

Single-trial event-related potentials with wavelet denoising

R. Quian Quiroga^{a,*}, H. Garcia^b

^a*Sloan-Swartz Center for Theoretical Neurobiology, California Institute of Technology, MC 216-76, Pasadena, 91125 CA, USA*

^b*Raul Carrea Institute for Neurological Research, FLENI, Montaneses 2325, 1428 – Buenos Aires, Argentina*

Accepted 23 October 2002

Abstract

The application of a recently proposed denoising implementation for obtaining event-related potentials (ERPs) at the single-trial level is shown. We study its performance in simulated data as well as in visual and auditory ERPs. For the simulated data, the method gives a significantly better reconstruction of the single-trial event-related responses in comparison with the original data and also in comparison with a reconstruction based on conventional Wiener filtering. Moreover, with wavelet denoising we obtain a significantly better estimation of the amplitudes and latencies of the simulated ERPs.

For the real data, the method clearly improves the visualization of both visual and auditory single-trial ERPs. This allows the calculation of better averages as well as the study of systematic or unsystematic variations between trials. Since the method is fast and parameter free, it could complement the conventional analysis of ERPs.

© 2002 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Wavelet; Denoising; Wiener; Electroencephalogram; Evoked potential; Single-trial

1. Introduction

Event-related potentials (ERPs) can be roughly defined as the changes of the ongoing electroencephalogram (EEG) due to stimulation. They have low amplitude in comparison with the background EEG activity and in consequence, they are hardly visualized in the single-trials. The usual way of improving the visualization of the ERPs is by averaging the responses over several trials. However, ensemble averaging relies on the basic assumption that the ERPs consist of an invariant pattern time-locked to the stimulus, laying on an independent stationary stochastic EEG signal. Although ensemble averaging has been successfully used since the '50 s, the previous assumption is in strict sense not valid (see e.g. Başar, 1980). In fact, averaging implies a loss of information related to systematic or unsystematic variations between the single-trials and furthermore, these variations (such as latency jitters) can affect the validity of the average ERP as representative of the single responses.

Several techniques have been proposed in order to improve the visualization of the ERPs. Most of these approaches involve filtering of the single-trial traces, in parti-

cular using techniques based on the Wiener formalism, which provides an optimal filtering in the mean square error sense (Walter, 1969; Doyle, 1975). However, these approaches have the common drawback of considering the signal as a stationary process and, since the ERPs are compositions of transient responses with different time and frequency localizations, they are not likely to give optimal results. A natural step forward is to implement time-varying strategies, as proposed by de Weerd (1981) and de Weerd and Kap (1981). In this respect, the Wavelet Transform (WT) provides a time-frequency decomposition that is shown to be very suitable for ERP analysis, due to its optimal resolution both in the time and in the frequency domain (Bartnik et al., 1992; Bertrand et al., 1994; Schiff et al., 1994; Demiralp et al., 1999; Quian Quiroga and Schürmann, 1999; Quian Quiroga et al., 2001). This study presents a recently proposed denoising implementation based on the WT (Quian Quiroga, 2000) to obtain the ERPs at the single-trial level.

2. Simulated data

ERPs were constructed by superimposing 3 Gaussian functions to simulated background EEG activity (see Fig. 1). As with real data, we will name the peaks: P1, N2 and P3. We further introduced some random fluctuation in the position of the peaks in order to resemble latency variability

* Corresponding author. Div. of Biology, MC: 216-76, California Institute of Technology, Pasadena, CA 91125, USA. Tel.: +1-626-395-8337; fax: +1-626-795-2397.

E-mail address: rodri@vis.caltech.edu (R. Quian Quiroga).

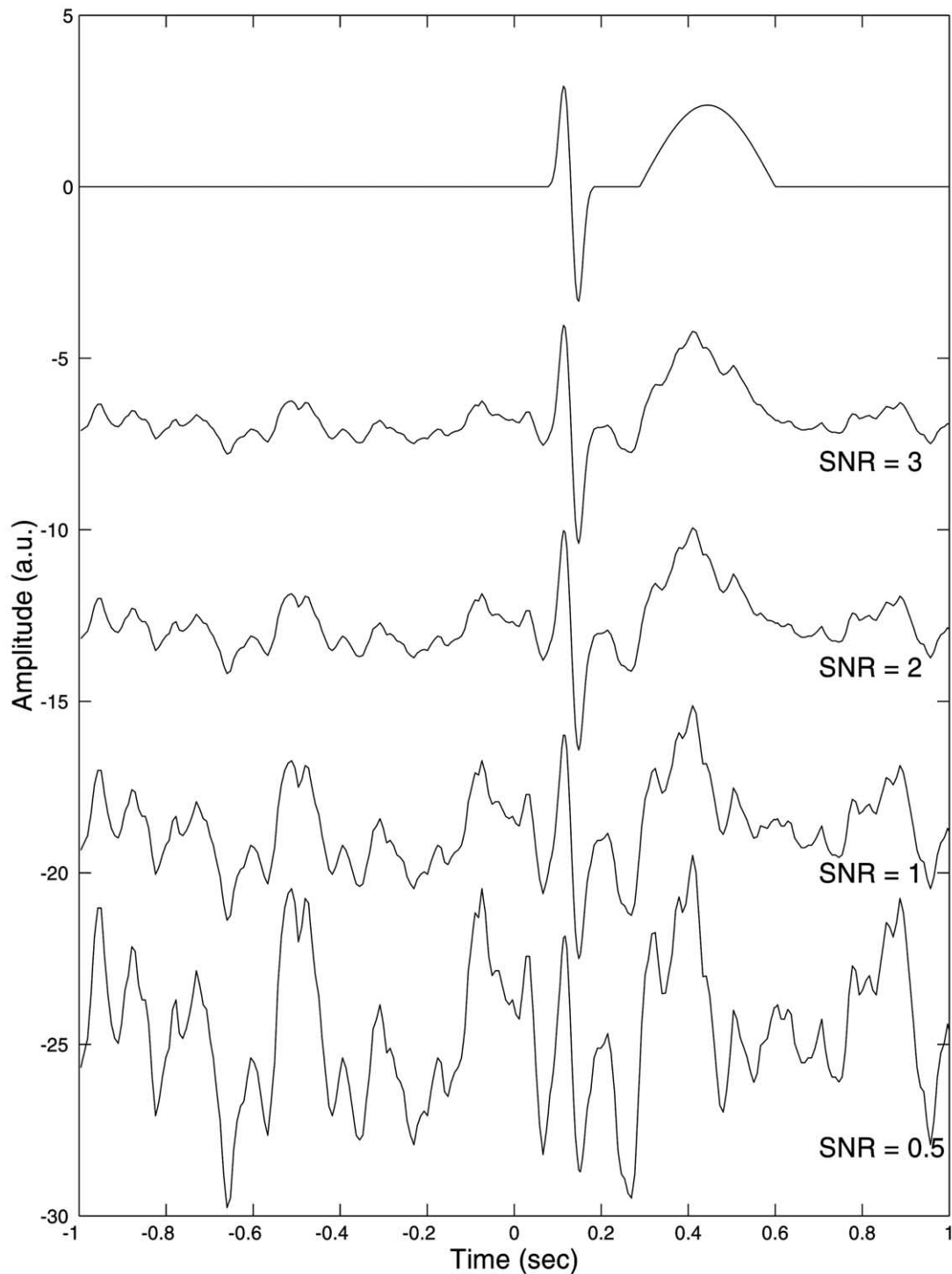


Fig. 1. Simulated event-related potential (upper plot) and single-trial realizations with added background EEG, for different SNR values.

seen in real data (ranges: 90–125, 120–155 and 400–650 ms, for the P1, N2 and P3, respectively). The background EEG signals were surrogates calculated from a short segment of ongoing EEG activity (without stimulation). In order to maintain the power spectrum, surrogates are usually constructed applying the Fourier transform, shuffling the phases, and then applying the inverse transform. In our

case we use an implementation proposed by Schreiber and Schmitz (2000) that besides the power spectrum, also maintains the amplitude distribution of the original signal. We generated 100 single-trials of 2 s each, with a sampling rate of 256 Hz (resembling the recording conditions of the next section) and different signal to noise ratios (SNR), as shown in Fig. 1. The SNR was defined as the ratio of standard

deviations of the ‘clean’ ERP signal and the one of the surrogate EEGs, in the time-range containing the ERPs (between 0 and 850 ms). Note that for SNR lower than 1 the ERPs are hardly recognizable in the single-trials. This is stressed by the fact that the shapes of the ERPs look very similar to the ones of the ongoing EEG.

Results with wavelet denoising will be compared to those obtained by applying conventional Wiener filtering, as described by Doyle (1975). For quantifying the performance of wavelet denoising, we calculated the root mean squared errors (RMS) of the single-trial denoised signals and compared them to the ones of the Wiener filtered signals and to the ones of the noisy original signals (in all cases, RMS values were calculated in the time range where the ERPs occur, i.e. between 0 and 850 ms). In order to test for statistical significance, for each SNR we performed analysis of variance (ANOVA) tests with factor signal type (original, denoised and Wiener filtered). When differences between groups were significant we further performed post-hoc comparisons using Scheffe’s correction.

We also quantified the error in the estimation of the amplitude and latency of the simulated peaks for the 3 type of signals (original, denoised and Wiener filtered). Each peak was identified in the single-trials as the maximum (minimum) in a proper time window (i.e. 55–170, 95–210 and 330–680 ms for the P1, N2 and P3, respectively). Then, the error in the single-trial amplitude (latency) estimation $\sigma = |x_i - \hat{x}_i|$ was defined, where for each peak and trial, x_i is the correct amplitude (latency) and \hat{x}_i is the estimated one (in the case of the amplitude, for each peak all x_i ’s are the same by construction). As in the previous case we performed ANOVA comparisons for each SNR level. Whenever significance at the $P < 0.05$ level was reached, post-hoc comparisons were performed using the Scheffe’s correction.

3. Subjects and experimental set up

We will also study ERP recordings obtained with an oddball paradigm upon two different stimulus modalities: visual event-related potentials (VERP) and auditory event-related potentials (AERP). The robustness of the method is stressed by the fact that although an (almost) identical implementation was used for both data sets, these were taken in different laboratories and under different recording settings. For both data sets, trials with artifacts were identified by visual inspection of the recordings and discarded for further analysis.

3.1. Pattern visual event-related potentials

In normal subjects, VERPs were obtained by using a checkerboard light pattern (sidelength of the checks 50’). Two different visual stimuli were presented in a pseudo-random order (oddball paradigm; $N = 200$ stimuli): 75% of the stimuli were the so called ‘non-target’ (a color reversal of the checks) and the other 25% were the deviant stimuli or ‘target’ (also a color reversal but with a half check displa-

cement of the pattern). Subjects were instructed to ignore the non-target stimuli and to count the number of appearances of the target ones (see Quian Quiroga and Schürmann, 1999 for more details on the experimental setup). Scalp recordings were obtained from frontal (F1, F2), central (C3, Cz, C4), parietal (P3, P4) and occipital (O1, O2) electrodes with linked earlobes references. In this study we will only show results on the O1 electrode. Sampling rate was 250 Hz and after bandpass filtering in the range 0.1–70 Hz, about 2 s of data (256 samples pre- and 256 samples post-stimulation) were saved on a hard disk. Inter-stimulus intervals varied randomly between 2.5 and 3.5 s.

3.2. Auditory event-related potentials

In two normal subjects, ERPs were obtained in an eyes-open condition from 20 scalp electrodes distributed according to the 10–20 system, with linked earlobes references. In this study, only results from the Cz electrode will be described. As in the case of VERP an oddball paradigm was used ($N = 100$ stimuli), the non-target stimuli (75%) being tones of 1000 Hz and the target ones (25%) being tones of 500 Hz. For each trial, after digital filtering in the range 1–70 Hz, 1 s of data was saved on a hard disc (from 0.2 s pre- to 0.8 s post-stimulation). Sampling rate was 204.8 Hz and the data of each trial was extended with symmetric border conditions in order to have 256 data points per trial.

4. Description of the method

4.1. Wavelet Transform: brief theoretical background

The WT gives a time-frequency representation of a signal that has two main advantages over previous methods: (a) an optimal resolution both in the time and in the frequency domains; and (b) the lack of the requirement of stationarity of the signal. It is defined as the convolution between the signal $x(t)$ and the wavelet functions $\psi_{a,b}(t)$

$$W_{\psi}X(a, b) = \langle x(t) | \psi_{a,b}(t) \rangle \quad (1)$$

where $\psi_{a,b}(t)$ are dilated (contracted) and shifted versions of a unique *wavelet function* $\psi(t)$

$$\psi_{a,b} = |a|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right) \quad (2)$$

(a, b are the scale and translation parameters, respectively). The WT gives a decomposition of $x(t)$ in different scales, tending to be maximum at those scales and time locations where the wavelet best resembles $x(t)$. Moreover, Eq. (1) can be inverted, thus giving the reconstruction of $x(t)$.

The WT maps a signal of one independent variable t onto a function of two independent variables a, b . This procedure is redundant and not efficient for algorithmic implementations. In consequence, it is more practical to define the WT only at discrete scales a and discrete times b by choosing the set of parameters $\{a_j = 2^{-j}; b_{j,k} = 2^{-j}k\}$, with integers j, k .

Contracted versions of the wavelet function match the high frequency components of the original signal and on the other hand, the dilated versions match the low frequency oscillations. Then, by correlating the original signal with wavelet functions of different sizes we can obtain its details at different scales. These correlations with the different wavelet functions can be arranged in a hierarchical scheme called multiresolution decomposition (Mallat, 1989). The multiresolution decomposition separates the signal into ‘details’ at different scales, the remaining part being a coarser representation of the signal called ‘approximation’.

In this study a 5 level decomposition was used, thus having 5 scales of details (d_1 – d_5) and a final approximation (a_5). The lower levels give the details corresponding to the high frequency components and the higher levels the ones corresponding to the low frequencies. Quadratic bi-orthogonal B-Splines (Cohen et al., 1992) were chosen as the basic wavelet functions due to their similarity with the event-related responses (thus having a good localization of the ERPs in the wavelet domain), and due to their optimal time-frequency resolution (for more details see Cohen et al., 1992; Chui, 1992).

4.2. Denoising of the event-related potentials

We will now describe the denoising implementation for a real dataset of 30 trials VERPs, corresponding to a single subject. The gray curves in Fig. 2 show the decomposition of the averaged VERP. On the left side we plot the wavelet coefficients and on the right side the actual components/decomposition. The sum of all the reconstructions gives again the original signal (gray curve of the uppermost right plot). Note that the P100–N200 response is correlated mostly with the first post-stimulus coefficient in the details d_4 – d_5 and the P300 is mainly correlated with the coefficients at about 400–500 ms in a_5 . This correspondence is easily identified because: (1) the coefficients appear in the same time (and frequency) range as the ERPs; and (2) they are relatively larger than the rest due to phase-locking between trials (coefficients related with background oscillations are diminished in the average). In consequence, a straightforward way to partially avoid the fluctuations related with the ongoing EEG is just by equaling to zero those coefficients not correlated with the ERPs. However, the choice of these coefficients should not be solely based on the average ERP; indeed it should also consider the time ranges in which the single-trial ERPs are expected to occur (i.e. some neighbor coefficients may be included in order to allow for latency jitters). In this respect, we propose first to choose the coefficients correlated with the event-related responses from the average signal and then to heuristically adjust this set of coefficients by comparing the outcomes of the denoised single-trial ERPs with the raw data. Then, we should mainly check that: (1) the denoising implementation will not filter ERPs appearing with latency shifts; and (2) the method does not introduce spurious changes in the peaks of interest (e.g.

as would happen if the set of coefficients is not sufficient for a proper reconstruction of the ERPs).

The black traces on the left side of Fig. 2 show the coefficients kept for the reconstruction of the P100–N200 and the P300 responses and the black curves on the right side show the contributions of each level obtained by eliminating all the other coefficients. Note that in the final reconstruction of the averaged response (black curve in the uppermost right plot) background EEG oscillations are filtered. We should remark that this is usually difficult to achieve with a Fourier filtering approach (especially in averages of a few number of trials) due to the different time and frequency localizations of the P100–N200 and P300 responses, and also due to the overlapping frequency components of these peaks and the ongoing EEG. In this context, the main advantage of wavelet denoising over conventional filtering is that we can select different time windows for the different scales.

Once the coefficients of interest are identified from the average ERP, we can apply the same procedure to each single-trial, thus diminishing the contribution of background activity not related to the ERPs. This, of course, does not filter spontaneous oscillations appearing in the same time range and with the same frequency composition as the ERPs. In fact, wavelet denoising may generate waveforms spuriously looking like ERPs just by filtering ongoing EEG oscillations. This effect is stronger when the reconstruction is done from very few wavelet coefficients, thus having a very narrow filtering both in time and frequency. We therefore propose to compare results with those obtained with background EEG test signals in order to discard misinterpretations due to an unfortunate selection of coefficients.

In summary, the method consist in the following steps:

1. the activity of the average ERP is decomposed in different scales (i.e. frequency bands) and times by using the wavelet multiresolution decomposition;
2. wavelet coefficients correlated with the ERPs are identified and the remaining ones are set to zero. The chosen coefficients should cover a time range in which the single-trial ERPs are expected to occur;
3. the inverse transform is applied, thus obtaining a denoised average ERP;
4. the denoising scheme defined by the previous steps (keeping the coefficients chosen in point 2) is applied to the single-trials;
5. validity of the method can be checked by applying the same procedure to background EEG test signals.

We remark that once the coefficients are chosen (steps 1–3), the method is parameter free and does not need to be adjusted for different EEG/ERP ratios, number of trials or other differences between subjects, electrodes, etc. A simple implementation of the method in Matlab as well as a sample data set can be obtained from <http://www.vis.caltech.edu/~rodrri>.

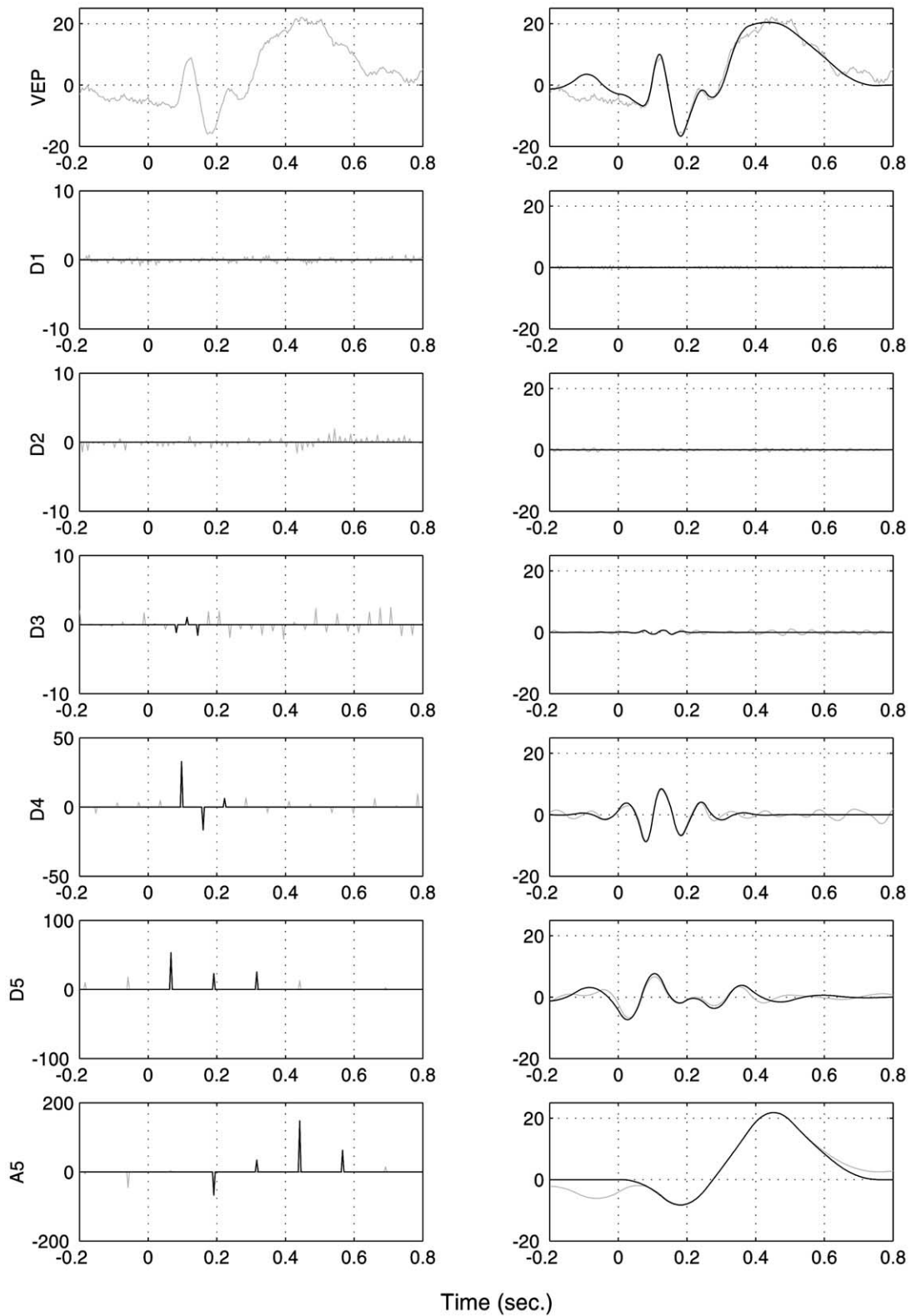


Fig. 2. Multiresolution decomposition and reconstruction of an averaged visual evoked potential (target stimuli, O1 electrode). Gray curves: original decomposition and reconstruction; black curves: denoised decomposition and reconstruction. D_1 – D_5 are the details at different scales and A_5 is the last approximation.

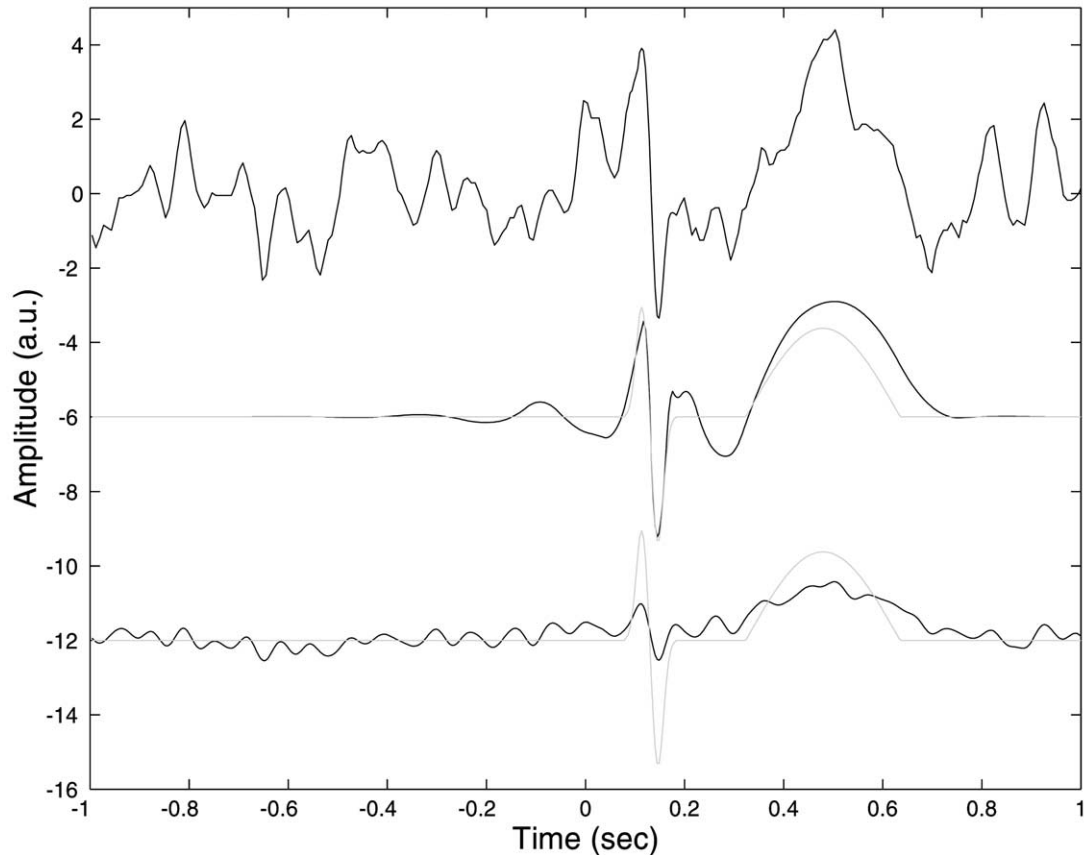


Fig. 3. Simulated single-trial ERP with SNR = 1. The upper trace corresponds to the ERP with added background EEG activity. The two lower black traces correspond to the reconstruction of the single-trial ERPs with wavelet denoising and Wiener filtering (lowest trace), respectively. The gray traces are the ERP without background EEG.

5. Results

5.1. Application to simulated data

Before applying the method to real data, we will quantify its performance in the simulated datasets and compare it to the one obtained with Wiener filtering. For denoising the data we used a similar implementation to the one shown in Fig. 2.

Fig. 3 shows one simulated single-trial with SNR = 1. The upper curve is the simulated ERP with added background EEG, and the two lower black traces are the reconstructions with wavelet denoising and Wiener filtering, respectively. Gray traces correspond to the original 'noise-free' ERP. We first note that the reconstruction of the original ERP with Wiener filtering is very poor. This is due to the time-varying characteristics of the ERP and as a consequence, there is no time-invariant filter that is suitable for all components. On the other hand, with wavelet denoising the amount of background activity is clearly reduced and the ERPs are much easily identified in the single-trials.

Fig. 4 shows the RMS values for the 3 signals, for different SNRs. Differences between the different type of signals (original, denoised and Wiener filtered) were highly significant for all SNRs ($P < 0.001$) and we further

performed post-hoc comparisons using Scheffe's correction (significant differences are shown with the horizontal segments in Fig. 4, all others were non-significant). In general, we observe that wavelet denoising gives the lowest RMS values. As expected, for high SNRs the single-trial ERP estimation with wavelet denoising is as good as with the original signal. On the contrary, the performance of Wiener filtering with high SNRs is significantly worse (as also seen in the previous figure). For SNR = 0.5 all signals have a relatively large RMS and the best performance is achieved with Wiener filtering. However, in this case Wiener filtering gives nearly a flat line, which of course is a bad estimation of the single-trial ERPs. The poor outcome of the Wiener filtered signals for both low and high SNRs will be more evident in the following analysis.

In Fig. 5 we show the error σ of the estimation of the amplitudes and latencies for the 3 peaks. Especially for the estimation of the single-trial amplitude of the 3 peaks, the performance with Wiener filtering is significantly worse than with wavelet denoising and with the original signals. On the contrary, wavelet denoising significantly improves the single-trial amplitude estimation for all SNR values for the P3, and for SNRs lower than 2 for the P1. For higher SNRs,

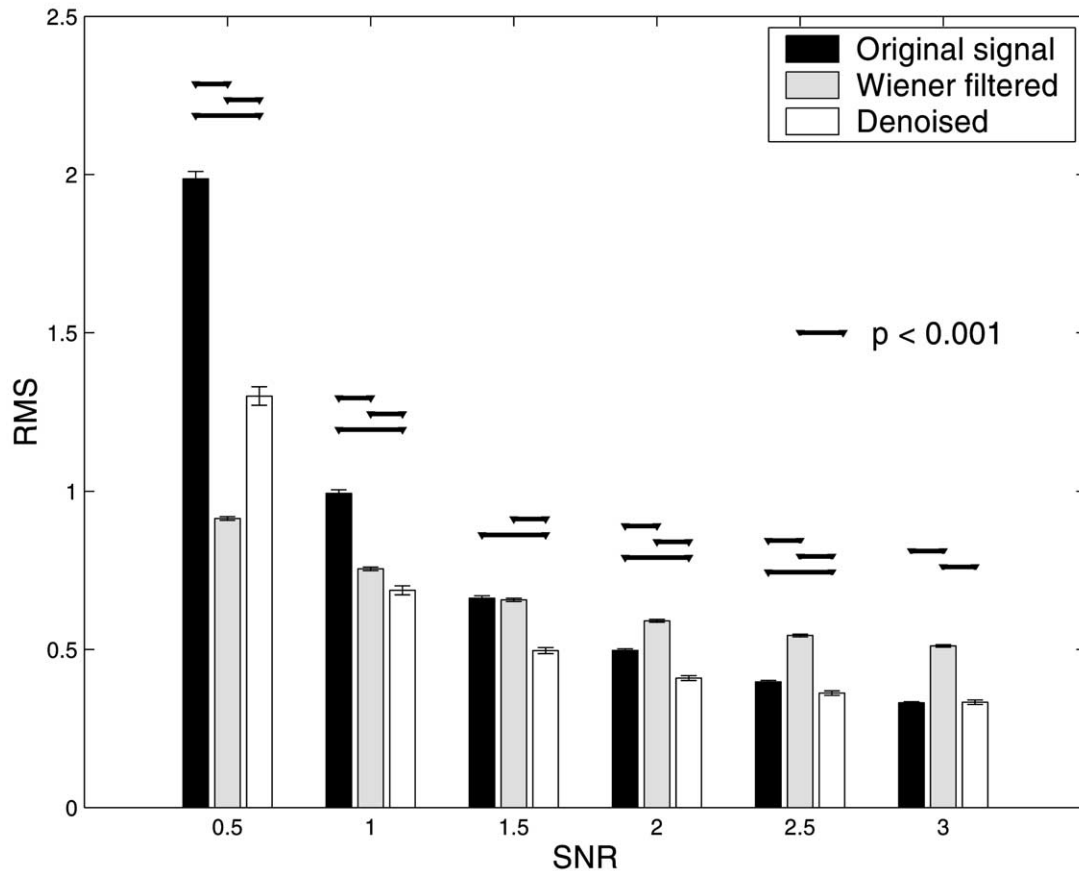


Fig. 4. RMS error for the original, denoised and Wiener filtered signals. Note the best performance of the denoised signals for nearly all SNRs.

the σ obtained with wavelet denoising for the P1 is lower than for the original signal, but without reaching significance. A similar tendency is observed for the N2, but in no case differences between denoised and original signals were significant.

For the estimation of the P3 latencies, wavelet denoising performs the best except for SNR = 0.5, where the performance of all methods was statistically the same. For the P1 and N2, latency estimations with wavelet denoising are comparable to those with the original signal, without any statistical difference. In general, latency estimations from the Wiener filtered signals are the worst for SNR = 2 and larger, and comparable to the other two signals for SNR < 2.

Table 1 summarizes the performance of the 3 signals (D: wavelet denoised; W: Wiener filtered and O: original signal) for the different measures and SNRs. In the table we mark whether one of the signals gave a significantly better performance than the other two, or whether two of them gave a better performance than the third one (but without significant differences between the first two). We observe that for most SNRs, wavelet denoising performs the best in terms of RMS, or based on the error of estimation of the single-trial amplitude and latencies, especially for the P3.

We finally checked that results of denoising ERP data differ from those of denoising ongoing EEG data (i.e. without ERPs). As mentioned in Section 4.2, we want to verify

that the denoised single-trials show real ERPs, rather than just ongoing EEG oscillations appearing in the same time and frequency ranges as the expected ERPs. For doing this, we applied the same denoising implementation to 100 simulated single-trials with SNR = 1, and to simulated trials of EEG data with the same variance. The difference between both results was quantified by calculating the maximum

Table 1

Performance of the 3 type of signals (D: wavelet denoised; W: Wiener filtered and O: original signal) for the different measures and SNRs^a

	SNR					
	0.5	1	1.5	2	2.5	3
RMS	W	D	D	D	D	D, O
P1 amplitudes	D	D	D	D, O	D, O	D, O
P1 latencies	–	–	–	–	–	D, O
N2 amplitudes	D, O	D, O	D, O	D, O	D, O	D, O
N2 latencies	–	–	–	–	–	D, O
P3 amplitudes	D, W	D	D	D	D	D
P3 latencies	–	–	D	D	D	D

^a We mark whether one of the signals gives a significantly better estimation than the other two, or whether two of them give a better performance than the third one. For RMS, differences are at $P < 0.001$ level and for amplitude and latency estimations, differences are at the $P < 0.05$ level.

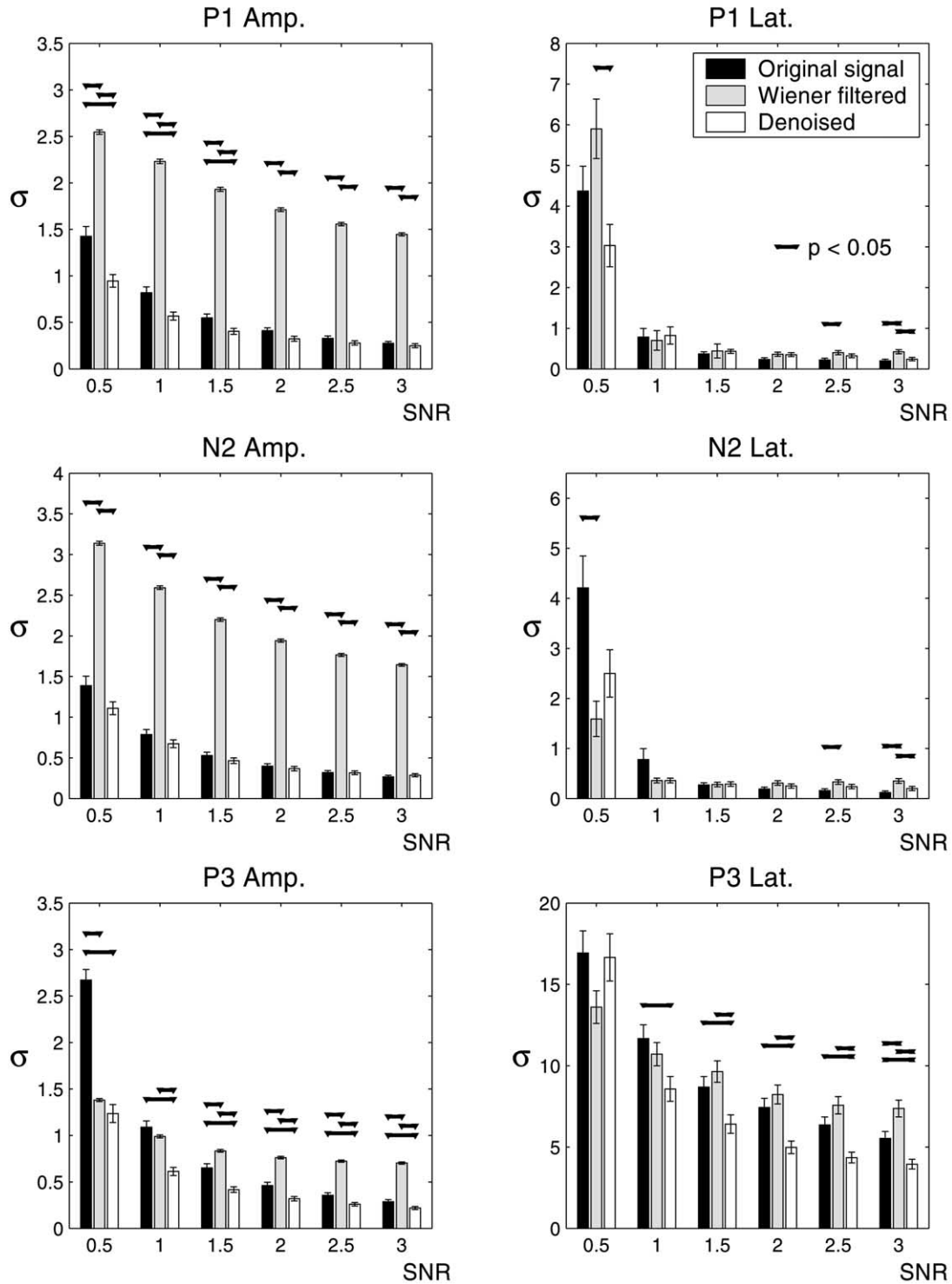


Fig. 5. Errors (σ) in the estimation of the single-trial amplitude and latencies of the 3 peaks from the original, denoised and Wiener filtered signals. Again, note the best performance of the denoised signal for most cases.

(minimum) of each peak (using the same time windows as before) and their latency variance. As expected, in all cases the amplitudes of the denoised ERPs were significantly larger than those of the denoised EEG ($P < 10^{-8}$) and the latency variances were significantly lower ($P < 10^{-5}$).

5.2. Application to visual event-related potentials

Fig. 6 shows the first 15 single-trials and the average ERP for a single subject (same subject and denoising implementation shown in Fig. 2). Note that with denoising

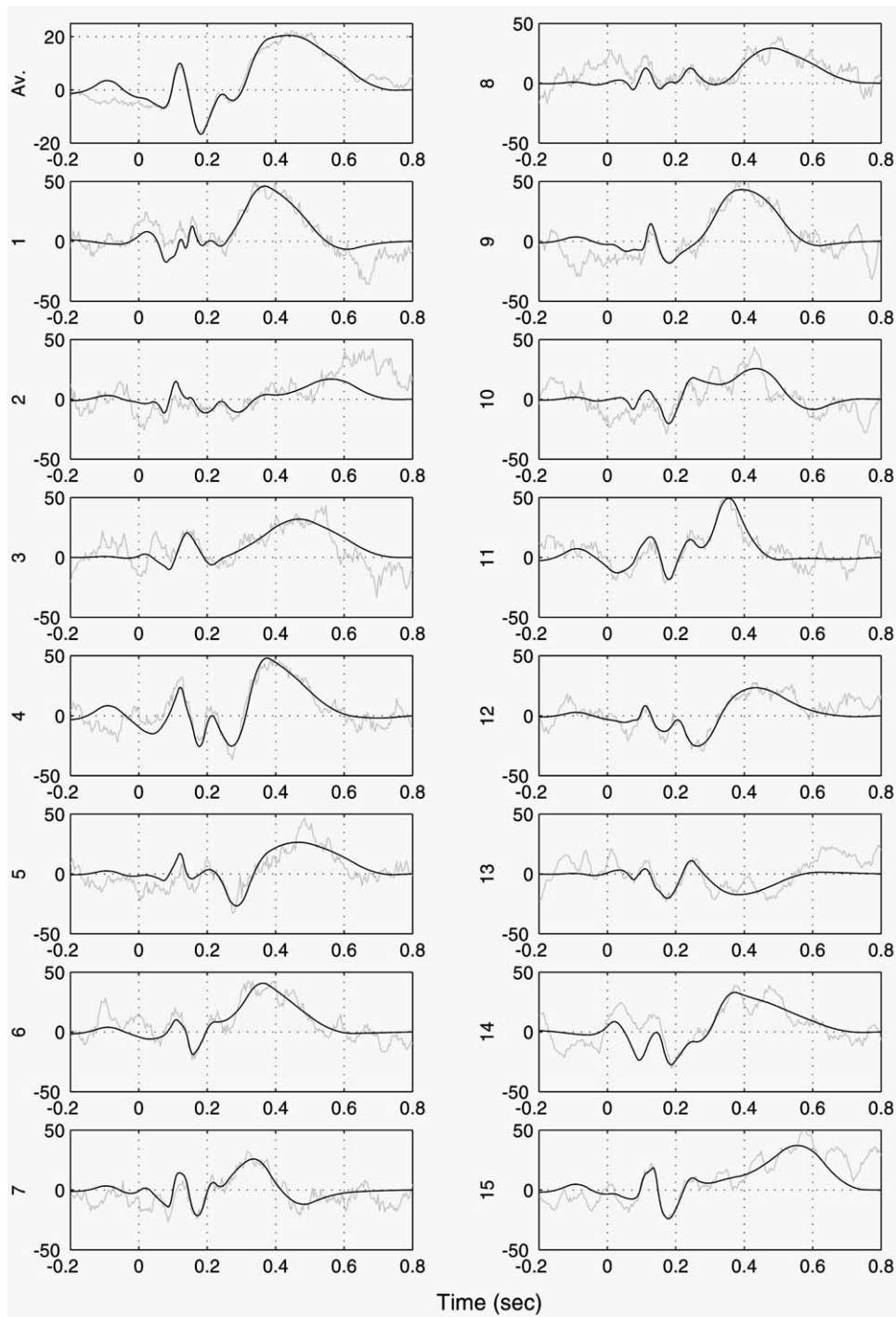


Fig. 6. Single-trials corresponding to the data of Fig. 2. Gray lines: original data; black lines: denoised data. Note how after denoising, the ERPs can be better visualized in the single trials. On the y-axis, values are in μV .

(black curves) we can distinguish the P100–N200 and the P300 in most of the trials. Note also that these responses are not easily identified in the original signal (gray traces) due to their similarity with the ongoing EEG. We can also observe some variability between trials, as for example in trials #2 and #13 where the ERPs are practically not present, or trials #1, #4, #6, #7, #9, #11, in which the P3

appears earlier than in the average. For an easier visualization, in Fig. 7 we plot the single-trial ERPs with and without denoising (left and right side plot, respectively) by using contour plots. In the denoised plot we observe between 100 and 200 ms a yellow/red pattern followed by a blue one corresponding to the P100–N200 peaks. The more unstable and wider yellow/red pattern at about

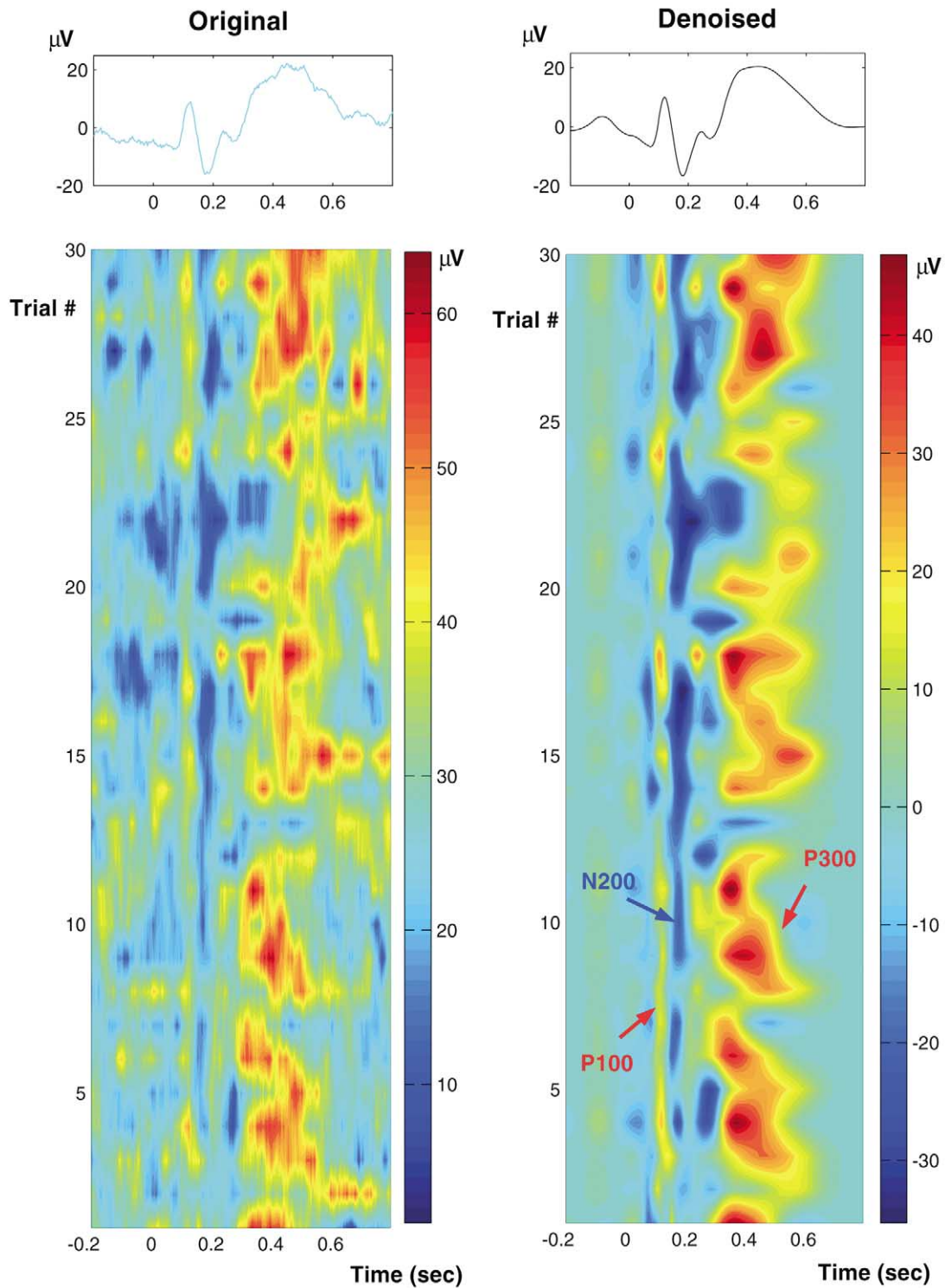


Fig. 7. Contour plot of the single trials shown in the previous figure. After denoising, it is easier to follow the evolution of the evoked responses with trial number.

400–600 ms corresponds to the P300. Noteworthy, all these patterns are more difficult to be recognized in the original signal (left plot).

Fig. 8 shows the outcome of the method for another subject. Again, we note that the ERPs are better visualized

after denoising. Indeed, in the original signal the P100–N200 peaks are hardly recognizable and the P300 is blurred by high frequency activity. We also note that in this case, there is a large latency variation of the P300 latency, something not seen in the average signal.

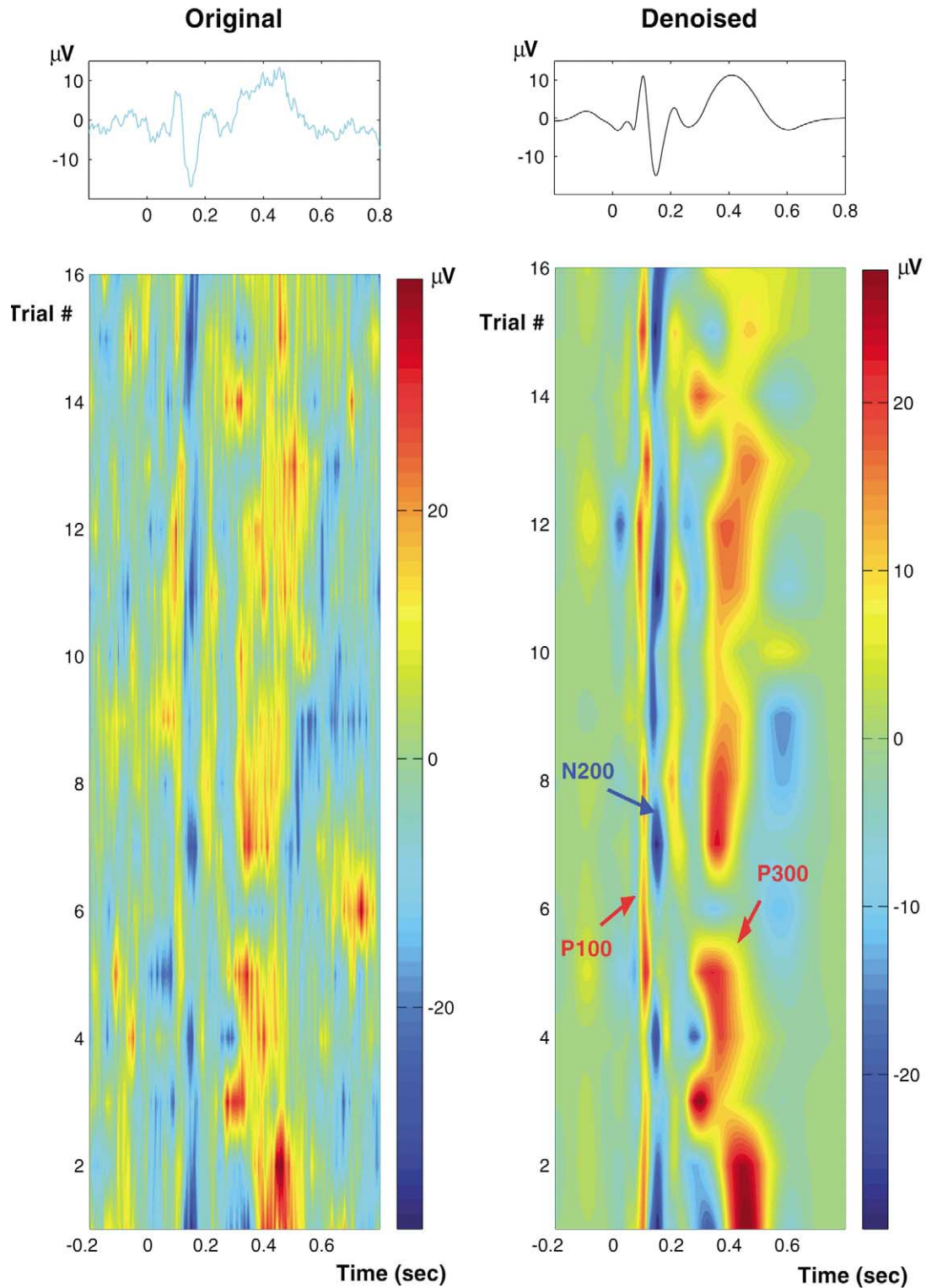


Fig. 8. Same as the previous figure for another subject. Note the higher latency variability of the P300.

5.3. Application to auditory event-related potentials

Fig. 9 shows the average AERPs of a typical subject (492) and the corresponding single-trial responses with and without denoising. In this case we show only the results upon

non-target stimuli, the ones upon the target ones being qualitatively similar to those shown in the previous section. In the denoised plots, we observe a blue pattern at about 100 ms after stimulation corresponding to the N100, followed by a yellow/red pattern corresponding to the P2. As in the case

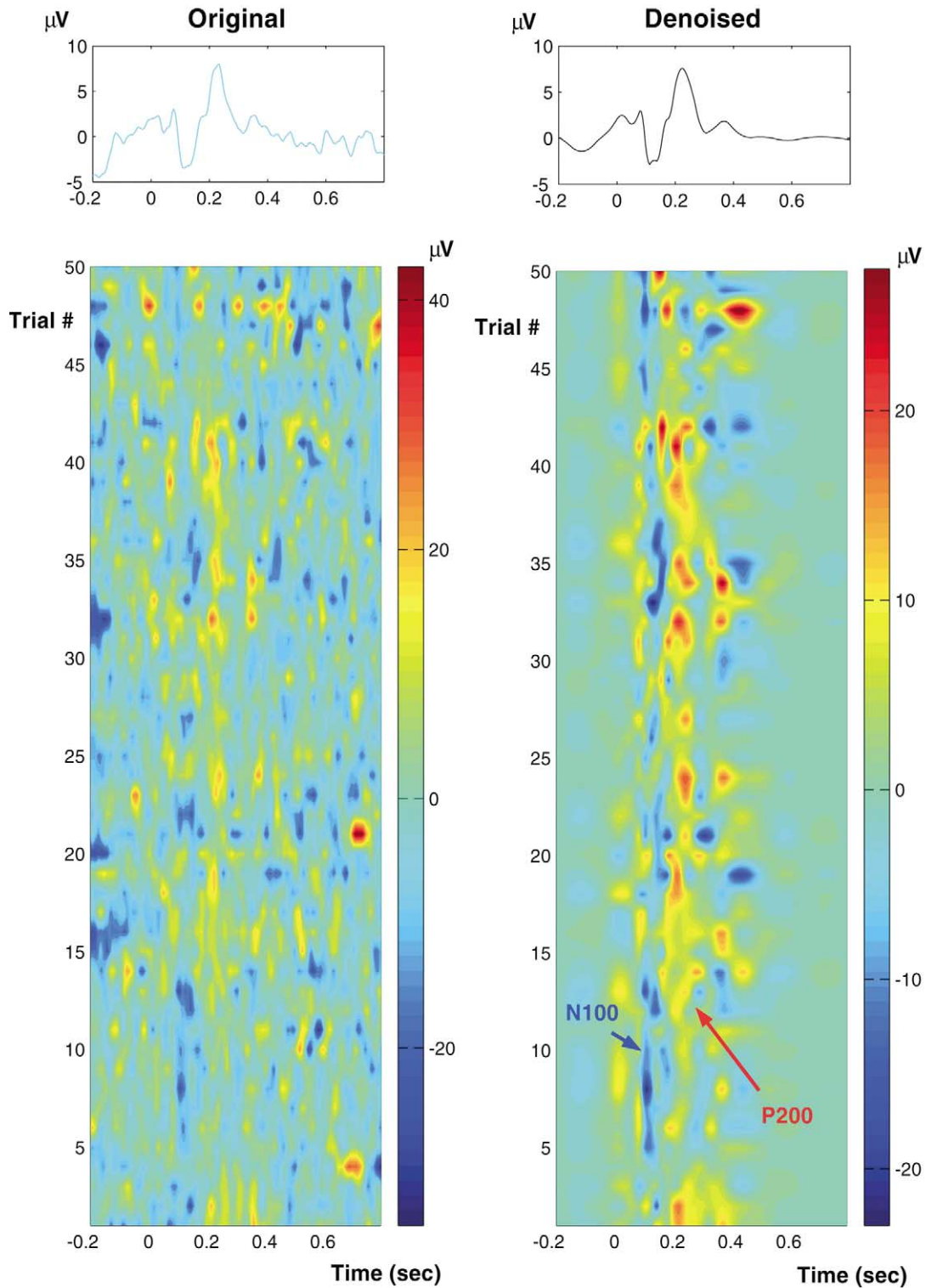


Fig. 9. Non-target AERPs from electrode Cz (subject 492). Note that single-trial ERPs are hardly recognizable from the original signal.

with visual stimulation, in many single-trials the ERPs are not well defined. We further note that between trials 15 and 40 the N100 response appears with some delay. Therefore, two ways of improving the averages are by: (1) selecting only those single-trials where the ERPs are well defined

(selective averaging); and (2) correcting for latency jitters between trials, as proposed by Woody (1967). For the selective averages, we calculated the cross-correlation between the denoised averages and the denoised single-trials and select those (denoised) trials with a cross-correlation larger

than a certain value (0.4). From these selected trials, we then calculated the jitter corrected averages by aligning the maximums of the N100 peaks to the maximum of the average.

In Fig. 10 we show the original and denoised averages,

the selective averages and the jitter corrected ones (all of them after denoising) for 3 different subjects and for a control EEG signal (without ERPs; also averaged over the same number of trials). For the first two subjects the selec-

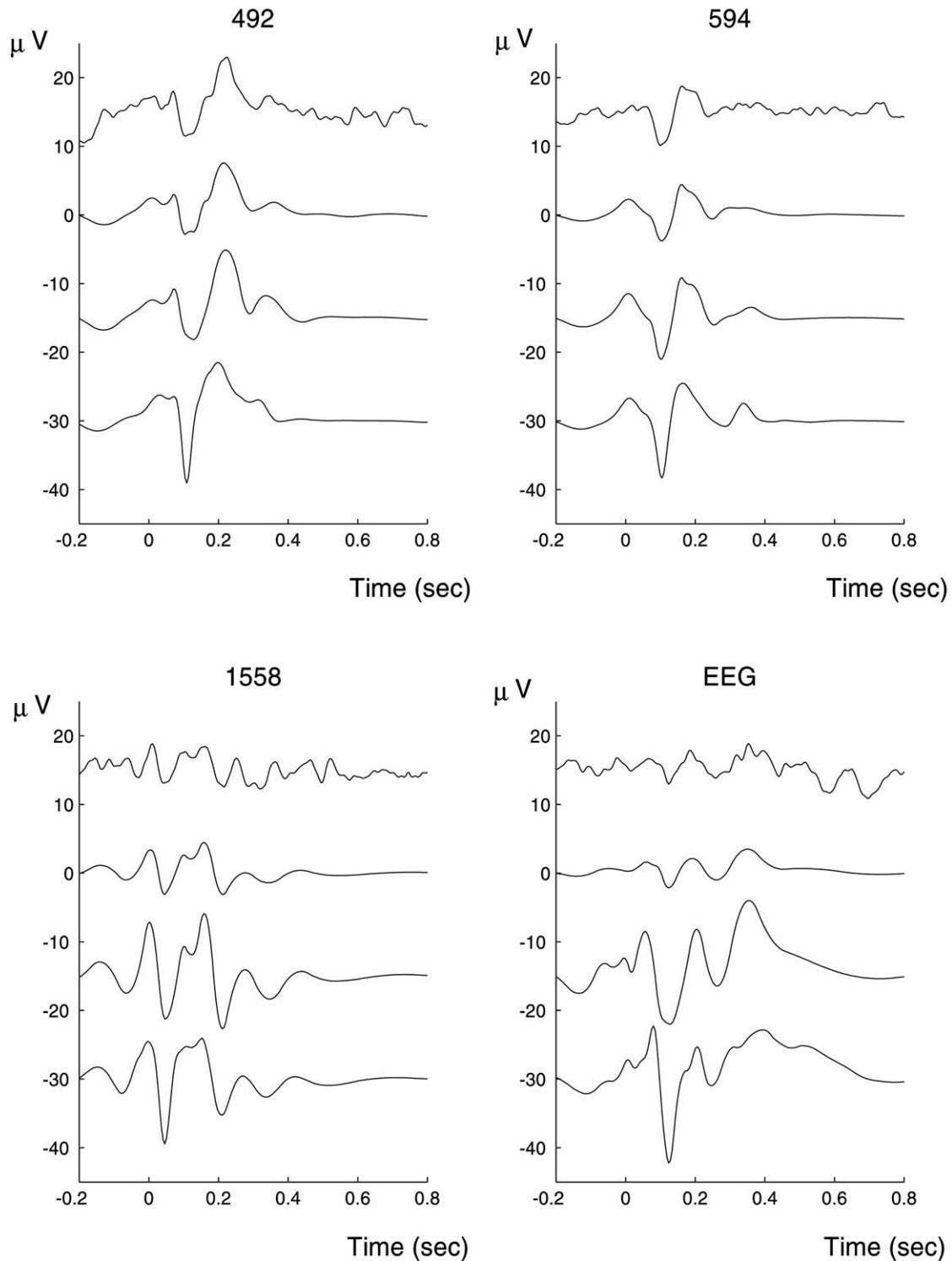


Fig. 10. From top to bottom: Averaged AERPs, denoised averages, selective averages and jitter corrected averages, for 3 subjects and for an ongoing EEG signal (without ERPs). Note how selective and jitter corrected averages increase the definition of the evoked potentials (except for subject 1558 where changes are comparable to the ones obtained in the control EEG signal).

Table 2

Amplitude and latencies of the N100 peak for 3 subjects and a test EEG signal obtained with the different type of averages

	Original averages	Denoised averages	Selective averages	Jitter corrected averages
492 amplitudes (μV)	- 3.5	- 2.8	- 3.1	- 9.0
594 amplitudes (μV)	- 4.8	- 3.8	- 6.0	- 8.3
1558 amplitudes (μV)	- 1.8	- 3.1	- 6.2	- 9.4
EEG amplitudes (μV)	- 2.2	- 2.1	- 7.0	- 12.2
492 latencies (s)	0.104	0.104	0.128	0.104
594 latencies (s)	0.099	0.104	0.104	0.099
1558 latencies (s)	0.040	0.045	0.045	0.040
EEG latencies (s)	0.123	0.123	0.123	0.123

tive averages show a better definition of the ERPs. The same holds for the jitter corrected averages, especially in the first case (492) due to its large latency variability, already shown in Fig. 9. Table 2 shows the amplitude and latency values for the N100 peaks obtained with the different averages. We first note that latency values do not change much for the different averages. On the contrary, the selective and jitter corrected averages clearly increases the amplitude of the peaks, as also seen in Fig. 10. The third subject (1558) should be treated more carefully because no clear ERPs are recognizable in the conventional average and also because the latency of the (eventual) N100 is too short. The lack of an average ERP can be due to a lack of responses in the single-trials or due to a high latency jitter between the trials. Although both the selective and the latency corrected averages seem to better localize some components, they also increase background oscillations. This shows that rather than improving the visualization of real ERPs, we are just aligning and filtering (with the denoising procedure) ongoing EEG activity. In fact, we obtain a similar result when applying these corrections to a test EEG signal (without ERPs), and we therefore conclude that for the last subject it is more likely that there were no real ERPs in the single-trials.

6. Discussion

We proposed a denoising implementation that improved the visualization of the single-trial ERPs and showed its application to simulated data as well as to different types of real ERPs. The usual way to visualize the ERPs is by means of ensemble averaging. However, when averaging, information about the variability between the single-trials is lost. Such information could be important in order to study processes such as habituation, sensitization, attention level, and it could also help to identify features of the ERPs when information from the average is not clear. In this respect, systematic changes related to habituation and sensitization processes have been recently reported using a denoising implementation similar to the one showed in this paper (Quian Quiroga and van Lujtelaar, 2002).

For the simulated data, wavelet denoising gave a significantly better estimation of the single-trial ERPs in comparison

with the original data and also in comparison with Wiener filtered data (Doyle, 1975). In general, differences between the original and denoised data became smaller when increasing the SNR. But not even in the case of very low noise (SNR = 3) the original data showed significantly better results than the denoised one. This stresses the fact that wavelet denoising does not introduce mayor artifacts in the reconstruction of the ERPs. In other words, the denoising of a clean ERP signal gives nearly the same ERP. This is not the case of Wiener filtering, which for high SNRs gave significantly higher errors in comparison with the original and denoised data.

Some of the most interesting features to study in single-trial ERPs are the changes in amplitude and latency of the peaks from trial to trial. In this respect, with wavelet denoising we obtained the best performance. This was remarkable when estimating the amplitude and latencies of the P3 and also the amplitudes of the P1. In the case of the N2 and also for the latency estimation of the P1, the performance of wavelet denoising was comparable to the one of the original signals. Again, errors in the estimations of amplitude and latencies with Wiener filtering were significantly the highest in most cases.

The variability between single-trials, besides providing interesting information per se, can also affect the validity of the average ERP as representative of the single-trial responses. Indeed, we showed how averages are improved by selecting those trials with good responses and also by correcting for latency jitters. These corrections led to narrower (i.e. better defined) peaks with a clear increase in their amplitudes. The use of selective averages as well as jitter corrected averages had been proposed long ago (Pfurtscheller and Cooper, 1975; Woody, 1967). The contribution of wavelet denoising at this respect is that it improves the identification of the single-trials responses, thus facilitating the construction of the selective or jitter corrected averages. We also showed that by comparison with the outcomes on ongoing EEG test signals, it could be checked whether such averages are not the spurious result of aligning and filtering of ongoing EEG oscillations. A similar check was also performed to validate the denoising implementation per se, where a very narrow filtering in time and frequency (i.e. selecting very few wavelet coefficients) could make filtered EEG oscillations look like real ERPs.

The main advantage of wavelet denoising over Wiener filtering is that wavelets give a time-variant decomposition. As a consequence, one can choose different filtering settings (i.e. wavelet coefficients) for different time ranges, thus matching the characteristics of the different event-related responses. On the other hand, with time-invariant approaches, such as Wiener filtering, it is not possible to find a unique implementation that will be suitable for all ERP components. A similar argument applies to other time-invariant filtering procedures. These limitations had already been noted by de Weerd (1981), who proposed a time-variant Wiener filtering of ERPs. In comparison with this approach, wavelet denoising has several advantages: (1) the wavelet decomposition offers an excellent resolution both in time and in frequency; (2) by applying the inverse WT, perfect reconstruction is achieved; and (3) the method is based on multiresolution decomposition, which is a very fast and recursive algorithm.

The use of wavelets for filtering average ERPs or for visualizing the ERPs in the single-trials has already been reported (Bartnik et al., 1992; Bertrand et al., 1994; Thakor et al., 1993). However, these works propose denoising implementations based exclusively on the average ERPs without considering latency variations in the single-trials. Efferen et al. (2000a–c) already noticed this caveat and proposed two solutions. In Efferen et al. (2000a), they introduced an initial latency correction (before denoising) by using an iterative procedure such as the one described by Woody (1967). This correction mainly assumes that event-related responses consist of a single component, or in the case of several components, that the latency variations from trial to trial are the same for all responses. A more sophisticated algorithm using denoising after embedding of the single-trials in a phase space was proposed in Efferen et al. (2000b,c). Briefly, by using a phase space reconstruction this algorithm looks for similarities in the shapes of the different single-trial traces (the range at which these similarities are searched depends on the embedding parameters). The main assumption is that the shape of the ERPs will be similar in the different trials even if appearing at different latencies. The caveat of such procedure is that single-trial ERP traces are in many cases not distinguishable from ongoing EEG activity. The implementation we used in this study is more straightforward and explicitly uses the knowledge of the time and frequency ranges in which the ERPs are expected to occur in the single-trials. The obvious disadvantage is that our implementation is not unsupervised and requires heuristic adjustment mainly by comparing the outcomes of the denoised single-trial ERPs with the raw data. However, we remark that once the wavelet coefficients are chosen, the method is parameter free and does not need to be adjusted for different EEG/ERP ratios, number of trials or other differences between subjects, electrodes, etc.

Acknowledgements

We thank Dr Martin Schuermann for the VERP data and to Farshad Moradi and Zoltan Nadasdy for comments on the manuscript. RQQ acknowledges support from the Sloan-Swartz foundation.

References

- Bartnik EA, Blinowska KJ, Durka PJ. Single evoked potential reconstruction by means of wavelet transform. *Biol Cybern* 1992;67:175–181.
- Başar E. EEG-Brain dynamics. Relation between EEG and brain evoked potentials. Amsterdam: Elsevier, 1980.
- Bertrand O, Bohorquez J, Pernier J. Time-frequency digital filtering based on an invertible wavelet transform: an application to evoked potentials. *IEEE Trans Biomed Eng* 1994;41:77–88.
- Chui C. An introduction to wavelets. San Diego: Academic Press, 1992.
- Cohen A, Daubechies I, Feauveau JC. Bi-orthogonal bases of compactly supported wavelets. *Comm Pure Appl Math* 1992;45:485–560.
- Demiralp T, Ademoglu A, Schuermann M, Basar-Eroglu C, Basar E. Detection of P300 waves in single-trials by the wavelet transform. *Brain Lang* 1999;66:108–128.
- de Weerd JPC. A posteriori time-varying filtering of averaged evoked potentials. I. Introduction and conceptual basis. *Biol Cybern* 1981;41:211–222.
- de Weerd JPC, Kap JJ. A posteriori time-varying filtering of averaged evoked potentials. II. Mathematical and computational aspects. *Biol Cybern* 1981;41:223–234.
- Doyle DJ. Some comments on the use of Wiener filtering for the estimation of evoked potentials. *Electroenceph clin Neurophysiol* 1975;38:533–534.
- Efferen A, Lehnertz K, Grunwald T, Fernandez G, David P, Elger CE. Time adaptative denoising of single-trial event-related potentials in the wavelet domain. *Psychophysiology* 2000a;37:859–865.
- Efferen A, Lehnertz K, Fernandez G, Grunwald T, David P, Elger CE. Single trial analysis of event related potentials: non-linear de-noising with wavelets. *Clin Neurophysiol* 2000b;111:2255–2263.
- Efferen A, Lehnertz K, Schreiber T, David P, Elger CE. Non-linear denoising of transient signal with application to event related potentials. *Physica D* 2000c;140:257–266.
- Mallat S. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans Pattern Anal Machine Intell* 1989;2:674–693.
- Pfurtscheller G, Cooper R. Selective averaging of the intracerebral click evoked responses in man: an improved method of measuring latencies and amplitudes. *Electroenceph clin Neurophysiol* 1975;38:187–190.
- Quian Quiroga R. Obtaining single stimulus evoked potentials with wavelet denoising. *Physica D* 2000;145:278–292.
- Quian Quiroga R, Schürmann M. Functions and sources of event-related EEG alpha oscillations studied with the Wavelet Transform. *Electroenceph clin Neurophysiol* 1999;110:643–655.
- Quian Quiroga R, van Luijtelaar G. Habituation and sensitization in rat auditory evoked potentials: a single-trial analysis with wavelet denoising. *Int J Psychophysiol* 2002;43:141–153.
- Quian Quiroga R, Sakowitz O, Basar E, Schuermann M. Wavelet Transform in the analysis of the frequency composition of evoked potentials. *Brain Res Protocols* 2001;8:16–24.
- Schiff SJ, Aldrouby A, Unser M, Sato S. Fast wavelet transformation of EEG. *Electroenceph clin Neurophysiol* 1994;91:442–455.
- Schreiber T, Schmitz A. Surrogate time series. *Physica D* 2000;142:346–382.
- Thakor N, Xin-rong G, Yi-Chun S, Hanley D. Multiresolution wavelet analysis of evoked potentials. *IEEE Trans Biomed Eng* 1993;40:1085–1094.
- Walter DO. A posteriori ‘Wiener filtering’ of average evoked responses. *Electroenceph clin Neurophysiol* 1969;27(Suppl.):61–70.
- Woody CD. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Med Biol Eng* 1967;5:539–553.