Signal analysis techniques for evoked and event-related potentials

2.1 Introduction

Event-related potentials (ERPs) and evoked potentials (EPs) have a low amplitude in comparison with the background EEG activity and, as a consequence, they are barely visible in single-trials. The usual way of improving the ratio between the E(R)P signal power and the background EEG power, i.e. the signal to noise ratio (SNR), is by averaging the response of several trials (Chiappa, 1997). Amplitude and latency are usually determined next. However, there are also other techniques to extract relevant information from E(R)Ps. This chapter gives an overview of the analysis techniques most used for evoked and event-related potentials. We will start this chapter with time domain analysis techniques. Subsequently, several techniques will be described for the frequency domain. Source localization, which is also often applied to E(R)P data, will be explained in the next chapter.

2.2 Time domain analysis techniques

2.2.1 Current Source Density / Surface Laplacian

The scalp potential distribution measured with EEG is blurred considerably relative to the potential distribution on the cortex due to the different conductivities of cortex, skull, and scalp (Nunez, 1990). Furthermore, the potential distribution depends strongly upon the location of the reference electrode or electrodes (Nunez et al., 1994; Gevins et al., 1989).

A method that can be used to deblur the EEG signal is to compute the reference-free Surface Laplacian (SL) (Hjorth, 1975; Nunez and Pilgreen, 1991; Yao, 2002b). The SL estimate is thought to be proportional to the current density flowing perpendicular to the skull into the scalp and is also called current source density (CSD) (Nunez et al., 1994) or scalp current density (SCD)(Perrin et al., 1989). It gives a more localized distribution of major components of electric activity as compared
Mathematically, the SL is the sum of the second order spatial derivatives of the scalp potential distribution (Yao, 2002b):

\[
\Delta V = \nabla^2 V = \nabla \cdot \nabla V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} \tag{2.1}
\]

The relation between the current source density and SL, when assuming no volumetric current sources in the scalp, is given by the Poisson equation:

\[
\nabla \cdot J = -\sigma \nabla^2 V \tag{2.2}
\]

where \(J\) is the current density, \(\sigma\) the conductivity of the scalp and \(V\) the scalp potential. According to this equation CSD is proportional to the divergence of the current density in the scalp (Perrin et al., 1987).

Two different SL estimation procedures have been proposed in literature: a local approach (Hjorth, 1975) and a more global approach (Perrin et al., 1989). The local SL is obtained by computing the difference between the potential at each electrode site and the average potential of its nearest four neighbours provided that the distance and angle between the electrodes are equal (Hjorth, 1975). Mathematically the second order derivative in 1D is then approximated by:

\[
\frac{\partial^2 V}{\partial x^2} \approx \frac{V(x_0 + \Delta x) - 2V(x_0) + V(x_0 - \Delta x)}{\Delta x} \tag{2.3}
\]

Now consider the surface of the head as a 2-dimensional space. By assuming equal distances between electrodes, assigning these distances unity and when the electrodes are positioned as in figure 2.2, equation 2.1 can be approximated by (Hjorth, 1975):

\[
\nabla^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} \approx (V_1 - V_0) + (V_2 - V_0) + (V_3 - V_0) + (V_4 - V_0) \tag{2.4}
\]

**Figure 2.1:** Potential map (left) and CSD map (right) of a visual evoked potential (VEP) recording. The CSD map shows more localized activity compared to the potential map.
where $V_0$ is the electric potential at electrode $P_0$ that is surrounded by electrodes $P_1, P_2, P_3, P_4$ with potentials $V_1, V_2, V_3, V_4$. A disadvantage of the local technique is that the SL cannot be estimated at each electrode, since the computation requires neighbouring electrodes (Hjorth, 1975).

The global approach uses all electrodes to determine the SL instead of only neighbouring electrodes. First a global interpolation function is constructed to represent surface potential distributions and then the SL is calculated for the global interpolation function (Yao, 2002a). Various functions have been used for the global interpolation e.g. 2D thin plate spline functions (Perrin et al., 1989), spherical spline functions (Perrin et al., 1989), and realistic spline Laplacians (Babiloni et al., 2001, 1996). In general, the global approach is a better approximation for the current source density than the local approach (Babiloni et al., 1995).

### 2.2.2 Principal component analysis and independent component analysis

In general, principal component analysis (PCA) and independent component analysis (ICA) are used for data compression and information extraction. In EEG analysis these techniques are mostly used for artifact rejection. They can easily be used to remove eye, electrocardiogram (ECG), muscle or MRI artifacts (Vigario et al., 2000; Jung et al., 2000; Iriarte et al., 2003; Vigario, 1997). PCA searches for components which are linear combinations of the signals observed, are uncorrelated and have maximum variation in the smallest number of components. The first principal component explains most of the variance, the second principal component explains most of the variance after the first component has been extracted, etc.

Mathematically, PCA relies on an eigenvector decomposition of the covariance matrix of the EEG signals. The covariance matrix contains all the covariances possible between the channels used and for a measured data matrix $M$ it can be defined by:

$$\text{cov}(M) = M^T M$$  \hspace{1cm} (2.5)
This equation assumes that the columns of $M$ have been mean centered, i.e. the mean is subtracted. The eigenvalues $\lambda$ of the square matrix $\text{cov}(M)$ can be obtained by solving:

$$\det(\text{cov}(M) - \lambda I) = 0$$  \hspace{1cm} (2.6)

where $I$ is the identity matrix. Solving equation 2.6 gives $n$ eigenvalues, where $n$ is the size of the covariance matrix. Subsequently, the eigenvectors $v_{\lambda_i}$ can be calculated by solving the following equation:

$$\text{cov}(M)v_{\lambda_i} = \lambda_i v_{\lambda_i}$$ \hspace{1cm} (2.7)

The eigenvectors represent the directions of the principal components in $n$-dimensional space and the eigenvalues represent a measure for the variance in that direction. Consequently, the eigenvector with the highest eigenvalue is the direction of the principal component that accounts for most of the variance, i.e. the eigenvector that corresponds to the strongest correlation in the dataset.

To reduce the data, the eigenvectors with a small eigenvalue can be ignored. For data extraction one can ignore other unwanted data such as eigenvectors that represent eye movements. By multiplying the chosen eigenvectors with the original data, the principal components are obtained:

$$Y = A_s M$$ \hspace{1cm} (2.8)

where $A_s$ is the orthogonal matrix containing the selected eigenvectors of the covariance matrix as the row vectors, and $Y$ contains the principal components. To return to the original dimensions, the transpose of matrix $A$ must be multiplied by $Y$:

$$M_s = A_s^T Y$$ \hspace{1cm} (2.9)

where $M_s$ is the compressed data.

An assumption made by PCA is that the data has a Gaussian distribution. If the data is not Gaussian, independent component analysis (ICA) can be used instead of PCA (Vigario et al., 2000). ICA searches for the most independent components instead of uncorrelated components. Statistically, uncorrelated is defined as:

$$E\{y_i y_j\} - E\{y_i\}E\{y_j\} = 0, \text{for } i \neq j$$ \hspace{1cm} (2.10)

and independent is defined as:

$$E\{g_1(y_i) g_2(y_j)\} - E\{g_1(y_i)\}E\{g_2(y_j)\} = 0, \text{for } i \neq j$$ \hspace{1cm} (2.11)
Signal analysis techniques for EPs and ERPs

Figure 2.3: PCA of a visual evoked potential (VEP) recording. Loadings (left) of each channel and the corresponding largest principal components (right). The loadings are defined by the eigenvectors.

where $g_1$ and $g_2$ can be any function and $E$ is the expected value. Independent variables are always uncorrelated, while uncorrelated does not always mean independent (Calhoun et al., 2001). Thus independence is a much stronger requirement than uncorrelatedness. Both requirements are equivalent concepts only in the case of Gaussian distributed signals (Vigario et al., 2000).

There are some assumptions that underlie ICA decomposition of EEG time series: 1) there must be at least as many channels as the number of independent sources, 2) sources are non-Gaussian 3) sources must be temporally independent (Vigario et al., 2000; Brown et al., 2001). ICA tries to find a matrix $W$ such that:

$$C = WM$$

(2.12)

where the components of $C$ are as independent as possible and $M$ is the original data matrix. Before using an ICA algorithm, some preprocessing must be performed on the data. As in PCA the data must be mean centered first. Secondly,
the observed data must be whitened, i.e. the data is linearly transformed to obtain uncorrelated components with variance that equals unity (Hyvarinen and Oja, 2000). This transformation can be performed by using PCA:

\[ \tilde{X} = D^{-\frac{1}{2}} A_s M \]  

(2.13)

where \( \tilde{X} \) is the transformed data and \( D \) is the diagonal matrix of the eigenvalues of the covariance matrix. The estimation of the matrix \( W \) is accomplished by minimizing or maximizing a cost function that represents dependence or independence: maximum non-Gaussianity estimation (kurtosis, negentropy), maximum likelihood estimation or minimum mutual information (Hyvarinen and Oja, 2000). To find the optimum of the cost function different iteration methods can be used depending on the choice of cost function, such as gradient and Newton-like methods.

For example, Infomax is an ICA procedure that is often used and that was introduced by Bell and Sejnowski (1995). This method calculates the matrix \( W \) by maximizing the entropy \( U \), which is a measure for mutual information:

\[ U = g(WM) \]  

(2.14)

where \( g \) is defined by:

\[ g(WM) = (1 + e^{-WM})^{-1} \]  

(2.15)

### 2.2.3 Cross-correlation

In statistics cross-correlation is a measure for the linear relationship between two variables. It is mostly used to examine the relation between two completely different parameters, for example arm length and median nerve somatosensory evoked potential (SEP) latency. In statistics the cross-correlation coefficient between two variables \( X \) and \( Y \) is defined by:

\[ \rho_{xy} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \]  

(2.16)

where \( \mu_X \) is the mean and \( \sigma_X \) is the standard deviation of variable \( X \). They are defined by:

\[ \mu_X = E(X) \]  

(2.17)

\[ \sigma_X = \sqrt{E(X^2) - E^2(X)} \]  

(2.18)

In the same way mean and standard deviation for variable \( Y \) can be calculated. The cross-correlation can have a value between -1 and 1. A value of -1 suggests a complete linear inverse correlation, a value of 1 complete linear direct correlation.
and a value of 0 a lack of linear relationship. If there is a positive relation the cross-correlation coefficient is also positive, while if there is a negative relation the cross-correlation coefficient is negative. In digital signal processing the cross-correlation is a measure of similarity of two signals as a function of the relative time $\tau$ between the signals:

$$
\rho_{xy}(\tau) = \frac{\sum_{t=1}^{T-\tau} [(X(t) - \mu_x)(Y(t + \tau) - \mu_y)]}{\sqrt{\sum_{t=1}^{T} [X(t) - \mu_x]^2 \sum_{t=1}^{T} [Y(t + \tau) - \mu_y]^2}}
$$

(2.19)

where $T$ is the total number of discrete time samples. The denominator of equation 2.19 is called the cross-covariance of variables $X$ and $Y$. Assuming that coordinated brain processing is reflected by a similar potential wave shape among the regions involved, covariance is sometimes used to investigate functional relationships between different brain areas (Gevins et al., 1987, 1989). For each electrode pair cross-covariance between their waveform segments is calculated. This gives a cross correlation function for each electrode pair. Next, the maximum as a function of $\tau$ of each function is determined and used as a measure for the functional relationship between the areas. Furthermore, cross-covariance and cross-correlation are both used in single trial analysis as described later in this chapter. Autocorrelation is the cross-correlation of a signal with itself. It is useful for finding repeating patterns in a signal.

### 2.2.4 Symmetry measures

A method to quantify symmetry of E(R)Ps is the laterality index. The laterality index is the amplitude difference between left and right hemisphere after stimulation or action of respectively right and left body side, divided by the sum of these amplitudes (Jung et al., 2003). This method is, for example, used for estimating the similarity of SEP peak amplitudes after stimulation of left and right median nerve. However, this method could also be used within a task recording bilateral activity.

A technique to investigate lateralized activity is by calculating event-related lateralizations (ERLs). The main goal of this technique is to enhance the detectability of lateralized activity by suppressing symmetric activity and all activity which is present in recordings of both sides. It computes the difference potentials between homologues electrodes contra and ipsilateral to the side of stimulus or action. As an example, the ERL for homologues electrode pair C3-C4 is defined as (Oostenveld et al., 2001):

$$
ERL = \frac{1}{2}[(C4 - C3)_L + (C3 - C4)_R]
$$

(2.20)
where $L$ and $R$ represent the measurements from the left and right task. Thus, for estimation of this difference potential left and right side need to be stimulated or moved one side at a time. Furthermore, it is assumed that activity in the contralateral hemisphere is similar for the left and right task. The same assumption is made for the ipsilateral hemisphere.

### 2.2.5 Single-trial analysis

When averaging single-trials for improving the SNR, one usually assumes that the E(R)P waveform is fixed and that its peak latency is constantly time-locked to the stimulus (Chiappa, 1997). However, information concerning individual trials and the variation in E(R)Ps is lost using this technique. Therefore, alternative techniques have been developed to determine single-trial latency, amplitude and/or other parameters.

The simplest single-trial analysis method used is “peak picking” (Smulders et al., 1994). This method takes the latency of the maximum amplitude in a specific window as latency of the peak of interest that normally occurs in this window. However, this technique often does not work, because the SNR is too low.

Rodionov et al. (2002) developed a more complex version of this technique, in which the number and amplitude of positive and negative deflections of the EEG is used to determine the variability of latency and amplitude of single-trials. This technique requires a lower SNR. A disadvantage of this technique is that the temporal resolution of the latency is very low due to summation of every five successive samples in the EEG signal.

Another single-trial analysis method calculates the cross-correlation or cross-covariance between the single-trials and a template. Next the maximum cross-correlation or cross-covariance is used as peak latency. In literature different templates are used:

1. half sine wave (Smulders et al., 1994; Pfefferbaum and Ford, 1988)
2. stimulus locked average ERP (Michalewski et al., 1986; Wastell and Kleinman, 1980; Thomas et al., 1989)
3. peak locked average ERP (Suwazono et al., 1994)
4. first two principal components of the average ERP (Kisley and Gerstein, 1999)

The stimulus locked average ERP may not be a perfect template for single-trial analysis, since the averaged peaks might be broadened in time and decreased in amplitude due to latency variation of the single-trial responses.

Another group of single-trial analysis methods models the ERP as a linear combination of multiple components (e.g. wavelets) whose component parameters (e.g.
Signal analysis techniques for EPs and ERPs

Latency and amplitude are calculated for each single-trial (Lange et al., 1997; Quian and Garcia, 2003; Truccolo et al., 2003; Heinrich et al., 1999) by using for example a least squares method (Lange et al., 1997). The components and initial parameters are often extracted from the average ERP (Lange et al., 1997; Quian and Garcia, 2003; Heinrich et al., 1999).

The quality of the results obtained by the described techniques depends mainly on the SNR of the signal. By filtering the signal before using it for single-trial analysis, the noise can be reduced and the SNR may be improved.

Single-trial analyses have been mainly used for ERP analyses, since ERP responses are sensitive to the subject’s performance, state of mind (Nishida et al., 1997), habituation, and have often a sufficient SNR (more than EPs).

![Figure 2.4: Single-trial plot of a standard oddball paradigm, whereby a target auditory stimulus is presented amongst more frequent standard auditory stimuli. Raw data is filtered (low pass 30 Hz, high pass 0.16 Hz) and segments with eye blinks have been removed. This plot only shows the target segments. Notice that the P300 peak is observable in most single-trials.]

2.3 Frequency domain analysis techniques

2.3.1 Fourier transform

The Fourier transform (FT) converts signals from the time domain to the frequency domain. It can be used to study frequencies in the signal or for filtering. Because EEG signals are digitally recorded nowadays, the discrete Fourier transform (DFT) is used to calculate the FT. The fast Fourier transform (FFT) is an efficient algorithm to compute the DFT. The (D)FT decomposes the original signal $s(t)$ in cosines and sines with a certain frequency, amplitude and phase. The frequencies of these cosines and sines are determined by the segment time $T_0$ and
the sample frequency $f_s$.

Mathematically the DFT for a signal $s(n)$ with $N$ sample points is given by (Roerdink, 1993):

$$S(k) = \sum_{n=0}^{N-1} s(n)e^{-\frac{i2\pi kn}{N}}, \quad k = 0, ..., N - 1 \quad (2.21)$$

where, according to Euler’s formula:

$$e^{-\frac{i2\pi kn}{N}} = \cos \left(\frac{-2\pi kn}{N}\right) + i \sin \left(\frac{-2\pi kn}{N}\right) \quad (2.22)$$

In 2.21 $s(n)$ is the original signal at sample point $n$, $i$ is the imaginary number ($i^2 = -1$) and $S(k)$ is the $k^{th}$ spectral coefficient. The index $k$ uniquely determines the frequency associated with the spectral component $S(k)$. Index $k$ can be transformed to a frequency $f$ by:

$$f = \frac{k}{T_0} \quad (2.23)$$

where $T_0$ is the segment time.

The resolution $\Delta f$ and the minimum frequency $f_{\text{min}}$ in the frequency domain are also determined by the segment time $T_0$:

$$f_{\text{min}} = \Delta f = \frac{1}{T_0} \quad (2.24)$$

The Nyquist frequency $f_N$, which is the maximum frequency that can be correctly represented in the sampled signal, depends on the sample frequency $f_s$:

$$f_N = \frac{f_s}{2} \quad (2.25)$$

Discrete Fourier transform algorithms assume that the data contain only frequencies that are an integer multiple of the minimum frequency. However, this is not the case with EEG data and this leads to spectral leakage; loss of power of a given frequency to other frequencies (Fig. 2.5).

The effect of spectral leakage can be diminished by multiplying the original signal by a window function before calculating the DFT. The Hanning window and Hamming window are often used (Fig. 2.6). They are defined by:

**Hanning:**

$$w(n) = 0.5 \left(1 - \cos \frac{2\pi n}{N - 1}\right), \quad n = 0, ..., N - 1 \quad (2.26)$$
Figure 2.5: Frequency spectrum of a signal consisting of a sinusoid with a frequency (3 Hz) that is an integer multiple of the minimum frequency (0.1 Hz) and a sinusoid of which the frequency is not an integer multiple of the minimum frequency (6.05 Hz). The latter causes spectral leakage.

Figure 2.6: Hanning Window (solid line) and Hamming window (dashed line) with N=10.
Hamming:

\[ w(n) = 0.54 - 0.46 \cos \frac{2\pi n}{N-1}, \quad n = 0, \ldots, N-1 \]  

(2.27)

When using the FT for filtering, unwanted frequencies are removed by using high-pass, low-pass, band-pass or band-stop filters. A filter commonly used for EEG is the Butterworth filter. It is designed to have a frequency response that is mathematically as flat as possible in the pass band and goes to zero in the stopband. The transfer function, which describes the amplification of the frequency amplitude by the filter, of the low-pass Butterworth filter is defined by:

\[ H(\omega) = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_c}\right)^{2p}}} \]  

(2.28)

where \( \omega_c \) is the cut-off angular frequency, \( \omega \) is the angular frequency of the signal in radians per second (\( \omega = 2\pi f \)) and \( p \) is the order of the filter. A steeper slope of the filter, indicating sharper cut-off, is obtained by increasing the order of the filter (Fig. 2.7). Equation 2.28 shows that when \( |\omega| < \omega_c \) the response is about unity, while if \( |\omega| > \omega_c \) the response drops rapidly to zero. The transfer function can be easily modified into a high-pass filter or a band-pass filter.

If the DFT is used for spectral analysis of E(R)Ps, the DFT is estimated for each segment and subsequently the DFTs of the segments are averaged. In this way the random noise is reduced.

**Figure 2.7:** Frequency responses of a first, second, third and fourth order Butterworth filter. \( \omega_0 = 1 \).
2.3.2 Coherence

Coherence analysis has been applied to EEG data for many years as a tool for studying the phase consistency or synchrony between two signals recorded from different scalp regions. High coherence between scalp areas has been interpreted as evidence for anatomical connections (Fein et al., 1988) and functional coupling (Thatcher et al., 1986) between the cortical structures underlying these areas. To study dynamic coupling between different brain areas in relation to specific motor, sensory or cognitive events, event-related coherence can be used (Andrew, 1996). Mathematically, coherence is analogous to a cross-correlation coefficient in the frequency domain. To calculate (event-related) coherence, the Fourier transform is calculated for each segment of the two channels. Secondly, the auto-spectra of channel X ($\hat{G}_{xx}$) and channel Y ($\hat{G}_{yy}$) and the cross-spectra of both channels ($\hat{G}_{xy}$) are determined as follows (Andrew, 1996):

\[
\hat{G}_{xx}(k) = \frac{1}{N} \sum_{n=1}^{N} |X_n(k)|^2
\]

\[
\hat{G}_{yy}(k) = \frac{1}{N} \sum_{n=1}^{N} |Y_n(k)|^2
\]

\[
\hat{G}_{xy}(k) = \frac{1}{N} \sum_{n=1}^{N} X_n(k)Y_n(k)
\]

where $N$ is the number of trials. $X_n(k)$ is the frequency amplitude of trial $n$ measured at channel $X$ for index $k$. Equation 2.23 describes the relation between index $k$ and the corresponding frequency. Next the coherence can be calculated as
Figure 2.9: Coherence between channel Pz and P4 (solid line) and between Pz and F4 (dashed line) of an oddball task recording. The coherence between Pz and P4 is considerably higher due to volume conduction effects: channel P4 lies much closer to Pz than F4.

the normalization of the cross-spectrum by the two auto-spectra:

\[
C^2_{xy}(k) = \frac{|\hat{G}_{xy}(k)|^2}{\hat{G}_{xx}(k)\hat{G}_{yy}(k)}
\]  

(2.32)

Or alternatively, a band-averaged coherence can be calculated according to (Andrew, 1996):

\[
C^2_{xy}(\bar{k}) = \frac{\left|\sum_{k=k_1}^{k_2} \hat{G}_{xy}(k)\right|^2}{\sum_{k=k_1}^{k_2} \hat{G}_{xx}(k) \sum_{k=k_1}^{k_2} \hat{G}_{yy}(k)}
\]

(2.33)

with the limits of summation \((k_1, k_2)\) related to the frequency band of interest \((f_1, f_2)\) and the segment time by equation 2.23 and \(\bar{k}\) the average of all indices \(k\). Additionally, it is possible to obtain a time course of coherence by shifting the segments over small periods from the start to the end of the trial (Pfurtscheller and Andrew, 1999). The latter technique can only be applied if the used segment length is much smaller than the trial length.

In coherence analysis the choice of reference is very important. Data with similar reference electrodes can increase the coherence compared to reference independent methods (Andrew, 1996). Reference methods that improve the coherence estimations are the closely spaced bipolar reference, the linked ear reference, the average reference, and the more complex surface Laplacian (see section 2.2.1) and cortical imaging (Nunez, 1997).

Alternative methods for coherence analysis are covariance analysis (see section...
2.2.3, phase locking statistics (Lachaux et al., 1999) and a method introduced by Nikolaev et al. (2001). The latter method first calculates the wavelet transform of averaged EPs and then computes the correlation coefficient between the wavelet transform curves of pairs of channels. Furthermore, Gross et al. (2001) designed a technique called DICS (Dynamic Imaging of Coherent Sources). DICS uses a spatial filter, which allows computation of coherence with respect to a reference point at any given location in the brain.

### 2.3.3 Wavelet transform

Similar to the FT, the wavelet transform (WT) is a method to analyze the frequency content of a signal. In addition, wavelet transforms can be used for filtering (Quian et al., 2001).

The main difference with FT is that wavelets are localized in both time and frequency, whereas the base function of the FT is only localized in frequency. Therefore, WT is used in EP and ERP studies to investigate changes in frequency bands due to stimulation. Başar et al. (1979b; 1979a) suggested that E(R)Ps appear as a reorganization of the spontaneous brain oscillations in response to a stimulus. Instead of sine and cosine functions, the wavelet transform uses wavelet functions to decompose the signal. In the continuous wavelet transform (CWT) method these "wavelets" are obtained from a mother wavelet by scaling:

\[
\varphi_{a,b}(x) = \frac{1}{\sqrt{|a|}} \varphi \left( \frac{x - b}{a} \right)
\]

where \(a\) is the dilation parameter, which governs the frequency of the wavelet and \(b\) is the position parameter, which governs the time displacement. However, the shape of the wavelets is always the same as that of the original mother wavelet \(\varphi\). For \(\varphi\) different wavelets may be chosen. Examples of mother wavelets are the Morlet wavelet and the Mexican hat wavelet (Fig. 2.10). The Morlet wavelet is defined by:

\[
\varphi = e^{i\omega_0 x} e^{-\frac{1}{2} x^2}
\]

The Mexican hat wavelet is defined by:

\[
\varphi = (1 - x^2) e^{-\frac{1}{2} x^2}
\]

The wavelet transform calculates the correlation between the wavelet and the original signal, i.e.:

\[
W_\varphi X(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} X(t) \varphi^* \left( \frac{t - b}{a} \right) dt
\]
Figure 2.10: Mexican hat wavelet (left) and real part of Morlet wavelet (right).

Figure 2.11: Time-frequency chart obtained with the continuous wavelet transform of a VEP recording at channel Oz. After 100 ms the frequency components of the P100 can be observed.

where $\varphi^*$ is the complex conjugate of the mother wavelet. As a result, the correlation coefficient is obtained as a function of frequency and time.

For discrete signals the discrete wavelet transform (DWT) can be used and like for the discrete Fourier transform, a fast operation, the fast wavelet transform (FWT) has been developed to compute the DWT. In case of the DWT, time frequency information is obtained by digital filtering techniques (Quian et al., 2001). The DWT is computed by low pass and high pass filtering of the discrete time domain signal. This results in two signals; the high pass signal contains frequencies between half the Nyquist frequency and the Nyquist frequency, the low pass signal consists of all frequencies lower than the Nyquist frequency. Next, the resulting high-pass and low-pass signals are downsampled by 2. This procedure is then repeated many times for the lowpass filter signal to further increase the frequency resolution. The signals that are high pass filtered are the outcome of the DWT.
2.3.4 ERD/ERS

When averaging single trials, one assumes that the evoked activity or signal of interest has a fixed time-delay to the stimulus, while the ongoing EEG activity behaves as additive noise. However, due to the stimulus, the ongoing EEG signal may change. The power in a specific frequency band can either decrease or increase. A decrease is called event-related desynchronization (ERD) (Pfurtscheller, 1977), since it is caused by a decrease in synchrony of the underlying neuronal populations. An increase is called event-related synchronization (ERS) (Pfurtscheller, 1992), since it is assumed to be caused by an increase in synchronization. These changes are time-locked, but not phase-locked, i.e. the power of some frequency bands of the ongoing EEG changes at a fixed time delay, but the phase of the ongoing EEG at that fixed time delay varies over the different trials. When single trials are averaged, this change in the ongoing EEG is no longer detectable. A prerequisite of calculating ERD or ERS is that the interval between two consecutive events should last at least 10 seconds, because event-related changes in ongoing EEG need time to develop and to recover (Pfurtscheller and Andrew, 1999).

ERD and ERS can be calculated using the following steps (see also Fig. 2.12):

1. band pass filtering of all event-related trials
2. calculating the power by squaring the amplitude
3. averaging the power across all trials
4. averaging of time samples to smooth the data and reduce the variability
5. calculating the power relative to a reference period. The reference period is a period of a few seconds before the event occurs.

A disadvantage of this technique is that phase locked (e.g. evoked potentials) cannot be distinguished from non-phase locked EEG, when both types of activity are in the same frequency band. To calculate only the non-phase locked power, the second step can be replaced by squaring the intertrial variance, which is the difference between the single trial and the averaged signal over all trials (Kalcher and Pfurtscheller, 1995). A limitation of this technique is that it does not account for latency and amplitude variability of the stimulus locked component within individual trials. As a result, this technique might remove too little or too much signal. Therefore McFarland et al. (2004) introduced the regression subtraction procedure. In this procedure the average that is subtracted was first corrected for latency and amplitude using the covariance between the averaged signal and the single trials.

Since the introduction of ERD and ERS, various alternative methods have been reported. For example, Makeig (1993) introduced a method called event-related spectral perturbation (ERSP). The advantage of this method is that it calculates