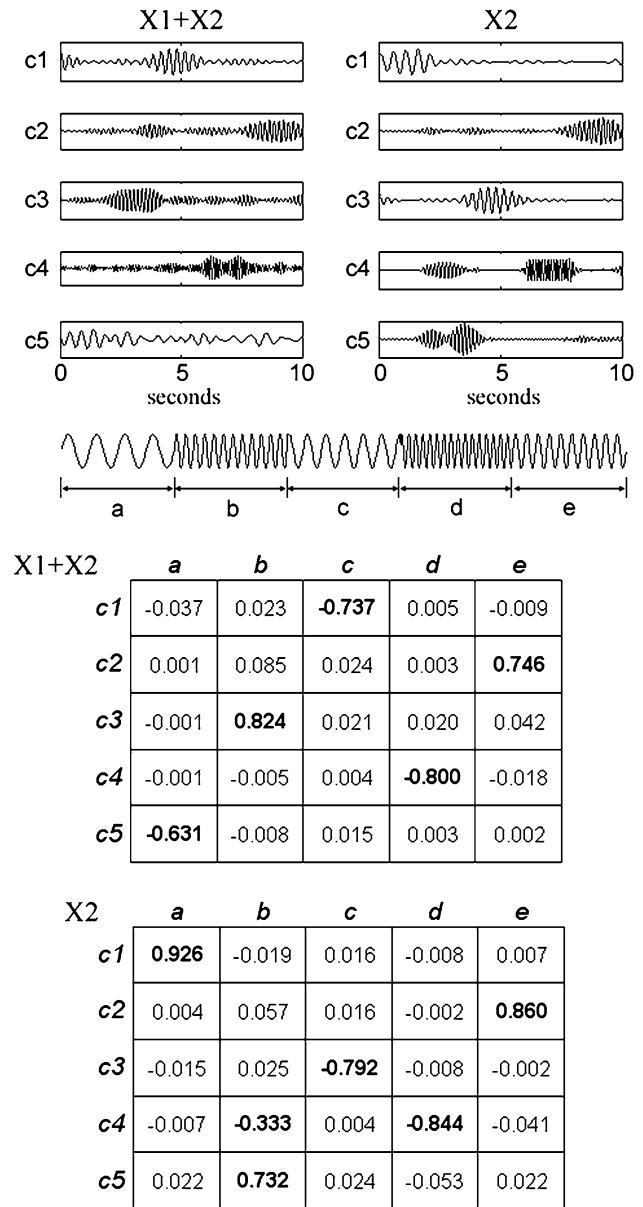


FIGURE 4. Separation of rhythmic components with ICA, following NRE wavelet packet extraction. The traces in the left column are the independent components isolated from pruned nodes following NRE extraction of rhythms from the test signal in Fig. 2. The traces in the right column pertain to ICA separation of the extracted elements from NRE performed solely on X2. The tables below correspond to the correlation coefficient (R) of each component with respect the different rhythmic segments (*a, b, c, d, e*) of X2.

recordings, the subthreshold activity and the spiking activity are both independently rhythmic. Therefore, the objective was to evaluate the statistical similarity between the wavelet packet filtered version of X1 (X1 WPF) against various methods of processing X3.

To quantify the signals, each signal was subdivided into 6 equal time segments and the sample average for the mean, variance, skewness and kurtosis were

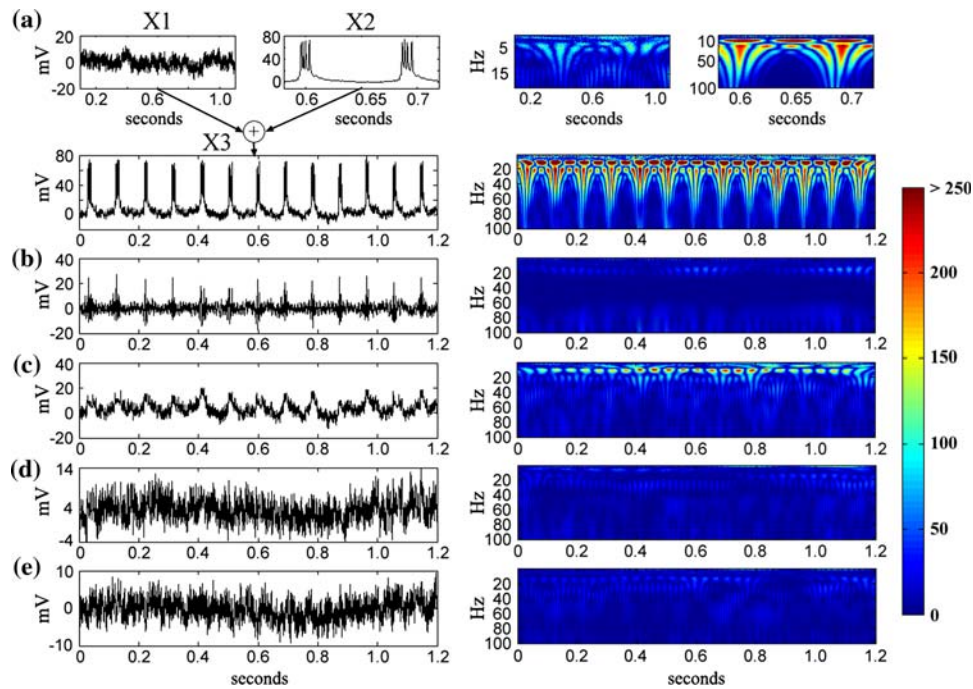


FIGURE 5. NRE processing of a signal with large amplitude spikes. (a) Composite signal (X3) created from a low-amplitude 5–8 Hz rhythmic component (X1), recorded intracellularly, and a 10–11 Hz bursting component (X2), simulated computationally. Continuous wavelet transforms of the signals revealing their time–frequency content are shown alongside the respective time series. (b) X3 wavelet packet filtered (WPF). (c) X3 with spike attenuation (SA) but no WPF. (d) X3 with SA followed by WPF. (e) X1 WPF.

TABLE 1. Statistical moments (sample mean \pm SEM)^a of unprocessed, spike attenuated (SA) and wavelet packet filtered (WPF) signals pertaining to Fig. 5.

Signal	Process	Pruning threshold	Mean	Variance	Skewness	Kurtosis
X1	WPF	0.0010	-0.34 ± 0.42	6.54 ± 0.23	0.031 ± 0.037	2.810 ± 0.042
	–	–	-0.39 ± 0.32	12.5 ± 1.5	0.173 ± 0.071	2.787 ± 0.083
X3	–	–	7.19 ± 0.31	181 ± 12	2.69 ± 0.11	11.05 ± 0.74
	WPF	0.0055	-0.25 ± 0.27	12.5 ± 1.0	0.94 ± 0.19	10.2 ± 1.6
	SA	–	3.70 ± 0.24	26.0 ± 3.1	0.394 ± 0.091	2.82 ± 0.12
	SA + WPF	0.0030	3.66 ± 0.36	6.45 ± 0.40	0.153 ± 0.041	2.829 ± 0.088

^a $n = 6$ sequential time series intervals of 2000 pts.

calculated (Table 1). For wavelet packet filtered signals, the initial normalized interpeak interval variance threshold (θ_λ) and power significance level (β_P) were set to 0.001 and 0.01, respectively. The subsequent pruning value of θ_λ determining the final tree was selected by minimizing the sum of squared errors between X1 WPF and the given signal, based on values of their four statistical moments. This was done to maximize similarity of the signals while at the same time avoiding an arbitrary choice of threshold, for purposes of comparison.

A two-sided paired *t*-test and a Wilcoxon signed ranked test³⁸ were performed on the individual and combined statistical moments, excluding the mean, to determine whether the difference between X1 WPF and

the various versions of X3 was statistically significant in each case (Table 2). The mean was excluded because the initial addition of X2 to X1 introduced a non-rhythmic offset that was not removed by NRE; furthermore, this offset can easily be subtracted to correct for the difference in the means. The results of the hypothesis tests confirm that only X3 processed with SA and WPF did not differ significantly from X1 WPF for all categories tested, therefore establishing the need for SA as a pre-processing stage to WPF when dealing with signals containing large amplitude spikes. The percentage errors and correlation coefficients of the processed signals compared to X1 WPF support this assessment (Table 3), particularly once the signal means have been subtracted to allow for equivalency in comparison.

TABLE 2. Two-sided paired *t*-test (TT) and Wilcoxon signed rank test (WT) *p* values^a.

Signal	Process	Variance		Skewness		Kurtosis		Combined	
		TT	WT	TT	WT	TT	WT	TT	WT
X1	–	0.011	0.031	0.16	0.31	0.75	0.68	0.025	0.014
X3	–	2.7×10^{-5}	0.031	3.7×10^{-6}	0.031	0.00011	0.031	0.0060	0.00020
	WPF	0.0013	0.031	0.0051	0.031	0.0051	0.031	4.9×10^{-5}	0.00020
	SA	0.0013	0.031	0.013	0.031	0.93	0.84	0.013	0.0038
	SA + WPF	0.85	0.84	0.13	0.22	0.87	1	0.89	0.37

Comparison is against X1 WPF with respect to statistical moments (excluding the mean).

^a*p* values that do not reject the null hypothesis at 5% significance are bolded.

TABLE 3. Correlation coefficient (\hat{R}) and percentage root mean squared error (*e*), with X1 WPF as the reference signal.

Signal	Process	\hat{R}	$(-\bar{X})^a$	<i>e</i> (%)	$(-\bar{X})^a$
X1	–	0.772	0.763	69.7	69.7
X3	–	0.116	0.198	458.0	397.8
	WPF	0.236	0.235	118.7	118.3
	SA	0.293	0.450	184.4	138.4
	SA + WPF	0.289	0.662	137.7	66.4

^aMeasures re-calculated following subtraction of the mean of each signal.

A survey of Fig. 5 reveals that X3 WPF without SA reduces much of the signature of the large amplitude bursts, but at the expense of removing too much of the signal, leaving a large empty swath in the time–frequency plane (Fig. 5b). SA alone also removes the high-frequency components of the bursts, but does not eliminate the low amplitude baseline rhythm associated with them (Fig. 5c).

Signal Separation in an Intracellular Recording with Spikes

For a practical demonstration of signal separation in the presence of large amplitude spikes, NRE was applied to an intracellular recording of a hippocampal CA3 pyramidal neuron under low Mg^{2+} conditions exhibiting rhythmic spiking activity. NRE successfully removed the action potentials and their underlying depolarizations following spike attenuation and wavelet packet filtering (Fig. 6a). The extracted rhythmic signal and spikes are shown in Fig. 6b, as well as their recombination to form a noise-reduced composite containing the restored action potentials as well as baseline depolarizations and sub-threshold rhythmic activity.

Extraction of EEG Source Rhythms

Extracellular field potentials reflect the summed activity of multiple neuronal populations in time and

space. As such, electroencephalogram recordings contain multiple embedded rhythms that originate from distributed source populations. Also, different rhythms predominate during different neurophysiological or pathophysiological states, which is an aspect that the EEG exploits as a diagnostic tool.³² EEG rhythms can be classified with reference to their frequency band localization, which by convention includes the delta (0.5–3.5 Hz), theta (3.5–7.5 Hz), alpha (7.5–12.5 Hz), beta (12.5–25 Hz), gamma (25–80 Hz) and super-gamma (>80 Hz) bands.²⁵ The difficulty with analyzing EEG signals is often in detecting weaker rhythms masked by predominant ones, and in distinguishing which signal features pertain to which rhythms.

The EEG signal in Fig. 7 is from the occipital region, and exhibits a well-defined alpha rhythm. Upon analysis with NRE, the time series was first decomposed into rhythmic and non-rhythmic signals, and the rhythmic signal was then further separated into three independent rhythms that were classified by Fourier spectral analysis as belonging to the delta, alpha and beta bands, respectively. These rhythms were reconstructed from pruned wavelet packet coefficient nodes grouped together according to their frequency position and order. The nonstationarity of the rhythms is evident in the spread of frequencies localized in each band. Without NRE to isolate the three rhythms, the lower-power delta and beta rhythms may have been overlooked. The alpha power exceeds the gamma and delta power by an order of magnitude, and as a result, it is difficult to visually distinguish between the Fourier spectrum of the extracted signal vs. the spectrum of the raw EEG (Fig. 7c).

Separation of Source Rhythms from Artifacts

Scalp EEGs are prone to being contaminated by electrical noise, muscle artifacts and cardiac rhythms. An example of an artifact-contaminated EEG signal from the forehead is shown in Fig. 8a (top trace). The signal, which presents with prominent cardiac QRS

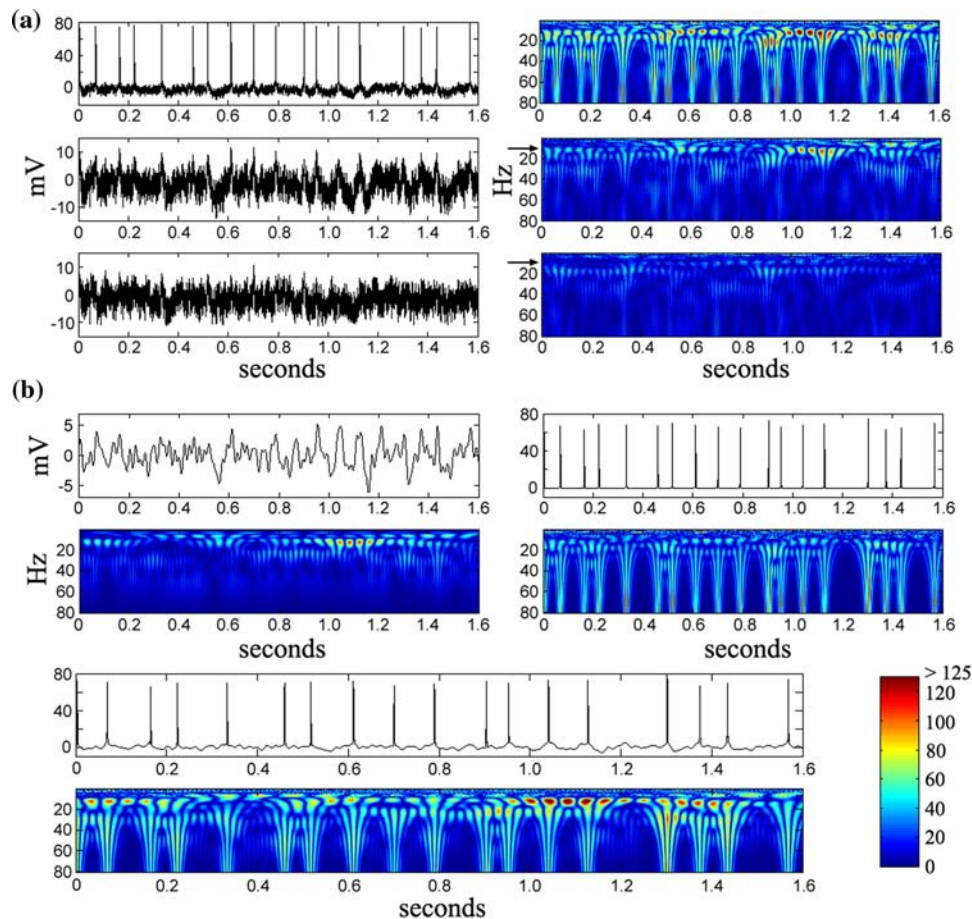


FIGURE 6. Demonstration of NRE on an *in vitro* rat hippocampal slice recording. Continuous wavelet transforms (CWT) revealing the time–frequency content of the signals are presented to the right of each time series. (a) *Top*: intracellular recording showing subthreshold and spiking activity. Note the dominant broad-spectrum spike signature in the CWT. *Middle*: signal recovered following spike attenuation. The broad-spectrum spike features have been eliminated, but a strong 10–20 Hz band remains (arrow). *Bottom*: signal recovered following spike attenuation and wavelet packet filtering ($\theta_\lambda = 0.001$ for analysis and pruning; $\beta_P = 0.01$). The spike-associated low-amplitude rhythm has been extracted without significant disruption to the remaining signal content. (b) The extracted rhythmic signal (*top left*), and the raw spikes subtracted from the signal by spike attenuation (*top right*) are merged to create a cleaned intracellular signal (*bottom*).

spikes, was analyzed by NRE and separated into non-rhythmic and rhythmic component signals (middle and bottom traces, respectively). Spike attenuation was not necessary before wavelet packet filtering because the spectral signature of the QRS complexes was fairly localized in frequency, and could therefore be removed with WPF (Fig. 8b). The algorithm also extracted a uniform band centered at 60 Hz attributed to alternating current electrical noise.

Low-amplitude neural source rhythms in the range of 0–20 Hz, however, were masked by the spectrally dominant QRS complexes. These rhythms shared the same time–frequency localization as the rhythmic QRS spikes; therefore, some of the pruned wavelet packet coefficient nodes contained a mixture of rhythmic artifact and genuine source features, and NRE was unable to separate them further. Consequently, independent component analysis (ICA) was undertaken as

an additional step to separate the extracted components based on the statistical differences between them, as was demonstrated in Fig. 4 on the test signal. From an initial pool of eight virtual channels created from the frequency-ordered pruned nodes, six sources were identified (Fig. 8c). Among the six components, c1 and c6 are likely cardiac in origin. Interspike interval analysis revealed the QRS complexes in c1 had a mean frequency of 47 beats per minute (bpm) whereas Fourier analysis of c6 yielded a spectral peak at 0.8 Hz, which translates to 48 bpm. Moreover, c1 and c6 are highly phase-correlated, which suggests a link between them. Fourier spectra of the other independent components were also computed. Component c2, which had a dominant peak at 60 Hz, pertains to the 60 Hz A/C electrical artifact. Components c3–c5 are putative neural source rhythms, with c3 situated in the alpha band (7.5–12.5 Hz), and c4 and c5 in the theta

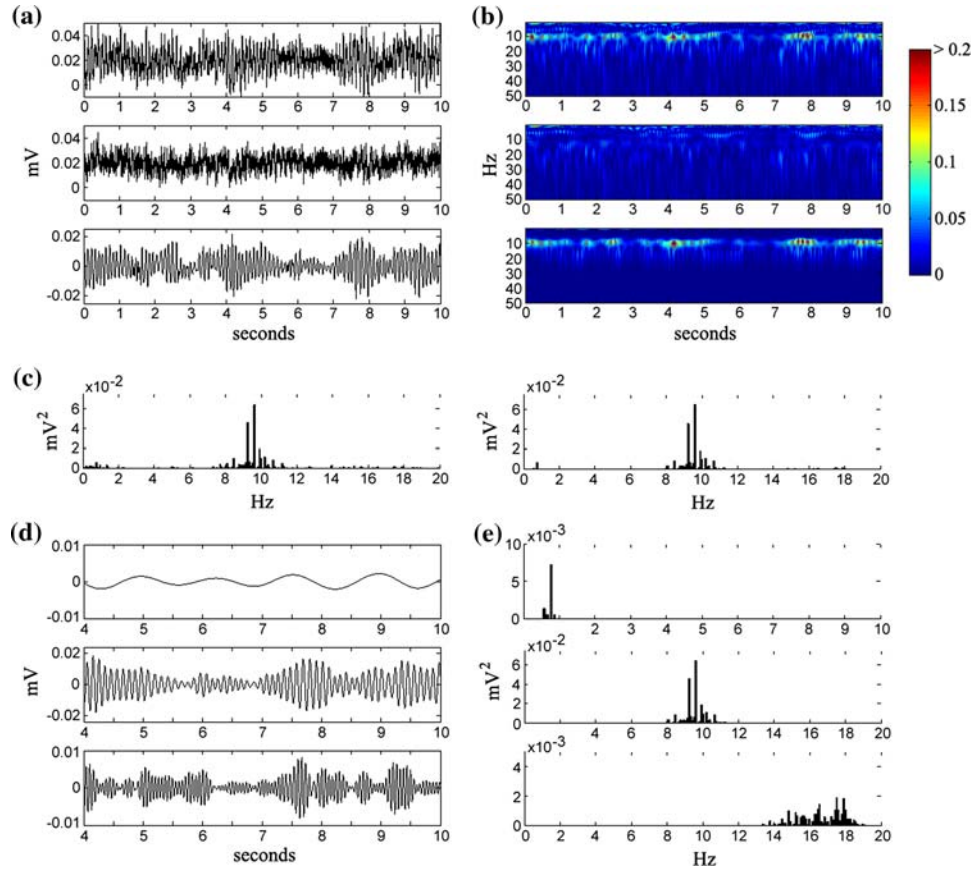


FIGURE 7. Extraction of neural source rhythms from a human scalp EEG signal. (a) *Top*: raw EEG. *Middle*: non-rhythmic EEG signal recovered following NRE wavelet packet filtering. *Bottom*: reconstruction of extracted components; $\theta_s = 0.005$ for analysis and pruning, $\beta_P = 0.01$. Of the 27 terminal nodes in the resultant wavelet packet tree, 7 were pruned. (b) Continuous wavelet transforms, respectively, of signals shown in (a). (c) Discrete Fourier power spectra of the raw EEG (*left*) and extracted signal (*right*). (d) Individual EEG-band rhythms reconstructed from pruned nodes. *Top*: node 5; *middle*: nodes 12–15; *bottom*: nodes 17 and 18. (e) Fourier spectra, respectively, of nodal reconstructions in (d). *Top*: 0.5–1 Hz (delta); *middle*: 8–12 Hz (alpha); *bottom*: 13–19 Hz (beta).

range (3.5–7.5 Hz). To demonstrate recovery of the observed signal following artifact removal, components c1 and c2 were eliminated and the observed signal reconstituted from the remaining independent components and the wavelet packet filtered signal (Fig. 8d).

DISCUSSION

The intricate texture of neural recordings, which are interwoven with time-varying rhythmicities, nonlinearities, random fluctuations and artifacts, presents a considerable challenge for signal analysis. Observation of certain features may be crucial for clinical monitoring or diagnosis. For instance, the relative power and coherence of the delta, theta, alpha or gamma band rhythms can be related to different functional states of the brain.^{2–4} Other abnormal rhythmic features, such as high-frequency (> 100 Hz) oscillations,

have been implicated in neurological disorders such as epilepsy^{29,36} and Parkinson's disease.^{17,29} Therefore, robust separation of signal features is not only desirable but necessary.

Wavelets are functions that are localized in both time and frequency, and hence can be used to construct specialized FIR quadrature mirror filter banks to decompose signals and separate out features much more effectively (see Appendix). The discrete wavelet transform (DWT), and its more general cousin, the discrete wavelet packet transform (DWPT) use orthogonal or biorthogonal wavelet bases for signal analysis and reconstruction.³⁰

The proposed neural rhythm extraction (NRE) algorithm makes use of the DWPT together with a threshold-based scheme to systematically detect and isolate rhythmic features from neural recordings. These features are then independently reconstructed or combined into contiguous rhythms. The intended objective of NRE as a tool, therefore, is to extract and

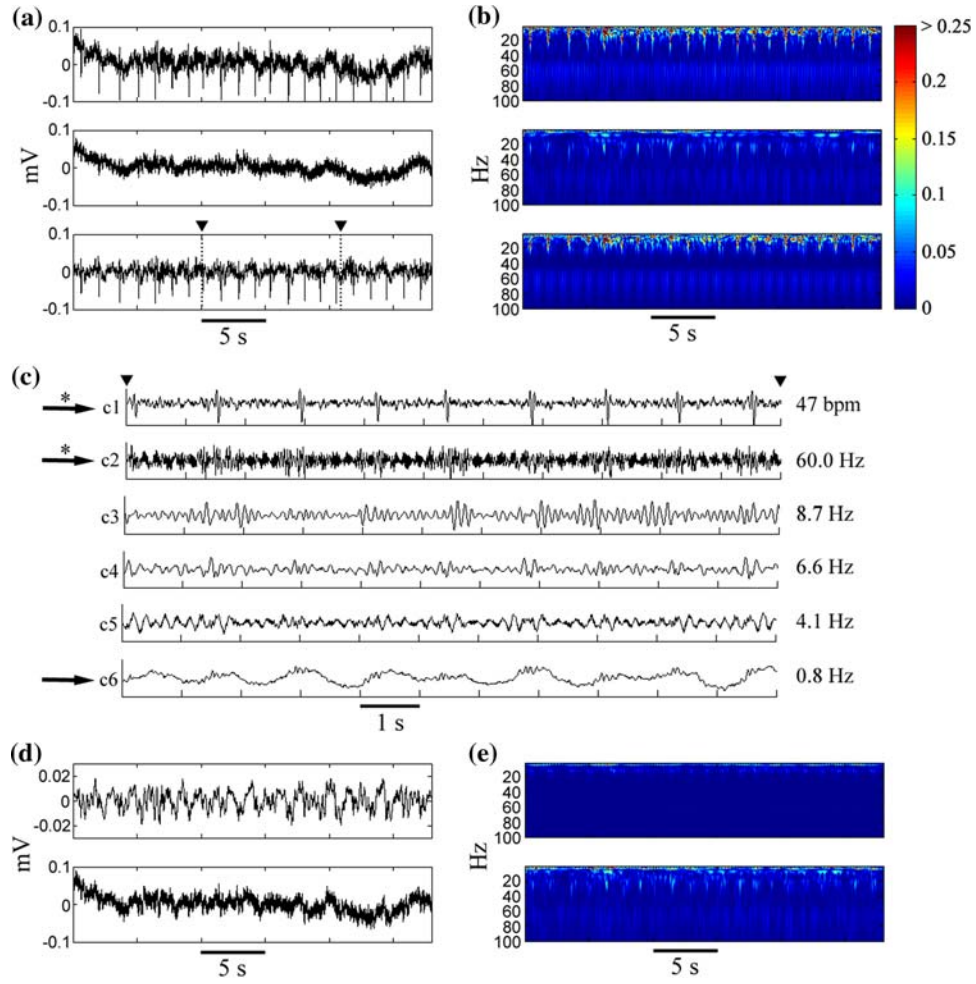


FIGURE 8. Separation of extracted rhythmic artifacts from neural source rhythms. (a) *Top*: raw EEG recorded from subject's forehead. *Middle*: non-rhythmic EEG signal recovered following NRE wavelet packet filtering (WPF). *Bottom*: extracted rhythmic signal; $\theta_\lambda = 0.001$ for analysis and 0.004 for pruning; $\beta_P = 0.01$. (b) Continuous wavelet transforms, respectively, of signals shown in (a). (c) Independent components (c1 to c6) estimated from rhythmic components extracted by WPF. View corresponds to expanded time-scale of segment denoted by inverted triangles in (a). The dominant Fourier spectral peak of each component is labeled to the right of the trace. Likely cardiac artifacts (c1, c6) and 60 Hz A/C electrical artifact (c2) are identified by arrows. Asterisks indicate components to be removed. (d) *Top*: reconstituted rhythmic signal (minus components c1 and c2), following ICA. *Bottom*: artifact-reduced composite signal composed of reconstituted rhythmic signal and NRE-isolated non-rhythmic signal. (e) Continuous wavelet transforms, respectively, of traces shown in (d).

enable separation of rhythmic features so that they can be used in other signal processing-related applications such as detection, classification and tracking. As results from Figs. 2, 6, 7, and 8 illustrate, NRE has the ability to isolate time-varying rhythms from signals even though the signals are noisy or contain many overlapping features in time or frequency. NRE is versatile because it can be applied to any signal recorded from any source, whether it be an intracellular single-neuron recording or a recording of the integrated field-potentials of multiple neuronal populations.

The NRE wavelet packet basis selection process relies on two parameters: the normalized interpeak interval variance threshold (θ_λ), which pertains to λ

defined in Eq. (1), and the power significance level (β_P) from Eq. (4). Two values of θ_λ may be assigned: one for the analysis stage, to guide wavelet packet tree formation; and the other for the pruning stage, to select which terminal tree nodes containing rhythmic features are to be extracted. Different parameter values offer varying degrees of signal decomposition and filtering. Decompositions may be shallow or deep, depending on the required application. β_P controls the relative depth of the tree, though the maximum depth is determined by the total length of the time series as constrained by multiresolution analysis. Therefore, the sampling rate is important, since a higher sampling rate for a recording of the same time duration will yield a longer time series, and hence allow for a deeper

wavelet packet tree analysis. This is a consequence of Nyquist's sampling theorem.

Shallow decompositions allow for isolation of rhythmicities with dominant spectral power, while deeper tree decompositions enable extraction of both dominant and weaker rhythmic features. Extraction of low-power, low-amplitude rhythms is possible even if their presence in a time series is not discernable by eye, or by Fourier spectral analysis. This is important because experimental evidence shows brain oscillations exhibit an inverse relationship in signal power with increasing frequency.¹¹ With EEG rhythms, higher frequency rhythms (e.g. beta and gamma) tend to be weaker in power compared to other more prominent low frequency rhythms (e.g. alpha or theta). Figure 7c demonstrates the insufficiency of Fourier methods for detecting or isolating rhythms, because of the manner in which the spectral power is averaged over time, thereby smoothing over and masking the time-dependent variation of rhythms. In contrast, the coefficients of the wavelet transform capture the temporal variation in the frequency composition of the signal.

As NRE parameter β_P is decreased, θ_λ should be proportionally decreased as well. This is evident in the color maps of Fig. 2, which show a slight shift in the optimal ridge for extraction toward lower values θ_λ for smaller β_P . The reason has to do with the narrowing of the frequency sub-bands analyzed by wavelet packet nodes at a given level, as tree depth increases. This implies, however, that choosing too small a value of β_P may lead to sub-optimal results, since the more rhythmic components of noise may also be extracted out. From empirical observation of the performance of NRE on different signals using different parameters, a rule of thumb for the selection of the threshold for pruning is that θ_λ should be roughly one order of magnitude smaller than the value of β_P , as suggested by the location of the optimal ridge for extraction in the color maps of Fig. 2. θ_λ in the range of 10^{-3} to 10^{-2} was also found to be appropriate for most signals, which makes sense because of the natural bounds on the interpeak interval variance measure ($\lambda = 0$ for a pure sinusoid; $\lambda = 1/3$ for a signal with uniformly distributed random interpeak intervals). An assessment of the efficacy of separation, however, is guided in practice by inspection of the continuous wavelet transform (CWT) of the output signals, to see whether the rhythmic features (that show up as distinct time-localized bands) have been successfully extracted.

Large amplitude spikes (LAS) are another common occurrence in neural signals, and they interfere with the extraction and separation of baseline rhythms because their dominant spectral power masks other smaller-amplitude features. Besides the utilitarian reasons for dispensing with LAS, isolation of the non-rhythmic

baseline signal may itself be of interest, due to active research into forms of (subthreshold) neuronal noise and their role in synaptic function and neuronal communication.^{14,15,22} The performance of WPF in the presence of LAS is comparable to the performance of conventional filtering, in that they are both inadequate. To address this shortcoming, a spike attenuation (SA) algorithm was developed and implemented as a pre-processing stage to WPF (Fig. 1). Several techniques have been devised to remove LAS from a time series. One relatively effective approach entails interpolating the baseline potential before and after each spike following its excision, then lowpass filtering the signal.^{33,39} A cruder method relies on bandpass filtering to attenuate the dominant-frequency spike components, but this approach usually distorts the underlying signal as well. In contrast to these techniques, NRE spike attenuation does not rely on excision or filtering; rather, it iteratively updates a mirror image of the target spike which it uses to selectively melt the spike away, thereby eliminating the spike's broad spectral footprint while preserving the baseline signal. The NRE approach to SA minimizes signal distortion and information loss, which is important for subsequent wavelet packet filtering, as demonstrated in Figs. 5 and 6. Also, information about the removed spikes' phases and amplitudes is not lost, so the spikes can be re-introduced to the post-processed time series (Fig. 6b). Though we have focused on implementation of the SA algorithm presented in this paper, this does not imply that other spike removal methods would not have produced acceptable results, and owing to the modularity of NRE, these can be substituted if desired.

In addition to LAS, neural recordings often contain unwanted artifacts that can obscure relevant portions of the signal. EEG artifacts, for instance, originate from various physiological and external sources, and can pose difficulties for EEG interpretation and analysis. External artifacts can include electromagnetic interference and thermal noise. Physiological artifacts are often biopotentials generated by the beating heart, or as the result of voluntary or involuntary muscle contractions and ocular movements. Rejecting artifacts from contaminated EEG activity using conventional filtering methods can result in significant distortion or loss of data. More selective and less disruptive means of removing artifacts from recordings have been devised, including statistical regression-based methods, principal component analysis (PCA), independent component analysis (ICA), wavelet-based denoising techniques and non-adaptive threshold methods.^{23,26,28} Though NRE is not specifically intended for artifact removal, it is able to extract rhythmic artifacts along with genuine source rhythms. The difficulty lies in trying to discriminate rhythmic

artifacts from genuine rhythmic features when they occupy the same time and frequency localization.

An example of this can be seen in Fig. 6, where an EEG recording taken from the forehead of a drowsy subject is heavily contaminated by cardiac QRS spike artifacts and 60 Hz A/C electrical noise. The wavelet transform of the extracted rhythmic signal (Fig. 8b, bottom) shows a clear band centered at 60 Hz that is easily identified and separated. But the dominant QRS spectral residues are all localized in the band from 0 to 20 Hz, which is where the less prominent theta and alpha rhythms are also located. Separating these rhythms from the QRS artifacts by clustering of frequency-ordered coefficient nodes does not work, since the pruned wavelet packet nodes pertaining to the 0–20 Hz band will contain a mixture of genuine rhythms and rhythmic artifacts. Supplementing NRE, however, with a technique such as ICA can improve performance to the extent that the signal separation achieved is superior to that of any one technique on its own. This was confirmed by attempting ICA on the straight wavelet packet decomposed signal, without normalized interpeak interval variance optimization (Fig. 3), and then on the optimized components extracted by NRE (Fig. 4). The non-optimized trials yielded poor convergence on several different runs using different ICA parameters, and on runs where convergence was attained, separation was fair but not exceptional. However, when NRE was combined with ICA, separation was improved, to the point that the independent components could be readily identified as being neural source rhythms or artifacts (Fig. 8c).

ICA may be considered as a tool that complements NRE, but the reverse also holds true. NRE serves as a pre-processing step to ICA by distilling out important signal features and reducing noise. Consequently, ICA is better able to statistically differentiate between the cleaner, processed features than raw features embedded in the original signal, as seen in Figs. 3 and 4. Mathematically, the complementariness of NRE and ICA is evident in the respective approaches of the two methods toward signal separation; with NRE, signals are classified by their degree of rhythmic variability, whereas with ICA, signals are distinguished based on their degree of non-Gaussianity. The combination of these two complementary approaches translates into improved artifact removal capabilities (Fig. 8d).

Despite the strengths of NRE, the method is limited at present by the inability to quantify its performance with respect to real signals, as well as the lack of a robust, automated method for selecting its operating parameters. Test signals have been useful for establishing criteria with regard to validation and parameter selection; nevertheless, biologically

recorded signals contain mixtures of rhythms, artifacts and noise components which are typically unknown or only partially identifiable *a priori*. One pre-processing technique that appears to improve the sensitivity of NRE involves bandpass-filtering the data into segments from low to high frequency, and performing parallel NRE on each band-limited segment to extract the rhythms.

By comparison, threshold selection methods have been developed for wavelet denoising, including minimization of Stein's unbiased risk estimate (SURE) and empirical Bayes estimates as applied to thresholding of wavelet coefficients^{7,13}; however, such methods assume statistical properties of the observed data that cannot be intrinsically verified, but only estimated, including the nature and underlying distribution of the noise or wavelet coefficients. Also, the choice of wavelet functions and the order of decomposition are left to the user, and these can affect the performance of the estimators. Independent component analysis likewise experiences difficulties from a validation standpoint, as the true number of independent sources and the amplitudes and polarities of the sources are in general unknown. Therefore, future work on NRE will involve finding ways of validating the algorithm for real biological signals, and devising techniques to determine optimal parameter values. One possible approach may be to formulate a nonlinear coherence measure to identify common signals from the wavelet packet node reconstructions which could then be cross-validated against the original data.

APPENDIX

The wavelet transform¹⁰ is the correlation of a signal, $x(t)$, with a mother wavelet function, $\psi(t)$, that is shifted and scaled, yielding wavelet coefficients $c_{\sigma,\tau}$:

$$c_{\sigma,\tau} = \frac{1}{\sqrt{\sigma}} \int_{-\infty}^{\infty} x(t) \psi\left(\frac{t-\tau}{\sigma}\right) dt. \quad (\text{A.1})$$

If discrete scaling and translation are applied such that $\sigma = 2^j$ and $\tau = 2^j k$, where k and j are integers, then a basis of orthonormal wavelet functions for $L^2(\mathbf{R})$ can be obtained³⁰:

$$\psi_{j,k} = \sqrt{2^{-j}} \psi(2^{-j}t - k). \quad (\text{A.2})$$

Restating the coefficients in terms of j and k , the signal can be represented analytically as a summation of wavelet coefficients and basis functions:

$$x(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_{j,k} \psi_{j,k}. \quad (\text{A.3})$$

Practical implementation of the discrete wavelet transform (DWT) is accomplished using a quadrature mirror filter (QMF) cascade.^{10,30} This requires the basis functions have compact support and the signal have finite energy and bandwidth. Let $\phi(t)$ be a lowpass scaling function associated with the wavelet function $\psi(t)$, such that the following expressions hold:

$$\phi(t) = \sqrt{2} \sum_k h_k \phi(2t - k) \quad (\text{A.4})$$

$$\psi(t) = \sqrt{2} \sum_k g_k \phi(2t - k) \quad (\text{A.5})$$

where h_k and g_k are lowpass and highpass QMF coefficients, respectively. The filter coefficients can be calculated directly by using Eqs. (A.4) and (A.5) and the scale and shift orthogonality of $\phi(t)$ and $\psi(t)$ to form the inner products

$$h_k = \langle \phi(t), \phi_{-1,k} \rangle = \sqrt{2} \int_{-\infty}^{\infty} \phi(t) \cdot \phi(2t - k) dt \quad (\text{A.6})$$

$$g_k = \langle \psi(t), \phi_{-1,k} \rangle = \sqrt{2} \int_{-\infty}^{\infty} \psi(t) \cdot \phi(2t - k) dt \quad (\text{A.7})$$

for all finite $k \in \mathbb{Z}$. The QMF pair has the special relationship

$$g_k = (-1)^k h_{1-k} \quad (\text{A.8})$$

in addition to satisfying the following orthogonality and normalization properties:

$$\sum_k h_{k-2m} h_{k-2m'} = \delta_{mm'} \quad (\text{A.9})$$

$$\sum_k g_{k-2m} g_{k-2m'} = \delta_{mm'} \quad (\text{A.10})$$

$$\sum_k g_{k-2m} h_{k-2m'} = 0 \quad (\text{A.11})$$

$$\sum_k h_k = \sqrt{2} \quad (\text{A.12})$$

$$\sum_k g_k = 0 \quad (\text{A.13})$$

for all $m, m' \in \mathbb{Z}$, where $\delta_{mm'}$ is the Kronecker delta.¹⁰ Alternatively, when the filter coefficients are known, the wavelet and scaling functions can be produced from the coefficients.

In general, for an N level multiresolution analysis using the DWT, the signal is represented by the expression:

$$\begin{aligned} x(t) &= A_N(t) + D_N(t) + \cdots + D_1(t) \\ &= \sum_k c_k^{A,n} \phi_{N,k} + \sum_n \sum_k c_k^{D,n} \psi_{n,k} \end{aligned} \quad (\text{A.14})$$

where $A(t)$ and $D(t)$ are approximation and detail signals, and $c_k^{A,n}$ and $c_k^{D,n}$ are the n th level approximation and detail coefficients, respectively. The $(n+1)$ level coefficients are computed by discrete convolution of the previous n th level approximation coefficients with the time-reversed QMF impulse responses, h and g , along with downsampling by 2³⁰:

$$c_k^{A,n+1} = [h_{-k} * c_k^{A,n}](\downarrow 2) = \sum_m h_{m-2k} c_m^{A,n} \quad (\text{A.15})$$

$$c_k^{D,n+1} = [g_{-k} * c_k^{A,n}](\downarrow 2) = \sum_m g_{m-2k} c_m^{A,n}. \quad (\text{A.16})$$

The downsampling operation removes redundant information in the coefficients, since the doubling of scale at each level halves the temporal resolution. By downsampling, the total number of coefficients remains roughly the same throughout the analysis. Time reversal of the filters for decomposition vs. reconstruction prevents aliasing. For signal reconstruction using the QMF cascade, the coefficients are upsampled by 2 before being passed through their respective filters and added together.

To obtain a finer resolution analysis than the DWT, and for better discrimination of higher frequency features, the discrete wavelet packet transform (DWPT) is used.^{8,18} Implementation of the DWPT is similar to the DWT, except that the scaling and wavelet functions of the DWT are extended into a set of orthonormal wavelet packet functions, $W_m(t)$, which are defined by the recursion relations¹⁸:

$$W_{2r}(t) = \sqrt{2} \sum_k h_k W_r(2t - k) \quad (\text{A.17})$$

$$W_{2r+1}(t) = \sqrt{2} \sum_k g_k W_r(2t - k) \quad (\text{A.18})$$

for integers $r \geq 0$, where $W_0(t) = \phi(t)$ and $W_1(t) = \psi(t)$. Multiresolution analysis gives rise to a binary tree structure, with coefficient nodes generated from a dictionary of wavelet packet atoms having discrete scaling, translation and sub-band localization:

$$W_{n,r,k} = \sqrt{2^{-n}} W_r(2^{-n}t - k) \quad (\text{A.19})$$

where the scale exponent $n = 0, 1, 2, \dots$, corresponds directly to the level in the tree, and index $r = 0, 1$,

2, ..., $2^n - 1$ corresponds to the horizontal node position. If tree nodes are re-ordered in terms of increasing frequency, then r becomes a frequency index relating to the sub-band analyzed by atoms $W_{n,r,k}$ at scale 2^n and time $2^n k$. For this paper, wavelet packet atoms were derived from Meyer scaling and wavelet functions, which are orthogonal, symmetric and compactly supported in the frequency domain.²⁴ Compact frequency support provides good frequency localization for purposes of separating rhythmic components, and symmetric filters ensure the phase behavior is linear. For purposes of computing the DWPT, the impulse responses of the Meyer functions were approximated to obtain compact support in time.

Signal decomposition via the DWPT is accomplished by sequentially splitting coefficient nodes into daughter node pairs, while synthesis involves the reverse operation. The leaves (terminal nodes) of the connected tree constitute an orthonormal basis from which the original signal can be represented by the following summation:

$$x(t) = \sum_{n'} \sum_{r'} x_{n',r'}(t) \quad (\text{A.20})$$

where the reconstructed nodal time series

$$x_{n',r'}(t) = \sum_k c_{n',r',k} W_{n',r',k} \quad (\text{A.21})$$

are defined for (n', r') belonging to the set denoting the leaves of the tree (with coefficients $c_{n',r',k}$).

ACKNOWLEDGMENTS

The authors thank Dr. Martin del Campo and the EEG Lab at the Toronto Western Hospital for providing human EEG data to analyze. This work has been funded by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes of Health Research (CIHR).

REFERENCES

- ¹Azimi-Sadjadi, M. R., D. Yao, Q. Huang, and G. J. Dobeck. Underwater target classification using wavelet packets and neural networks. *IEEE Trans. Neural Netw.* 11(3):784–794, 2000. doi:10.1109/72.846748.
- ²Basar, E., C. Basar-Eroglu, S. Karakas, and M. Schurmann. Gamma, alpha, delta, and theta oscillations govern cognitive processes. *Int. J. Psychophysiol.* 39(2–3):241–248, 2001.
- ³Bendat, J. S., and A. G. Piersol. *Random Data Analysis and Measurement Procedures*. 3rd ed. New York: John Wiley & Sons, Inc., 2000.
- ⁴Buzsaki, G., D. L. Buhl, K. D. Harris, J. Csicsvari, B. Czeh, and A. Morozov. Hippocampal network patterns of activity in the mouse. *Neuroscience* 116(1):201–211, 2003. doi:10.1016/S0306-4522(02)00669-3.
- ⁵Buzsaki, G., and A. Draguhn. Neuronal oscillations in cortical networks. *Science* 304(5679):1926–1929, 2004. doi:10.1126/science.1099745.
- ⁶Canolty, R. T., E. Edwards, S. S. Dalal, M. Soltani, S. S. Nagarajan, H. E. Kirsch, M. S. Berger, N. M. Barbaro, and R. T. Knight. High gamma power is phase-locked to theta oscillations in human neocortex. *Science* 313(5793):1626–1628, 2006. doi:10.1126/science.1128115.
- ⁷Chang, S. G., B. Yu, and M. Vetterli. Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans. Image Process.* 9(9):1532–1546, 2000. doi:10.1109/83.862633.
- ⁸Cody, M. A. The wavelet packet transform. *Dr. Dobbs's J.* 19(4):44–46, 50, 52, 54, 100, 1994.
- ⁹Coifman, R. R., and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Inform. Theor.* 38(2):713–718, 1992. doi:10.1109/18.119732.
- ¹⁰Daubechies, I. Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* 41(7):909–996, 1988. doi:10.1002/cpa.3160410705.
- ¹¹Demanuele, C., C. J. James, and E. J. Sonuga-Barke. Distinguishing low frequency oscillations within the 1/f spectral behaviour of electromagnetic brain signals. *Behav. Brain Funct.* 3:62, 2007. doi:10.1186/1744-9081-3-62.
- ¹²Derchansky, M., S. S. Jahromi, M. Mamani, D. S. Shin, A. Sik, and P. L. Carlen. Transition to seizures in the isolated immature mouse hippocampus: a switch from dominant phasic inhibition to dominant phasic excitation. *J. Physiol.* 586(2):477–494, 2008. doi:10.1113/jphysiol.2007.143065.
- ¹³Donoho, D. L., and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc.* 90:1200–1224, 1995. doi:10.2307/2291512.
- ¹⁴Faisal, A. A., L. P. Selen, and D. M. Wolpert. Noise in the nervous system. *Nat. Rev. Neurosci.* 9(4):292–303, 2008. doi:10.1038/nrn2258.
- ¹⁵Faure, P., and H. Korn. A nonrandom dynamic component in the synaptic noise of a central neuron. *PNAS* 94(12):6506–6511, 1997. doi:10.1073/pnas.94.12.6506.
- ¹⁶Garcia, C., G. Zikos, and G. Tziritas. Wavelet packet analysis for face recognition. *Image Vis. Comput.* 18:289–297, 2000. doi:10.1016/S0262-8856(99)00056-6.
- ¹⁷Hammond, C., H. Bergman, and P. Brown. Pathological synchronization in Parkinson's disease: networks, models and treatments. *Trends Neurosci.* 30(7):357–364, 2007. doi:10.1016/j.tins.2007.05.004.
- ¹⁸Hess-Nielsen, N., and M. V. Wickerhauser. Wavelets and time–frequency analysis. *Proc. IEEE* 84(4):523–540, 1996. doi:10.1109/5.488698.
- ¹⁹Hyvarinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* 10(3):626–634, 1999. doi:10.1109/72.761722.
- ²⁰Hyvarinen, A., and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Comput.* 9(7):1483–1492, 1997. doi:10.1162/neco.1997.9.7.1483.
- ²¹Hyvarinen, A., and E. Oja. Independent component analysis: algorithms and applications. *Neural Netw.* 13(4–5):411–430, 2000. doi:10.1016/S0893-6080(00)00026-5.
- ²²Jacobson, G. A., K. Diba, A. Yaron-Jakobovitch, Y. Oz, C. Koch, I. Segev, and Y. Yarom. Subthreshold voltage noise of rat neocortical pyramidal neurones. *J. Physiol.* 564(Pt 1):145–160, 2005. doi:10.1113/jphysiol.2004.080903.

- ²³James, C. J., and C. W. Hesse. Independent component analysis for biomedical signals. *Physiol. Meas.* 26(1):R15–R39, 2005. doi:[10.1088/0967-3334/26/1/R02](https://doi.org/10.1088/0967-3334/26/1/R02).
- ²⁴Jawerth, B., and W. Sweldens. An overview of wavelet based multiresolution analyses. *SIAM Rev.* 36(3):377–412, 1994. doi:[10.1137/1036095](https://doi.org/10.1137/1036095).
- ²⁵John, E. R. The neurophysics of consciousness. *Brain Res. Rev.* 39(1):1–28, 2002. doi:[10.1016/S0165-0173\(02\)00142-X](https://doi.org/10.1016/S0165-0173(02)00142-X).
- ²⁶Khosravani, H., C. R. Pinnegar, J. R. Mitchell, B. L. Bardakjian, P. Federico, and P. L. Carlen. Increased high-frequency oscillations precede in vitro low-Mg seizures. *Epilepsia* 46(8):1188–1197, 2005. doi:[10.1111/j.1528-1167.2005.65604.x](https://doi.org/10.1111/j.1528-1167.2005.65604.x).
- ²⁷Klausberger, T., P. J. Magill, L. F. Marton, J. D. Roberts, P. M. Cobden, G. Buzsaki, and P. Somogyi. Brain-state- and cell-type-specific firing of hippocampal interneurons in vivo. *Nature* 421(6925):844–848, 2003. doi:[10.1038/nature01374](https://doi.org/10.1038/nature01374).
- ²⁸Krishnaveni, V., S. Jayaraman, L. Anitha, and K. Ramadoss. Removal of ocular artifacts from EEG using adaptive thresholding of wavelet coefficients. *J. Neural Eng.* 3(4): 338–346, 2006. doi:[10.1088/1741-2560/3/4/011](https://doi.org/10.1088/1741-2560/3/4/011).
- ²⁹Le Van Quyen, M., I. Khalilov, and Y. Ben-Ari. The dark side of high-frequency oscillations in the developing brain. *Trends Neurosci.* 29(7):419–427, 2006. doi:[10.1016/j.tins.2006.06.001](https://doi.org/10.1016/j.tins.2006.06.001).
- ³⁰Mallat, S. G. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 11(7):674–693, 1989. doi:[10.1109/34.192463](https://doi.org/10.1109/34.192463).
- ³¹Meyer, F. G., A. Z. Averbuch, and J.-O. Stromberg. Fast adaptive wavelet packet image compression. *IEEE Trans. Image Process.* 9(5):792–800, 2000. doi:[10.1109/83.841526](https://doi.org/10.1109/83.841526).
- ³²Nunez, P. L., and R. Srinivasan. *Electric Fields of the Brain: The Neurophysics of EEG*. 2nd ed. New York: Oxford University Press, pp. 5–19, 2006.
- ³³Sakuranaga, M., Y. Ando, and K. Naka. Dynamics of the ganglion cell response in the catfish and frog retinas. *J. Gen. Physiol.* 90(2):229–259, 1987. doi:[10.1085/jgp.90.2.229](https://doi.org/10.1085/jgp.90.2.229).
- ³⁴So, P., J. T. Francis, T. I. Netoff, B. J. Gluckman, and S. J. Schiff. Periodic orbits: a new language for neuronal dynamics. *Biophys. J.* 74(6):2776–2785, 1998.
- ³⁵Tikkanen, P. E. Nonlinear wavelet and wavelet packet denoising of electrocardiogram signal. *Biol. Cybern.* 80(4): 259–267, 1999. doi:[10.1007/s004220050523](https://doi.org/10.1007/s004220050523).
- ³⁶Urrestarazu, E., R. Chander, F. Dubeau, and J. Gotman. Interictal high-frequency oscillations (100–500 Hz) in the intracerebral EEG of epileptic patients. *Brain* 130(Pt 9): 2354–2366, 2007. doi:[10.1093/brain/awm149](https://doi.org/10.1093/brain/awm149).
- ³⁷Walczak, B., and D. L. Massart. Noise suppression and signal compression using the wavelet packet transform. *Chemometr. Intell. Lab. Syst.* 36(2):81–94, 1997. doi:[10.1016/S0169-7439\(96\)00077-9](https://doi.org/10.1016/S0169-7439(96)00077-9).
- ³⁸Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bull.* 1(6):80–83, 1945. doi:[10.2307/3001968](https://doi.org/10.2307/3001968).
- ³⁹Zaghloul, K. A., K. Boahen, and J. B. Demb. Different circuits for ON and OFF retinal ganglion cells cause different contrast sensitivities. *J. Neurosci.* 23(7):2645–2654, 2003.