

RESEARCH

Open Access



Time-frequency feature extraction for classification of episodic memory

Rachele Anderson* and Maria Sandsten

*Correspondence:
rachele.anderson@matstat.lu.se
Division of Mathematical Statistics,
Centre for Mathematical Sciences,
Lund University, Sölvegatan 18, 221
00, Lund, Sweden

Abstract

This paper investigates the extraction of time-frequency (TF) features for classification of electroencephalography (EEG) signals and episodic memory. We propose a model based on the definition of locally stationary processes (LSPs), estimate the model parameters, and derive a mean square error (MSE) optimal Wigner-Ville spectrum (WVS) estimator for the signals. The estimator is compared with state-of-the-art TF representations: the spectrogram, the Welch method, the classically estimated WVS, and the Morlet wavelet scalogram. First, we evaluate the MSE of each spectrum estimate with respect to the true WVS for simulated data, where it is shown that the LSP-inference MSE optimal estimator clearly outperforms other methods. Then, we use the different TF representations to extract the features which feed a neural network classifier and compare the classification accuracies for simulated datasets. Finally, we provide an example of real data application on EEG signals measured during a visual memory encoding task, where the classification accuracy is evaluated as in the simulation study. The results show consistent improvement in classification accuracy by using the features extracted from the proposed LSP-inference MSE optimal estimator, compared to the use of state-of-the-art methods, both for simulated datasets and for the real data example.

Keywords: Time-frequency features, Classification, Non-stationary signals, Neural networks, EEG signals, Locally stationary processes, Optimal spectral estimation

1 Introduction

The analysis of electroencephalography (EEG) signals is one of the main methodological tools for understanding how the electrical activity of the brain supports cognitive functions [1]. One of the main reasons for the importance of the EEG is the outstanding temporal resolution. Additionally, the procedure of measuring EEG on the scalp surface is non-invasive and low-cost, which makes it excellent both for clinical practice and for cognition studies.

Extensive literature supports the possibility to investigate the cognitive functions, such as episodic memory, through neural activity recordings [2–4]. In cognitive psychology, event-related potentials are indicators of cognitive processes. For the transient responses which are not specifically time-locked, the time-frequency (TF) images of EEG signals have become one of the more popular techniques of today's research. The TF images are

often used for extracting the features to feed a neural network classifier, and most TF methods are based on the short-time Fourier transform (spectrogram) and the wavelet transform (scalogram). The quality of the TF representation is crucial for the extraction of robust and relevant features [1, 5–7], thus leading to the demand for high-performing spectral estimators. The wavelet transform, using the Morlet wavelet, is the most popular TF method today [8–11].

Recently, it has been argued that the sine waves and the wavelets of the spectrogram and scalogram should be replaced with physiologically defined waveform shapes which could be related to an individual-based physical model [12, 13]. With more sophisticated techniques and advanced applications, the increase in the individual-based tailoring of the methods becomes even more critical. We consider an individual-based stochastic model for the signals, based on the definition of locally stationary processes (LSPs), introduced by Silverman in [14]. LSPs are characterized by a covariance function that is the modulation in time of an ordinary stationary covariance function.

The optimal kernel for the estimation of the Wigner-Ville spectrum (WVS) for a certain class of LSPs is obtained in [15]. Based on this result, we derive the mean square error (MSE) optimal WVS for our model covariance. The use of a proper time-frequency kernel offers a valuable approach to address the fundamental problem in stochastic time-frequency analysis, which is the estimation of a reliable time-frequency spectrum from only a few realizations of a stochastic process. The derivation of the optimal time-frequency kernels of stochastic models has been a research field of signal processing interest for a long time [16, 17]. An efficient implementation is based on an equivalent optimal multitaper spectrogram [18–20]. The kernels are parameter-dependent, and thanks to the HAnkel-Toeplitz Separation (HATS) inference method introduced in [21], we are able to estimate the MSE optimal WVS from measured data. The derivation of the corresponding multitapers for the kernel is of great interest in the signal processing community, as their main advantages include computational efficiency of spectral properties and real-time use.

In a simulation study, we compare the MSE optimal WVS with state-of-the-art TF representations. Some preliminary results are presented in a previous conference paper [22]. Here, we extend the study by including the classical estimator of the WVS and the wavelet transform among the TF representations compared. Since the wavelet transform is one of the most popular representations used for EEG signal classification, its inclusion among the state-of-the-art methods considered in the comparison is especially relevant for evaluating the advantages offered by the proposed approach. Additionally, we extend the quantitative comparison in terms of MSE with the evaluation of the classification accuracy of simulated signals based on the TF features extracted with the different methods. The controlled setting of a simulation study allows evaluating how the quality of the TF estimate affects the classification accuracy of classes of signals indistinguishable in the time-domain but having distinct TF spectra. The classifier implemented is a multilayer perceptron neural network, which has been previously used for the classification of EEG signals based on TF features, e.g., [5, 23, 24]. The real data example considered consists of three categories of EEG signals measured during a memory encoding task [8, 25, 26].

The paper is structured as follows. In Section 2, we present the different TF representations, the model of LSPs used both in the simulation study and in the real data case study, the expression for the MSE optimal WVS, and the neural network used for classification.

The results for the simulation studies and for the real data case are presented in Section 3. In Section 3.1, the performance of the derived spectral estimator is evaluated in terms of MSE, whereas in Section 3.2, the performance is evaluated as classification accuracy in a simulated study. In Section 3.3, we present the results of the classification of EEG signals collected within a study on human memory encoding and retrieval. Final comments and directions for further research are given in Section 4.

2 Methods

This section is dedicated to the exposition of the theory and methods. First, we present the different TF representations considered. Then, the definition of LSP [14] is recalled, and the inference problem discussed. We proceed by presenting the MSE optimal kernel for the general LSP case, based on [15, 27], together with the MSE optimal kernel derived for the parametric model used in this paper. Finally, we specify the neural network classification approach.

2.1 Time-frequency representations

The spectrogram is a natural extension of the periodogram for non-stationary signals. The spectrogram of a zero-mean signal $x(t)$, $t \in \mathbb{R}$, is calculated as:

$$S_x(t, \omega) = \left| \int_{-\infty}^{\infty} x(s) h^*(s - t) e^{-i\omega s} ds \right|^2, \quad (1)$$

where $h(t)$ is a window function, commonly a Hanning window or Gaussian window centered around zero. The Welch method [28], first introduced for stationary spectral estimation, aims at reducing the variance of the spectral estimate by calculating the average of the spectrum estimates obtained on shorter segments of the data, possibly overlapping. It can be implemented in TF analysis for computing the spectral estimates of each shorter sub-sequence of the spectrogram.

We can write any TF representation belonging to Cohen's class, including the spectrogram, as:

$$W_x^C(t, \omega) = \int \int x\left(u + \frac{\tau}{2}\right) x^*\left(u - \frac{\tau}{2}\right) \phi(\theta, \tau) e^{i(\theta t - \omega \tau - \theta u)} du d\tau d\theta, \quad (2)$$

where all integrals run from $-\infty$ to ∞ and $\phi(\theta, \tau)$ is an ambiguity kernel [29]. With $\phi(\theta, \tau) = 1$ for all values of θ and τ , the classical Wigner-Ville distribution is received as:

$$W_x(t, \omega) = \int_{-\infty}^{\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-i\tau\omega} d\tau. \quad (3)$$

An extension of the uncertainty principle holds in TF analysis, implying a trade-off between frequency and temporal resolution. The wavelet transform is an approach to address the resolution limits posed from the uncertainty principle, by prioritizing the time resolution for high-frequency events and the frequency resolution for low-frequency events. The continuous wavelet transform is based on the correlation between the signal and the wavelet translated at time τ and scaled by a . The scalogram is defined as:

$$\mathcal{S}_x(t, a) = \left| \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(u) \psi^*\left(\frac{u - t}{a}\right) du \right|^2, \quad (4)$$

where the unite energy function ψ should fulfill the admissibility condition:

$$c_\psi = \int_{-\infty}^{\infty} \frac{|\mathcal{F}\psi(\omega)|^2}{|\omega|} d\omega < \infty, \quad (5)$$

with \mathcal{F} representing the Fourier transform. Larger scale values offer higher frequency resolution and are used for determining the frequency of long-term events, such as baseline oscillations. Smaller scales provide a high temporal resolution, necessary for determining short-time events, such as spikes and transients.

A scale-to-frequency conversion allows recovering a time-frequency representation from the scalogram, where the constant $\beta = \frac{a}{\omega}$ depends on the wavelet.

2.2 Locally stationary processes

The stochastic model proposed for episodic memory data and used for the simulation study belongs to the class of locally stationary processes in Silverman's sense [14]. A zero mean stochastic process $X(t)$ is a locally stationary process (LSP) in the wide sense if its covariance:

$$C(s, t) = \mathbb{E}[X(s)X(t)^*] \quad (6)$$

can be written as:

$$C(s, t) = q\left(\frac{s+t}{2}\right) \cdot r(s-t) = q(\eta) \cdot r(\tau), \quad (7)$$

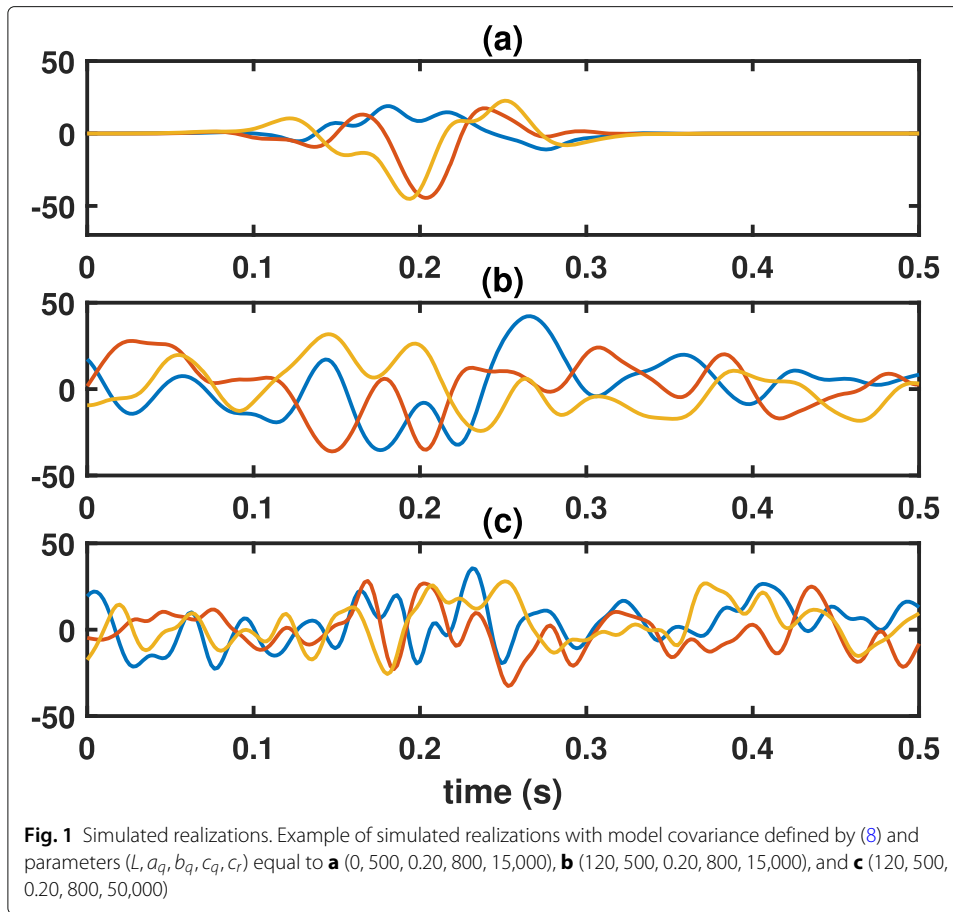
with $s, t \in [T_0, T_f] \subseteq \mathbb{R}$, where $q(\eta)$ is a non-negative function and $r(\tau)$ is a stationary covariance function. When $q(\eta)$ is a constant, (7) reduces to a stationary covariance. The proposed model is determined by the functions $q(\eta)$ and $r(\tau)$ as:

$$\begin{aligned} q(\eta) &= L + a_q \cdot \exp\left(-c_q (\eta - b_q)^2 / 2\right) \text{ with } \eta = \frac{t+s}{2}, \\ r(\tau) &= \exp\left(-\frac{c_r}{8} \cdot \tau^2\right) \text{ with } \tau = t-s, \end{aligned} \quad (8)$$

and parameters $L > 0$, $a_q > 0$, $b_q \in [T_0, T_f]$, $c_r > c_q > 0$. The latter assumption is necessary to assure that the resulting covariance is positive semi-definite. This choice of functions is motivated by the case study presented in Section 3.3.

In Fig. 1, we illustrate how different parameter settings affect the LSP realizations obtained from the model covariance (8). Each set of realizations presented has power centered at time $b_q = 0.20$ s, but different parameters determining the functions q and r result in a slowly varying behavior in “a” and “b,” compared to a much faster variation in “c” given by a larger value of c_r . The models generating the realizations shown in “a” and “b” differ only for the parameter L , representing the minimum energy level. A higher parameter L , as in “b” and “c,” is more realistic and illustrates the possibility of using the level L to include the additional on-going spontaneous EEG activity.

For the practical use of the LSP model, the parameters defining the covariance need to be estimated. A maximum likelihood approach is not feasible since a closed-form expression of the process distribution is not known. A natural approach is the least square fitting of the model covariance to the classical sample covariance matrix (SCM). Unfortunately, the SCM is not a reliable estimator when the sample size is smaller than the number of elements to be estimated [30]. Additionally, even when the sample size would be sufficient to produce reliable estimates, the initialization of the starting point for the optimization problem is computationally expensive. The inference method HATS [21], based on the separation of the two factors defining the LSP covariance function in order to take advantage of their individual structure, allows to overcome the mentioned disadvantages of the least square fitting to the SCM.



2.3 Mean square error optimal kernel

The Wigner-Ville spectrum (WVS) is the stochastic version of the classical Wigner-Ville distribution in (3) [29], defined as:

$$W(t, \omega) = \int_{-\infty}^{\infty} \mathbb{E} \left[X \left(t + \frac{\tau}{2} \right) X^* \left(t - \frac{\tau}{2} \right) \right] e^{-i\tau\omega} d\tau. \quad (9)$$

The traditional estimate of the WVS based on one realization of the process is equivalent to estimating the Wigner-Ville distribution (WV) (3). For LSPs, the WVS assumes the advantageous expression:

$$W(t, \omega) = q(t) \cdot \mathcal{F}r(\omega) \quad (10)$$

and the ambiguity spectrum, defined as:

$$A(\theta, \tau) = \int_{-\infty}^{\infty} \mathbb{E} \left[X \left(t + \frac{\tau}{2} \right) X^* \left(t - \frac{\tau}{2} \right) \right] e^{-it\theta} dt, \quad (11)$$

is given by:

$$A(\theta, \tau) = \mathcal{F}q(\theta) \cdot r(\tau) \quad (12)$$

[14, 15]. Any TF representation member of Cohen's class can be expressed in terms of the ambiguity spectrum and kernel as:

$$W^C(t, \omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(\theta, \tau) \phi(\theta, \tau) e^{-i(\tau\omega - t\theta)} d\tau d\theta. \quad (13)$$

The MSE optimal kernel for LSP having Gaussian functions as factors of the covariance is expressed as:

$$\phi_0(\theta, \tau) = \frac{|\mathcal{F}q(\theta)|^2 |r(\tau)|^2}{|\mathcal{F}q(\theta)|^2 |r(\tau)|^2 + (\mathcal{F}|r|^2(\theta))(\mathcal{F}^{-1}|\mathcal{F}q|^2(\tau))} \quad (14)$$

[15].

Thanks to (14), we are able to compute the parameter-dependent optimal kernel $\phi_0(\theta, \tau)$ for the introduced model (8), as:

$$\phi_0(\theta, \tau) = \frac{|A(\theta, \tau)|^2}{|A(\theta, \tau)|^2 + B(\theta, \tau)},$$

with

$$\begin{aligned} |A(\theta, \tau)|^2 &= |\mathcal{F}q(\theta)|^2 |r(\tau)|^2 \\ &= \left(L^2 \delta_0(\theta) + \frac{2\pi a_q^2}{c_q} e^{-\frac{\theta^2}{c_q}} + 2a_q L \sqrt{\frac{2\pi}{c_q}} \delta_0(\theta) e^{-\frac{\theta^2}{2c_q}} \right) e^{-\frac{c_r \tau^2}{4}} \end{aligned}$$

and

$$\begin{aligned} B(\theta, \tau) &= (\mathcal{F}|r|^2(\theta))(\mathcal{F}^{-1}|\mathcal{F}q|^2(\tau)) \\ &= \left(2\sqrt{\frac{\pi}{c_r}} e^{-\frac{\theta^2}{c_r}} \right) \left(\frac{L^2}{2\pi} + a_q^2 \sqrt{\frac{\pi}{c_q}} e^{-\frac{c_q \tau^2}{4}} + a_q L \sqrt{\frac{2\pi}{c_q}} \right), \end{aligned}$$

where δ_0 denotes the Dirac delta function.

The efficient implementation and estimation are based on multitapers, i.e., a weighted sum of windowed spectrograms, as:

$$W^C(t, \omega) = \mathbb{E} \left[\sum_{k=1}^K \alpha_k \left| \int_{-\infty}^{\infty} X(s) h_k^*(t-s) e^{-i\omega s} ds \right|^2 \right], \quad (15)$$

with weights α_k and windows $h_k(t)$, $k = 1 \dots K$ [18–20]. The weights and windows are derived from the solution of the eigenvalue problem:

$$\int_{-\infty}^{\infty} \Psi^{rot}(s, t) h(s) ds = \alpha h(t), \quad (16)$$

where the rotated time-lag kernel is Hermitian and defined as:

$$\Psi^{rot}(s, t) = \Psi \left(\frac{s+t}{2}, s-t \right), \quad (17)$$

with

$$\Psi(t, \tau) = \int_{-\infty}^{\infty} \phi(\theta, \tau) e^{it\theta} d\theta. \quad (18)$$

The multitaper spectrogram solution is an efficient solution from implementation aspects since only a few α_k differ significantly from zero. In the implementation for real signals, the corresponding analytic signal is used.

2.4 Pattern recognition neural networks

Pattern recognition networks are feed-forward networks, i.e., networks in which the information flow is only directed forward, which can be trained to classify inputs according to target classes [31]. The input and target vectors are divided into three separate sets

for training, validation, and testing. After the training of the neural network, the validation phase is necessary to ensure that the network is generalizing and not overfitting, and the testing phase consists of a completely independent test of the network.

As in standard networks used for pattern recognition, in this study, we consider a multilayer perceptron, with the input layer where the TF features are inserted, two hidden layers each consisting of 20 neurons, and an output layer for classification of the signals into three categories. The network structure is exemplified in Fig. 2. The activation function of the nodes in the hidden layers is the logistic function, also called sigmoid function, defined as:

$$a_{\text{sigmoid}}(z) = \frac{1}{1 + e^{-z}}. \quad (19)$$

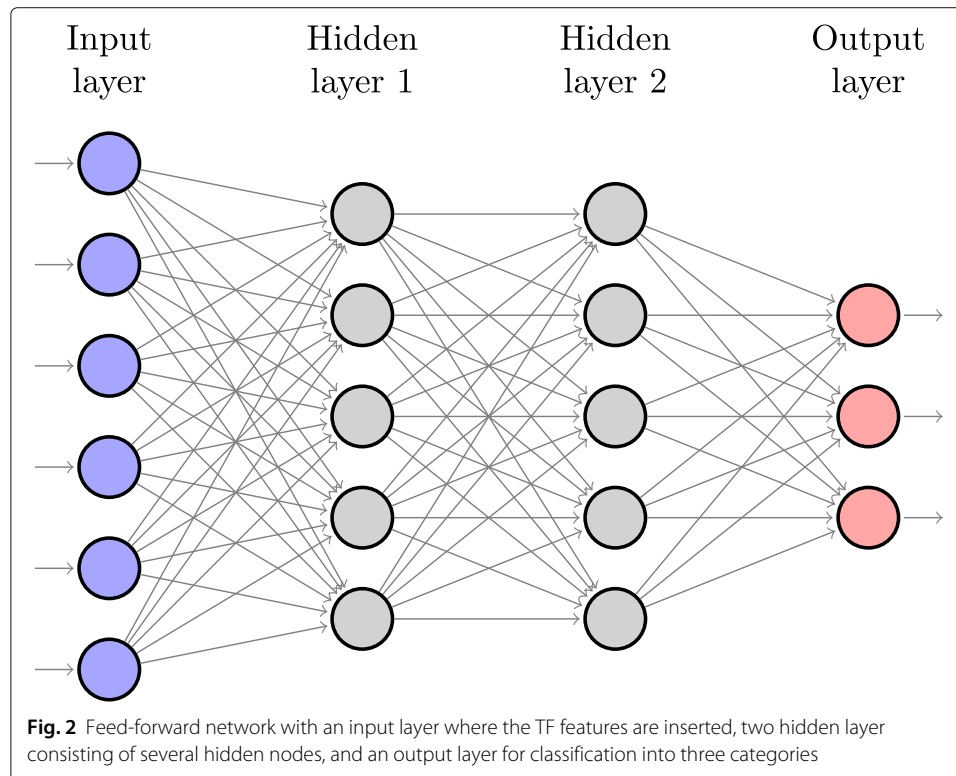
As usual for multiclass classification, the output node j converts its total input z_j into a class probability p_j by using the softmax function, defined as:

$$a_{\text{softmax}}(z_j) = p_j = \frac{e^{z_j}}{\sum_{k=1}^{N_c} e^{z_k}}, \quad (20)$$

where N_c is the total number of classes. The cost function is the categorical cross-entropy between the target probabilities p_k and the outputs of the softmax q_k :

$$C = - \sum_{k=1}^{N_c} p_k \log(q_k), \quad (21)$$

and the learning technique is back-propagation [32, 33].



3 Results and discussion

In this section, first we present the results of the evaluation of the proposed method in a simulation study where the true WVS, as given in (9), is known, and the MSE of the estimators can be calculated. The second part of the simulation study is the classification of simulated datasets based on features extracted from the different TF estimators. The same approach is then used in a real data application to classify the EEG signals measured during a memory encoding task.

3.1 Mean square error evaluation

The first part of this simulation study focuses on evaluating the proposed method performance in terms of MSE, in comparison with state-of-the-art estimators.

We simulate $M = 100$ realizations of an LSP with covariance function defined by (8), sampled with $f_s = 512$ Hz in 256 equidistant points during the time interval $[T_0, T_f] = [0, 0.5]$ seconds. The model parameters for simulating the data are $(L, a_q, b_q, c_q, c_r) = (120, 500, 0.20, 800, 15,000)$, and a few realizations from this parameter setting can be seen in Fig. 1b. The parameters (L, a_q, b_q, c_q, c_r) are estimated with the inference method HATS. Based on the parameter estimates, the MSE optimal ambiguity kernel and corresponding multitapers according to (15) are calculated. In Fig. 3, the MSE optimal multitapers and weights for this parameter set are shown.

The state-of-the-art estimators considered for comparison are the single Hanning window spectrogram (HANN), calculated as in (1); the Welch 50% overlapped segment averaging with Hanning windows (WOSA); the classical Wigner-Ville spectrum estimate

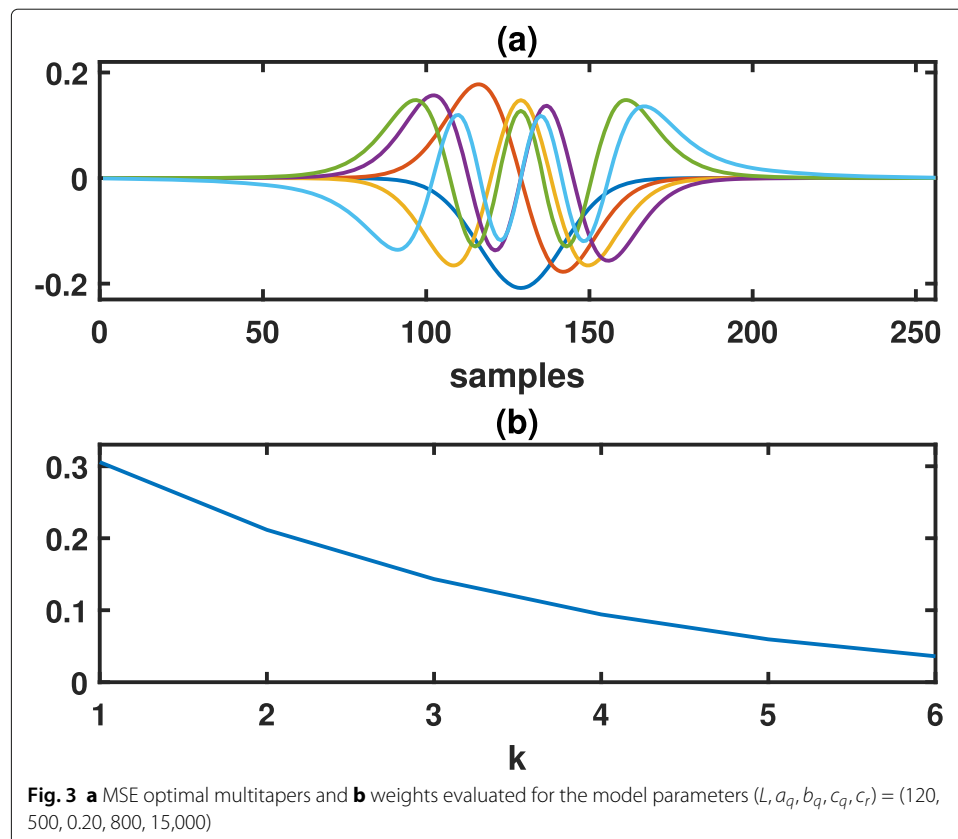


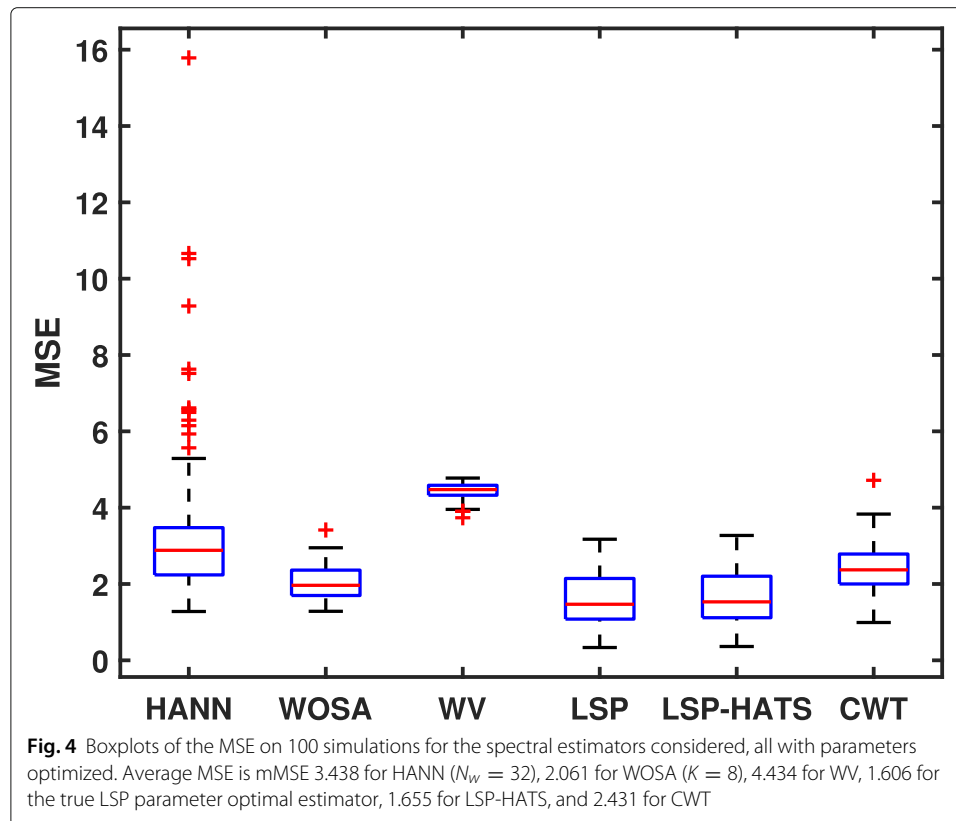
Fig. 3 **a** MSE optimal multitapers and **b** weights evaluated for the model parameters $(L, a_q, b_q, c_q, c_r) = (120, 500, 0.20, 800, 15,000)$

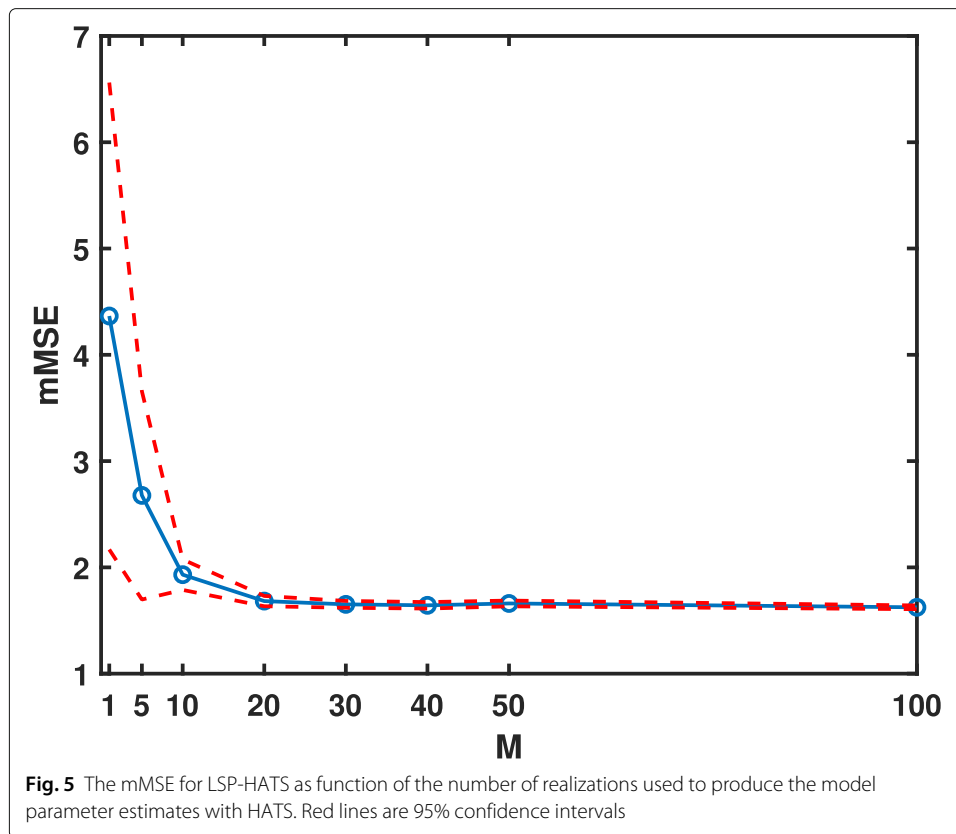
(WV); and finally the continuous wavelet transform using the Morlet wavelet (CWT). To allow the comparison with the other TF methods, the scales for evaluating the scalogram (4) are computed as $a = \frac{\beta}{f}$, where f denotes the frequency in hertz and the constant β is the center frequency of the Morlet wavelet [34].

Each method is optimized to evaluate it at its best performance in terms of MSE. For HANN, the window length $N_w \in \{16, 32, 64, 128, 256\}$ is optimized, while for WOSA, the optimized parameter is the number of windows $K \in \{2, 4, 8, 12, 16\}$, where the total length of all included windows is the total data length of 256 samples. For WV and CWT, a spectral magnitude scaling parameter is used to adjust the estimate magnitude to the true spectrum magnitude.

The expected value of the MSE, or mean MSE (mMSE), is computed approximated as the average of 100 independent realizations. The boxplots of the MSE achieved with the different methods in the 100 simulations are presented in Fig. 4. The mMSE value for the MSE optimal estimator with the true parameters (LSP) is 1.606 and with parameters estimated from the 100 realizations with HATS (LSP-HATS) is 1.655. Not only the spectral estimate obtained using LSP-based MSE optimal kernels achieves the best mMSE as expected, but using the true parameters or those estimated with HATS leads to almost the same result.

The optimal mMSE for HANN and WOSA is 3.438 and 2.061, respectively, obtained with $N_w = 32$ for HANN and $K = 8$ for WOSA. The worst mMSE of 4.434 is, as expected, obtained with WV. The CWT performance is in between HANN and WOSA, with a mMSE of 2.431.





In order to establish how the number of realizations used to estimate the model parameters with HATS affects the results in the TF domain, we study the decrease of the MSE of LSP-HATS based on the parameter estimates obtained using an increasing number of realizations $M \in \{1, 5, 10, 25, 50, 100\}$. The mMSE, computed as average on 100 independent simulations, is plotted as function of the number of realizations used with corresponding 95% confidence interval in Fig. 5. The values are reported in Table 1. The mMSE is extremely low even with only $M = 10$ realizations. However, the number of realizations required for a reliable estimate in real data cases might be higher especially in the case of a low signal-to-noise ratio.

3.2 Classification of simulated stochastic signals

Three classes of stochastic signals are simulated from the model (8), using slightly different parameters and center frequency f_0 , reported in Table 2. The parameters L and c_r are the same for the three classes, whereas different parameters define the function q describing the instantaneous power, plotted in Fig. 6. The center frequency f_0 for each class is

Table 1 mMSE and standard errors (SE) for LSP-HATS with an increasing number of realizations M used

	M							
	1	5	10	20	30	40	50	100
mMSE	4.365	2.677	1.931	1.683	1.652	1.643	1.660	1.624
SE	1.121	0.500	0.073	0.025	0.016	0.016	0.014	0.009

Table 2 Parameters for the three classes of simulated signals.

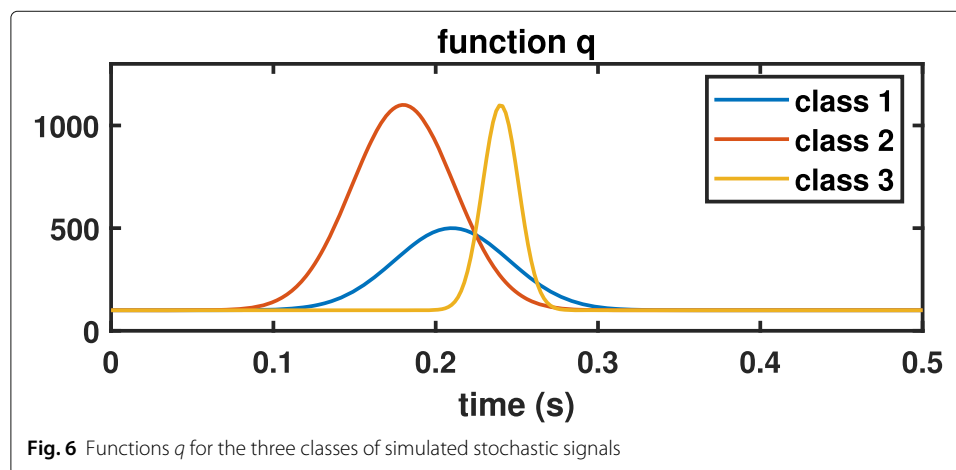
	L	a_q	b_q	c_q	c_r	f_0
Class 1	100	400	0.21	800	60,000	$\mathcal{U}(18, 22)$
Class 2	100	1000	0.18	1000	60,000	$\mathcal{U}(12, 16)$
Class 3	100	800	0.24	8000	60,000	$\mathcal{U}(6, 10)$

uniformly distributed around a mean \bar{f}_0 , with a jitter frequency from the mean of maximum 2 Hz, i.e., $f_0 \sim \mathcal{U}(\bar{f}_0 - 2, \bar{f}_0 + 2)$. A few random realizations from the three classes and their true WVS are presented in Figs. 7 and 8, respectively. Even though the classes cannot be distinguished by looking at the time-domain realizations, their true WVS is different.

The simulated dataset consists of 100 realizations for each class. A random partition in 80 realizations for training and 20 realizations for testing the neural network classifier is repeated 10 times. The LSP parameters are re-estimated from the 20 realizations used for independent testing and used in the computation of their LSP-HATS spectral estimates. The use of 20 realizations assures reliable estimates as shown in the simulation study, see Fig. 1, and it is compatible with the typical scenario in studies on physiological signals as EEG, where multiple trials of an experiment are available. The classification is repeated with the 10 different independent random sets of testing realizations.

The neural network classifier described in Section 2.4 is fed with TF features extracted with HANN ($N_w = 32$), WOSA ($K = 8$), WV, LSP-HATS, and CWT, where each feature is the spectral power at each TF point in the time interval $[0, 0.5]$ seconds and frequency up to 40 Hz. For computing the LSP-HATS kernels, the model parameters L, a_q, b_q, c_q, c_r were inferred from training and testing realizations independently.

The total classification accuracies of each method are summarized in Table 3, from which is evident how the LSP-HATS outperforms the state-of-the-art estimators with an accuracy of 86%. Similar performance is achieved with HANN and CWT, with accuracy of 65.0% and 69.7%, respectively. The methods with the worst classification performance are WOSA and WV, with an accuracy just above 50%. The confusion matrices for each method are reported in Tables 4, 5, 6, 7, and 8. The signals correctly classified appear on the diagonal of each confusion matrix; therefore, an ideal classifier would have 200 on each diagonal entry and 0 otherwise.



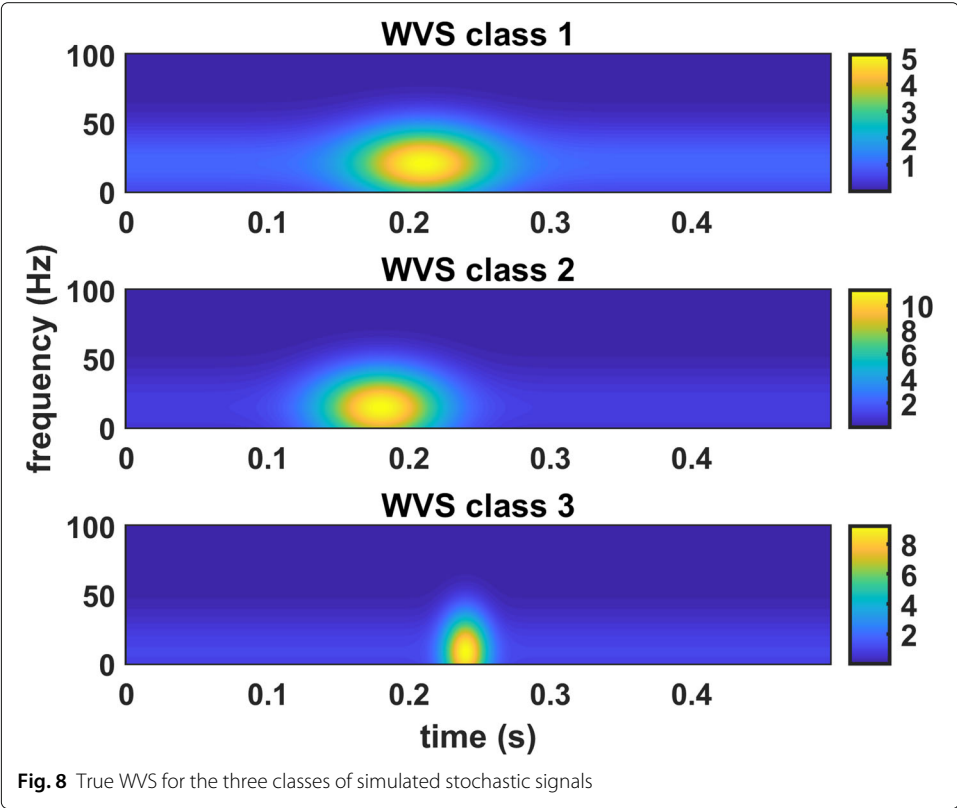
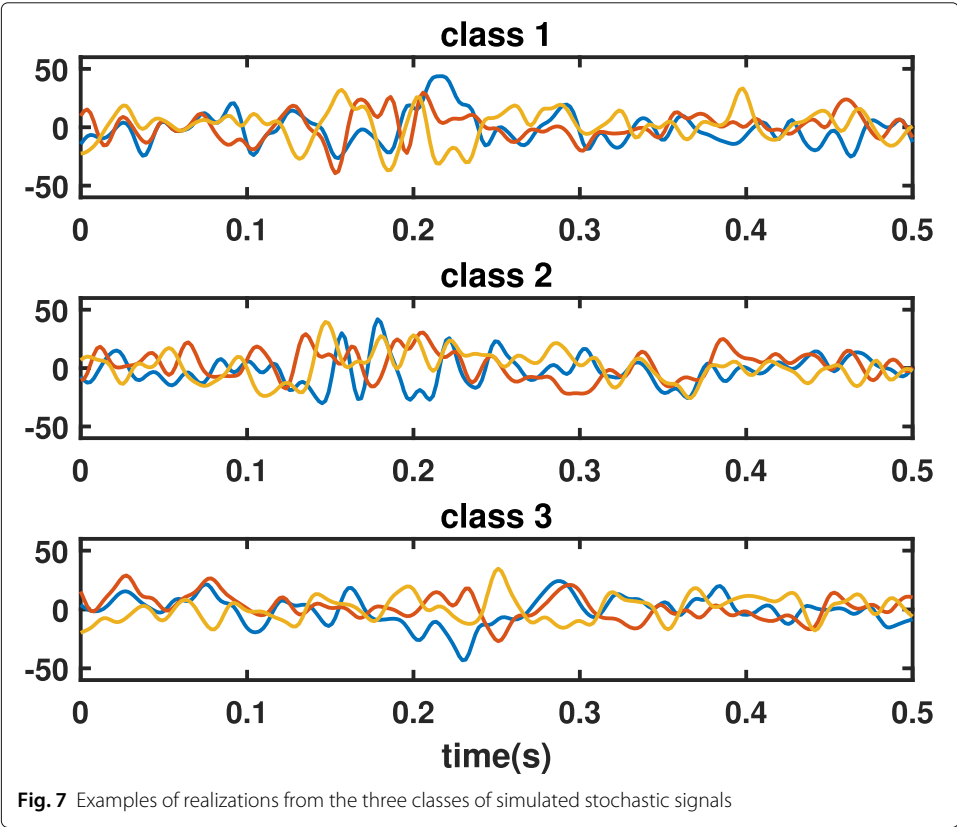


Table 3 Total classification accuracy obtained using features extracted with the different TF estimators in the simulation study, obtained on 10 independent simulations

	Method				
	HANN	WOSA	WV	LSP-HATS	CWT
Accuracy	65.0%	50.7%	54.8%	86.0%	69.7%

Table 4 Confusion matrix using features obtained with HANN ($N_w = 32$). Total classification accuracy is 65.0%

		Actual			
		Class 1	Class 2	Class 3	
Predicted	Class 1	108	32	59	54.3%
	Class 2	31	143	2	81.3%
	Class 3	61	25	139	61.8%
		54.0%	71.5%	69.5%	65.0%

Table 5 Confusion matrix using features obtained with WOSA ($K = 8$). Total classification accuracy is 50.7%

		Actual			
		Class 1	Class 2	Class 3	
Predicted	Class 1	136	25	21	74.7%
	Class 2	32	161	172	44.1%
	Class 3	32	14	7	13.2%
		68.0%	80.5%	3.5%	50.7%

Table 6 Confusion matrix using features obtained with WV. Total classification accuracy is 54.8%

		Actual			
		Class 1	Class 2	Class 3	
Predicted	Class 1	100	58	61	45.7%
	Class 2	39	109	19	65.3%
	Class 3	61	33	120	56.1%
		50.0%	54.5%	60.0%	54.8%

Table 7 Confusion matrix using features obtained with LSP-HATS. Total classification accuracy is 86.0%

		Actual			
		Class 1	Class 2	Class 3	
Predicted	Class 1	168	33	13	78.5%
	Class 2	27	162	1	85.3%
	Class 3	5	5	186	94.9%
		84.0%	81.0%	93.0%	86.0%

Table 8 Confusion matrix using features obtained with CWT. Total classification accuracy is 69.7%

		Actual			
		Class 1	Class 2	Class 3	
Predicted	Class 1	116	30	28	66.7%
	Class 2	35	138	8	76.2%
	Class 3	49	32	164	66.9%
		58.0%	69.0%	82.0%	69.7%

3.3 Application: classification of EEG signals

The data considered have been collected within a study on human memory retrieval, conducted at the Department of Psychology of Lund University, Sweden, during the spring of 2015. The EEG signals have been measured from one subject participating in the experiment. The encoding task consisted in associating a non-related word with a target picture belonging to one of three categories (“Faces,” “Landmarks,” “Objects”). A total of 180 trials were performed, 60 for each class. The EEG measurements were recorded from channel O1 according to the International 10–20 system, as primary visual areas can be found below the occipital lobes, and sampled with $f_s = 512$ Hz. Analogously to the simulation study, each time series has 256 equidistant samples during the time interval $[0, 0.5]$ seconds. A few signals from each class are shown in Fig. 9.

Similarly to the simulation study, the LSP model parameters L, a_q, b_q, c_q, c_r were inferred from 40 randomly selected realizations out of the total 60 for each class. The estimated

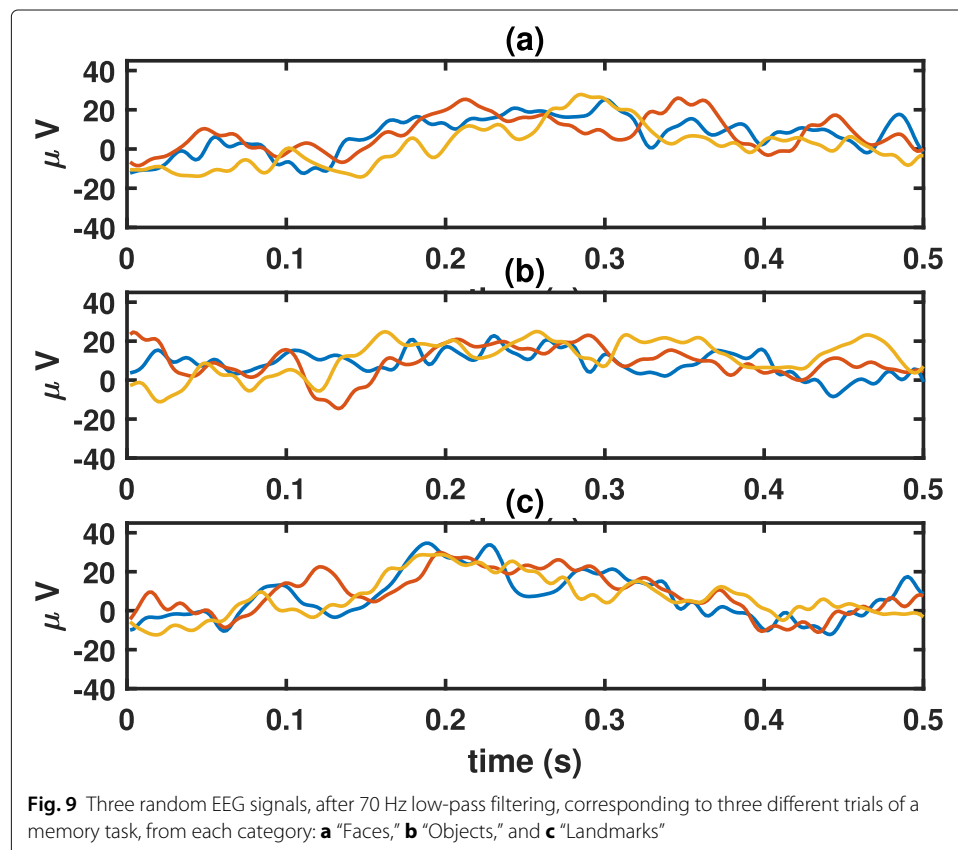
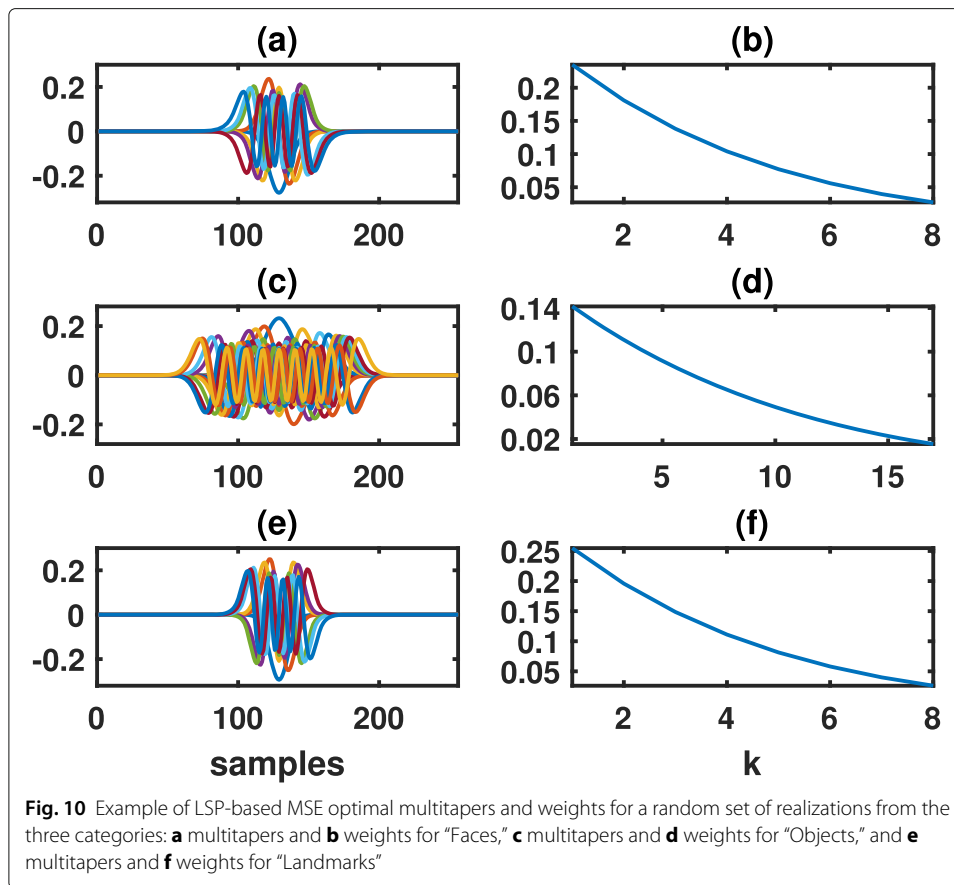


Fig. 9 Three random EEG signals, after 70 Hz low-pass filtering, corresponding to three different trials of a memory task, from each category: **a** “Faces,” **b** “Objects,” and **c** “Landmarks”



parameters are used to compute the LSP-HATS kernels and corresponding optimal multitapers to be used for computing the training data TF spectra. The multitapers and weights for the three classes, estimated from all available trials, are illustrated in Fig. 10. The spectral estimates obtained with LSP-HATS, HANN ($N_w = 128$), WOSA ($K = 8$), WV, and CWT are used to extract TF features, where each feature is the spectral power at each TF point in the time interval $[0, 0.5]$ seconds and frequency up to 40 Hz.

The remaining 20 realizations are used for independent testing of the network. The LSP parameters are re-estimated from these realizations and used in the computation of their LSP-HATS spectral estimates. The random partition in 40 realizations for training and 20 for testing is repeated 10 times, and the test is repeated with the different random sets of testing realizations. Classification accuracy is based on the 10 independent splits of the testing data.

The total classification accuracy of each method is reported in Table 9. The use of the TF features obtained with the proposed MSE optimal LSP-HATS estimator has resulted in significantly higher classification accuracy, compared to the use of the other estimators.

Table 9 Total classification accuracy of the memory encoding EEG signals of one subject, obtained on 10 independent different partition of the data

	Method				
	HANN	WOSA	WV	LSP-HATS	CWT
Accuracy	52.5%	39.8%	51.0%	70.8%	48.2%

As a note of caution, despite that the proposed approach is, in theory, optimal in terms of MSE, the performance in real applications depends both on how appropriate is the model for the data and on the purpose of the application. Nevertheless, in our case study, the higher quality of the TF representation has improved the accuracy of classification.

4 Conclusion

The purpose of this paper is to show how the MSE optimal WVS offers a significant improvement in practical applications, leading to a higher classification accuracy thanks to the greater quality of the TF features extracted with the proposed approach. The estimation of the model parameters thanks to a novel inference method allows the explicit computation of the corresponding MSE optimal kernel. The kernel is transformed into a robust and computationally efficient multitaper spectrogram, and a complete procedure for the LSP-inference MSE optimal spectral estimator from data is achieved.

In a simulation study evaluating the performance of the method in terms of MSE, the spectral estimate obtained using the optimal kernel achieves the best average MSE compared to state-of-the-art TF estimators, such as the Hanning spectrogram, the Welch method, the classical Wigner-Ville spectrum estimator, and the Morlet wavelet-based scalogram, as expected. More critical, computing the MSE optimal kernel from the parameters estimated with the novel HAnkel-Toeplitz Separation (HATS) inference method gives a similar result, which holds for parameter estimation from the small number of 10 realizations.

The performance of the proposed estimator is also evaluated in a classification simulation study, where it outperforms state-of-the-art estimators in terms of classification accuracy. The higher classification accuracy is also achieved in the episodic memory case study, where the parameter estimation on a suitable LSP model for EEG signals allows the extraction of improved features for classification.

Extensions of this research could consider a multidimensional model for extracting features from multichannel measurements.

Abbreviations

CWT: Continuous wavelet transform using the Morlet wavelet; EEG: Electroencephalography; HANN: Single Hanning window spectrogram; HATS: Hankel-Toeplitz separation; mMSE: Mean MSE; MSE: Mean square error; LSP: Locally stationary process; LSP-HATS: LSP with parameters estimated with HATS; TF: Time-frequency; WOSA: Welch 50% overlapped segment averaging with Hanning windows; WV: Classical Wigner-Ville spectrum estimate; WVS: Wigner-Ville spectrum

Acknowledgements

The data in the example was collected by Ines Bramao and Mikael Johansson, Department of Psychology, Lund University.

Authors' contributions

The idea for the paper was conceived jointly. RA investigated, implemented, and had the main responsibility for writing the manuscript. During the whole process, MS has provided feedback and guidance. Both authors read and approved the final manuscript.

Funding

The work has been funded by the eSENCE Strategic Research Program. Open access funding provided by Lund University.

Availability of data and materials

The dataset used in the example are not publicly available as they are part of a larger study outside the scope of this paper but are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The procedures followed to collect the data were in accordance with the Helsinki Declaration of 1975, as revised in 2000 and 2008.

Competing interests

The authors declare that they have no competing interests.

Received: 19 November 2019 Accepted: 6 April 2020

Published online: 01 May 2020

References

1. M. X. Cohen, R. Gulbinaite, Review: Five methodological challenges in cognitive electrophysiology. *NeuroImage*. **85**(Part 2), 702–710 (2014)
2. X. Sun, C. Qian, Z. Chen, Z. Wu, B. Luo, G. Pan, Remembered or forgotten?—an EEG-based computational prediction approach. *PLoS ONE*. **11**(12), 1–20 (2016)
3. S. Michelmann, B. P. Staresina, H. Bowman, S. Hanslmayr, Speed of time-compressed forward replay flexibly changes in human episodic memory. *Nat. Hum. Behav.* **3**(2), 143–154 (2019)
4. Z. Kurth-Nelson, G. Barnes, D. Sejdinovic, R. Dolan, P. Dayan, Temporal structure in associative retrieval. *ELife*. **4**, 1–18 (2015)
5. K. Samiee, P. Kovacs, M. Gabbouj, Epileptic seizure classification of EEG time-series using rational discrete short-time Fourier transform. *IEEE Trans. Biomed. Eng.* **62**(2), 541–552 (2015)
6. J. Meng, L. M. Merino, N. B. Shamlo, S. Makeig, K. Robbins, Characterization and robust classification of EEG signal from image RSVP events with independent time-frequency features. *PLoS ONE*. **7**(9), 1–13 (2012)
7. F. Kai, Q. Jianfeng, C. Yi, D. Yong, Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM. *Biomed. Sign. Process. Control*. **13**, 15–22 (2014)
8. I. Bramão, M. Johansson, Neural pattern classification tracks transfer-appropriate processing in episodic memory. *eNeuro*. **5**(4) (2018). <https://doi.org/10.1523/ENEURO.0251-18.2018>
9. A. Jafarpour, L. Fuentemilla, A. J. Horner, W. Penny, E. Duzel, Replay of very early encoding representations during recollection. *J. Neurosci.* **34**(1), 242–248 (2014)
10. A. Jafarpour, A. J. Horner, L. Fuentemilla, W. D. Penny, E. Duzel, Decoding oscillatory representations and mechanisms in memory. *Neuropsychologia*. **51**, 772–780 (2013)
11. M. Navarrete, J. Pyrzowski, J. Corlier, M. Valderrama, M. Le Van Quyen, Original research paper: Automated detection of high-frequency oscillations in electrophysiological signals: methodological advances. *J Phys Paris*. **110**(Part A), 316–326 (2016)
12. S. R. Cole, B. Voytek, Review: Brain oscillations and the importance of waveform shape. *Trends Cogn. Sci.* **21**, 137–149 (2017)
13. M. X. Cohen, Opinion: Where does EEG come from and what does it mean? *Trends Neurosci.* **40**(4), 208–218 (2017)
14. R. Silverman, Locally stationary random processes. *IRE Trans. Inf. Theory*. **3**(3), 182 (1957)
15. P. Wahlberg, M. Hansson, Kernels and multiple windows for estimation of the Wigner-Ville spectrum of Gaussian locally stationary processes. *IEEE Trans. Sig. Process.* **55**(1), 73–84 (2007)
16. L. Stanković, D. Mandić, M. Daković, M. Brajović, Time-frequency decomposition of multivariate multicomponent signals. *Signal Process.* **142**, 468–479 (2018)
17. G. Matz, F. Hlawatsch, *IEEE Trans. Inf. Theory*. **52**(3), 1067–1086 (2006)
18. G. S. Cunningham, W. J. Williams, Kernel decomposition of time-frequency distributions. *IEEE Trans. Signal Process.* **42**, 1425–1442 (1994)
19. F. Cakrak, P. J. Loughlin, Multiple window time-varying spectral analysis. *IEEE Trans. Signal Process.* **49**(2), 448–453 (2001)
20. M. Hansson-Sandsten, Optimal multitaper Wigner spectrum estimation of a class of locally stationary processes using Hermite functions. *EURASIP J. Adv. Signal Process.* **2011**, 980805 (2011)
21. R. Anderson, M. Sandsten, Inference for time-varying signals using locally stationary processes. *J. Comput. Appl. Math.* **347**, 24–35 (2019). <https://doi.org/10.1016/j.cam.2018.07.046>
22. R. Anderson, M. Sandsten, in *2018 26th European Signal Processing Conference, EUSIPCO 2018, vol. 2018-September*, Classification of EEG signals based on mean-square error optimal time-frequency features (EUSIPCO European Signal Processing Conference, Rome, Italy, 2018), pp. 106–110. <https://doi.org/10.23919/EUSIPCO.2018.8553130>
23. J. Yang, H. Singh, E. L. Hines, F. Schlaghecken, D. D. Iliescu, M. S. Leeson, N. G. Stocks, Channel selection and classification of electroencephalogram signals: an artificial neural network and genetic algorithm-based approach. *Artif. Intell. Med.* **55**(2), 117–126 (2012). <https://doi.org/10.1016/j.artmed.2012.02.001>
24. A. Craik, Y. He, J. L. Contreras-Vidal, Deep learning for electroencephalogram (EEG) classification tasks: a review. *J. Neural Eng.* **16**(3), 031001 (2019). <https://doi.org/10.1088/1741-2552/ab0ab5>
25. R. Hellerstedt, M. Johansson, Competitive semantic memory retrieval: temporal dynamics revealed by event-related potentials. *PLoS ONE*. **11**(2), 0150091 (2016)
26. I. Bramão, M. Johansson, Benefits and costs of context reinstatement in episodic memory: an ERP study. *J. Cogn. Neurosci.* **29**(1), 52–64 (2017)
27. A. M. Sayeed, D. L. Jones, Optimal kernels for nonstationary spectral estimation. *IEEE Trans Signal Process.* **43**(2), 478–491 (1995)
28. P. D. Welch, The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoustics*. **AU-15**(2), 70–73 (1967)
29. L. Cohen, *Time-Frequency Analysis*. (Prentice-Hall, Upper Saddle River, 1995)
30. S. T. Smith, Covariance, subspace, and intrinsic Cramér-Rao bounds. *IEEE Trans. Signal Process.* **53**(5), 1610–1630 (2005)

31. C. M. Bishop, *Pattern Recognition and Machine Learning. Information science and statistics*. (Springer, Springer-Verlag Berlin, Heidelberg, 2006)
32. G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, B. Kingsbury, Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Proc. Mag.* **29**(6), 82–97 (2012). <https://doi.org/10.1109/MSP.2012.2205597>
33. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*. (MIT Press, Cambridge, USA, 2016)
34. H.-G. Stark, *Wavelets and Signal Processing - An Application-Based Introduction*. (Springer, Springer-Verlag Berlin, Heidelberg, 2005)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
