Time Frequency Analysis of EEG Measured When Performing the Flanker Task

Johan Brynolfsson

June 15, 2012

Abstract

This thesis handles time frequency analysis of EEG signals measured on participants performing the so-called flanker task. The analysis is done mainly using multitapering techniques on the quadratic class. Using multiple orthonormal windows when estimating the spectra of a process, one lowers the variance of estimate.

A class of locally stationary processes (LSP) is presented to use as a model of EEG which can then be used to evaluate the different time-frequency methods that are presented. This LSP contains only one component is used to model only one part of the EEG signal. When analyzing the set of EEG signals of this thesis one is most interested in the so-called N2 event and the model is therefore applied to this event. Having this model one can then find the optimal multitapers in the mean square error sense.

Different sets of multitapers are used to analyze the time-frequency representation of the EEG-signals. These are evaluated on LSPs where the true spectra are known.

Spectra are then estimated for the EEG-signals. As there are multiple channels and different methods are used, only a selected set of these spectra are presented here.

Acknowledgments

I would primarily like to thank Maria Sandsten who has supervised and guided me through this thesis. Many thanks also go out to Gerd Waldhauser, with the Department of Psychology at the University of Konstanz, Konstanz, Germany, who has collected the data, and his PhD supervisor Mikael Johansson with the Department of Psychology at the University of Lund.

Contents

1 Introduction

1.1 Purpose

The main focus of this thesis is to investigate and implement different time-frequency analysis methods and applying them to grand averaged EEG signals. Multitaper methods are used to estimate the time-frequency representation. As the tapers are orthonormal they generate uncorrelated estimations and one therefore lowers the variance of the estimate. The purpose is to evaluate the results of different windows.

1.2 EEG

The brain is a very complex structure that we still today lack the full understanding of and we are still locking into good ways to interpret the signals from the brain. The brain is built up of neurons that communicate via electric impulses; measuring these individual impulses would give us a quite clear picture of what is going on, but of course this is (today) impossible. However when millions of neurons communicate at the same time they give rise to an electric field that is even measurable from the scalp. By applying multiple electrodes around the scalp, one can measure the activity at different locations of the brain which will give a rough picture of what parts of the brain that are momentarily more or less active. This is called the EEG signal, or Electroencephalogram. Generally the signals are rhythmic and one categorizes them depending on at what frequency they are oscillating and their relative amplitude. Theses classifications can be found in table 1.

The EEG is a cheap and non-invasive method that renders a high time-resolution, which makes it a popular method of measuring the brain activity. However, as relatively few electrodes can be fitted onto the scalp one gets a quite poor spatial resolution. Two adjacent electrode will even record much of the same information. Additionally, one can not tell from what depth of the brain the signal is coming from.[11]

Rhythm	Frequencies	A coarse explanation of when rhythms occur.
Delta rhythm	<4 Hz	Typically occurs during sleep and is generally correlated with a high amplitude.
Theta rhythm	4-8 Hz	Occurs when a person is drowsy, going to sleep or waking up.
Alpha rhythm	8-13 Hz	When a person is relaxed or reflecting, the alpha rhythm arises and is most prominent at the posterior parts of the scalp.
Beta rhythm	13-30 Hz	Occurs when a person is awake and alert and is correlated with a low amplitude. Generally observed in frontal and central regions of the scalp.
Gamma rhythm	>30 Hz	This rhythm is observed when a person is processing information from the cortex which is involved in the more complex functions of the brain.

Table 1: Rhythm classifications

1.2.1 Evoked potentials, grand averages and difference curves.

The EEG signal is due to its poor spatial resolution very noisy and hard to interpret. A popular method to increase the signal-to-noise ratio is to use evoked potentials. The idea is that one performs the same task repeatedly, with a waiting time between every repetition. Assuming that the brain responds in the same way each time one can form the average which will then reduce the noise levels greatly. The waiting time between each task performance can be varied to make the test less predictable for the participants. This will make the experiment more robust as patterns become less prominent and therefore the experiments effects in the brain become more clear. The grand average is calculated as a point-wise mean of all signals. Assume K evoked potentials, $X_1(n), X_2(n) \dots X_N(n)$, are measured at N points. The grand average signal, X_{GA}, is then found as

$$X_{GA}(n) = \frac{1}{K} \sum_{k=1}^{K} X_k(n) \qquad for\ n = 0,\ 1,\ \dots,\ N-1.$$

In the test analyzed in this thesis, one is interested in the difference in brain responses triggered by two different tasks. To compare how the responses differ, one may look at two different methods. The first method is to perform frequency analysis on the two signals and compare the difference in frequency responses. The second is to compute the difference curve, i.e. compute the difference between the grand average responses of the two tasks. This method retains the phase of the two signals, which also may contain valuable information.

1.3 The flanker task

The experiment, that was performed by Gerd Waldhauser, PhD, who then was with the Department of Psychology at Lund University, is a version of the so-called flanker task [2]. Participants are asked to identify the *middle letter* of a five letter string, shown on a screen. Only the letters H or S are used and there are four different strings available. Either the string is congruent, i.e. all the five letters are the same (HHHHH or SSSSS), or the string is incongruent, i.e. the *middle* letter is different from the surrounding letters (HHSHH or SSHSS). Subjects identify the middle letter by pushing an H or S button. The letters may also be shown on the left, middle or right part of the screen. The point of interest in this experiment is the so-called $N2$ response, or second brain response. This usually occurs somewhere between 200 and 500 ms after the stimulus. The hypothesis is that there will be a difference in the brain responses at $N2$ on the posterior part at the opposite side of where the strings are shown. In other words, if a participant is shown one congruent and one incongruent string on the *right* side of the screen, a difference in brain response is expected to be seen on the posterior *left* side.

There were 27 participants in the experiment but the results where only used from 22 of them. The EEG signals are measured using 38 electrodes, where 36 electrodes measures EEG and two measures the eye movement. A rough image of the electrode setup can be seen in figure 1.1.

The signals were sampled at 500 Hz.

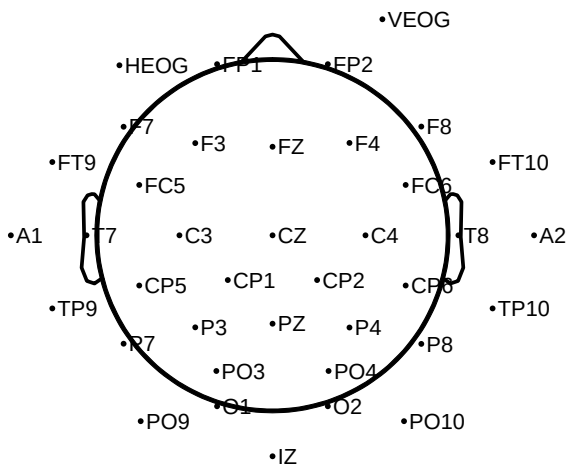


Figure 1.1: Map of locations of the electrodes placed on the scalp. The 39:th electrode is just a reference electrode.

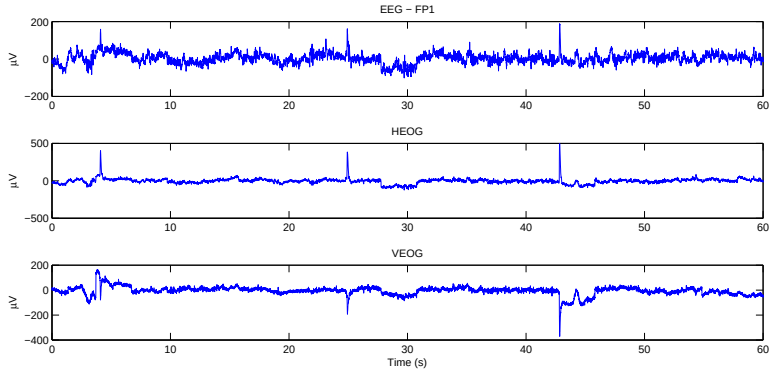


Figure 2.1: The eye-movement is shown in the two lower plots. Clear blinking- and eye-movement artifacts are seen in the EEG signal (top plot).

2 Theoretical Concepts and methods

2.1 Removing eye movement artifacts

One of the more prominent artifacts in the EEG signals derives from eye movements; both from moving the eyes themselves and from blinking. To remove this information, which is unwanted, from the EEG signal, one usually measures both of these kinds of eye movements using two additional electrodes. The first, called HEOG (Horizontal ElectroOculuGram), measures the horizontal eye movement and the second, VEOG, measures the vertical eye movement. An example of how these eye movement artifacts affect the EEG can be seen in figure 2.1. The data is taken from the tutorial of the EEG analysis software eeglab [10] and shows the EEG from a frontocentral electrode of a subject performing a visual attention experiment. Hence there are quite a lot of eye movements and therefore artifacts in the EEG.

There are many ways to remove these eye movement artifacts from the EEG. The method chosen here is a recursive least square filter proposed by He, P. et al. [4].

An example of the result of the algorithm can be seen in figure 2.2. The peaks correlated to blinking are gone but one can also see that the power of the EEG signal is slightly lowered due to the noise correlation in all three signals.

The EEG signals acquired from the experiment in this thesis however, are already EOG-filtered using standard methods, but there still seems to be some correlation between frontal EEG and EOG-signals. Running the proposed recursive least square filter on these signals removes some baseline wander that is seen simultaneously in EEG and EOG and I therefore chose to filter all EEG signals.

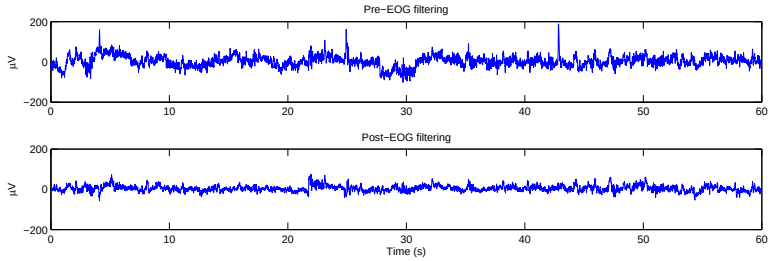


Figure 2.2: The prominent blinking artifacts that are seen in the top plot are removed after EOG-filtering is done.

2.2 Locally Stationary Process

To evaluate methods and investigate how they perform when analyzing EEG signals, a model has to be made. One can then simulate signals from this model and test the performances of the different methods.

One proposition is that EEG can be modeled as a locally stationary process (LSP). There are a few different definitions of a locally stationary process but the one used in this thesis is the one defined by Silverman in 1957 [9]. It may not be a very accurate model but a decent model to test the performance of different analysis methods.

Definition 1. A zero-mean random process $x(t)$ is called a **Locally Stationary Process** (in the wide sense) if its covariance function can be written on the form

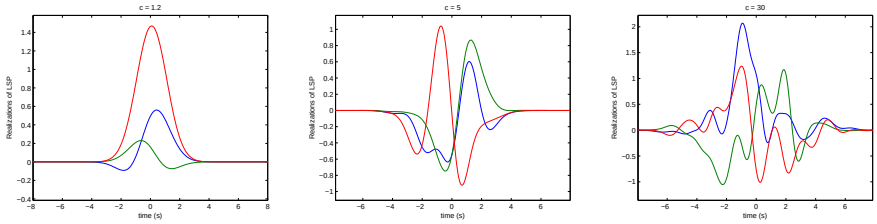
$$r_x(s,t) = q\left(\frac{t+s}{2}\right) \cdot r\left(t-s\right). \tag{2.1}$$

The function $r(\tau)$ must fulfill the criteria of a covariance function and $q(\tau)$ may be any function. In the model used in this thesis, the functions q and r are chosen as

$$\begin{cases} q(\tau) = a \cdot e^{-\tau^2/2}, \\ r(\tau) = e^{-c\tau^2/8}. \end{cases} \tag{2.2}$$

The variable c reflects how stationary the process is. As $c \to \infty$, the process becomes a stationary process. A restriction on c is that $c \geq 1$, for $r(\tau)$ to be a covariance function. The parameter a is the amplitude of the covariance function. In the simulations this will be chosen as $a = 1$. Some realizations of LSPs, with different values of c, can be seen in figure 2.3.

Note that the variables in eq. 2.1 are in t and s, two time variables. One may want the covariance expressed in the variables t and τ, time and time-lag. This is achieved by rotating the coordinate system 45 degrees which is done using the coordinate change



(a) (b) (c)

Figure 2.3: Realizations of LSPs for different values of c

$$\begin{cases} t &= \frac{t+s}{\sqrt{2}} \\ \tau &= \frac{t-s}{\sqrt{2}} \end{cases}.$$

This gives the LSP covariance function to be

$$r_x(t,\tau) = q\left(\frac{t}{\sqrt{2}}\right) \cdot r\left(\tau\sqrt{2}\right).$$

One downside with this model is that all frequencies are concentrated around origo. This is a problem when modeling EEG, as components of the real signal may be concentrated around a frequency, f_0, separated from $0\,Hz$. Assuming this frequency f_0 is constant, one can model this as an oscillation of the signal. The real EEG signal may also have its center of mass at time 0. This however is easily fixed by estimating the time mass center and shifting the time-scale.

2.2.1 Estimating parameters

If one wants to model the EEG signals as LSP one needs to know the parameters in the model. There are 6 parameters that need to be estimated;

- c - The parameter reflecting how stationary the signal is.
- $F1$ - Scaling frequency. This is used to scale the windows that eventually are used.
- f_0 - The oscillating frequency. This must be done as the LSP only contain frequencies close to 0. If one wants to model a signal with frequencies with a frequency mass center, this will be modeled as a sinusoidal signal multiplied with the LSP.
- a - The amplitude of the covariance function.
- σ^2 - Assuming the signal is observed under white noise, one will also want to estimate the variance of this noise. Assuming only white noise however is a very strong assumption which may not really be accurate; especially when analyzing EEG signals where the sources of noise are multiple.

When fitting a model, the estimation of these parameters will be performed in four main steps.

1. First one must estimate the covariance matrix of the signal. To get a good estimation, one needs multiple realizations, especially if the signal is observed under noise. Then find the center of mass of the covariance function. Assuming the signal is a LSP, due to symmetry it will be sufficient to find the center of mass of the diagonal of the covariance matrix. This makes computations easier since the problem then is 2-dimensional. Lastly move the process in time so that the center of mass is at time 0.

2. Estimation of the Wigner-Ville distribution of the Hilbert-transformed signal. This will return an image of how frequencies change over time. As one expects the signal to have one main component, cross-terms are not expected to be a problem (see section 2.4.4 for more on this). The most powerful frequency will be taken as an initial estimation of the oscillating frequency f_0.

3. Now assume that the signals has a covariance that can be written on the form of eq. 2.1 & 2.2. On the diagonal $t = s$, the $r(t-s)$ term will then be identical to 1 and therefore one will only observe the $q(\tau)$-term. Assuming the noise is white this diagonal will have the base at σ^2. The estimations of $F1$, σ^2 and a will then be the numerical least square optimal fit to the the theoretical curve of $q(\tau)$ as in eq. 2.2. An analytic expression that could be minimized would be preferable but in this thesis the numerical solution given by the built-in function `fminsearch` in `MATLAB` will suffice. The σ^2 parameter will, if indeed the noise is white, be estimated quite accurately as it will only raise the diagonal base level from 0 to σ^2.

4. Lastly, look at the anti-diagonal $t = -s$. By the same arguments as in step 3, the $q(\frac{t+s}{2})$ will be identical to 1 and one therefore only observes the $r(\tau)$-term. `fminsearch` can then again be used to estimate the remaining parameter, c.

This multi-step estimation of the parameters is of course not optimal but will return a descent estimation, given that the model assumption is correct.

2.3 Frequency analysis

By assuming that the signal one wants to analyze is oscillating one can estimate with what frequency or frequencies this is occurring. There are two main group of estimation techniques, parametric and non-parametric models. In this thesis, only non-parametric models will be used. The result of a non-parametric technique will be a spectrum or power spectral density (PSD) .

2.3.1 Fourier-transform and spectrum

The most fundamental tool used in frequency analysis is the Fourier transform which transforms the signal into the frequency domain.

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-i2\pi ft}dt. \tag{2.3}$$

The Fourier transform is designated as

$$\begin{cases} X(f) &= \mathscr{F}\left(x(t)\right) \\ x(t) &= \mathscr{F}^{-1}\left(X(f)\right). \end{cases}$$

The spectrum $\phi(f)$ can then be found in two ways. Either as the squared absolute value of the Fourier transformed signal or from the Fourier transformed auto-covariance function, $r(\tau)$, of the signal.

$$\begin{aligned} S_x(f) &= \left|\int_{-\infty}^{\infty} x(t)e^{-i2\pi ft}dt\right|^2 \\ &= \int_{-\infty}^{\infty} r(\tau)e^{-i2\pi f\tau}d\tau. \end{aligned} \tag{2.4}$$

2.3.2 Periodogram

As one is usually studying real data which at some point has to be sampled, the discrete Fourier transform is of more interest. Assume one has N samples of the signal $x(n)$, where $0 \leq n \leq N-1$. The discrete Fourier transform is then found as

$$X(f) = \sum_{n=0}^{N-1} x(n)e^{-i2\pi fn}. \tag{2.5}$$

The periodogram was the one of the first methods to estimate periodicities, or frequencies, in a signal. It is simply the squared absolute value of the discrete Fourier transformed signal.

Definition 2. The periodogram is found as

$$\hat{S}_x(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n)e^{-i2\pi fn}\right|^2, \tag{2.6}$$

where N is the length of the signal which also determines the resolution of the frequency estimate.

If one wants higher resolution, ie. estimate the frequency on a larger grid, this can be achieved by so-called zero-padding. By adding zeros to the signal one does not add any information but the frequency grid is made larger. Another benefit of zero padding is that the discrete Fourier transform, that is implemented in `MATLAB`, is optimized in computational time if the signal has a length of a power of two. Therefore it is preferable to zero pad up to 2^k for some integer k.

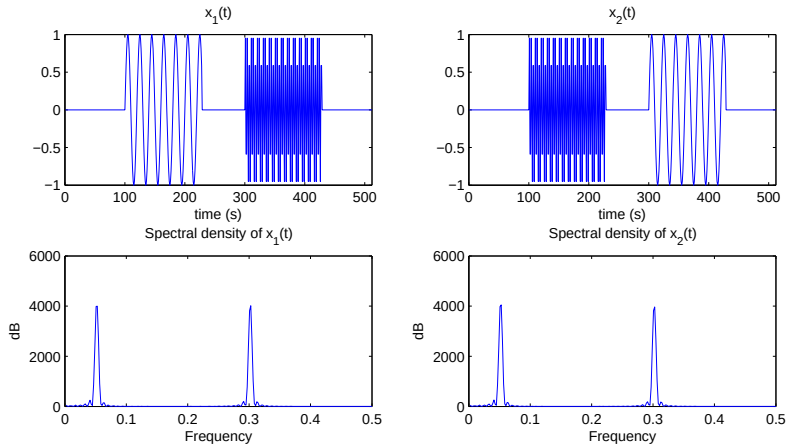


Figure 2.4: The signals $x_1(t)$ and $x_2(t)$ have the same two components, one faster and one slower sinusoidal signal. Even though they come in different order the two signals have the same spectral density.

S_x is a periodic function with the period of 1 in the normalized frequency, and one can therefore find all information in the interval $f \in [-\frac{1}{2}, \frac{1}{2}]$. For the periodogram it holds that $S_x(f) \geq 0$, which is reasonable as one estimates the density of frequencies; a quantity that can not be negative.

To use the periodogram one must assume the signal to be stationary, ie. that the signal has the same mean value throughout the signal and that the covariance function of the signal only depends on τ, the distance in time between two points. One then estimates what frequencies can be found in the signal. If the signal is non-stationary, and therefore changes over time, it will not be reflected in the periodogram. An example of this is seen in figure 2.4.

2.3.3 Windowed periodogram

One big problem with the periodogram is the so called spectral leakage. An example of this phenomenon can be seen in figure 2.5 where the periodogram is estimated for a pure sinusoidal signal with a normalized frequency of 0.25 Hz. The expected result is then a spike at frequency 0.25 Hz and 0 otherwise but as can be seen in the figure, there are multiple spikes propagating from the main peak.

Another problem with the periodogram is the large variance of the estimates. Both these problem, the bias in the form of spectral leakage and the variance, may be reduced by introducing the windowed periodogram,

$$\hat{S_x}(f) = \left| \sum_{n=0}^{N-1} w(n)x(n)e^{-i2\pi fn} \right|^2 , \tag{2.7}$$

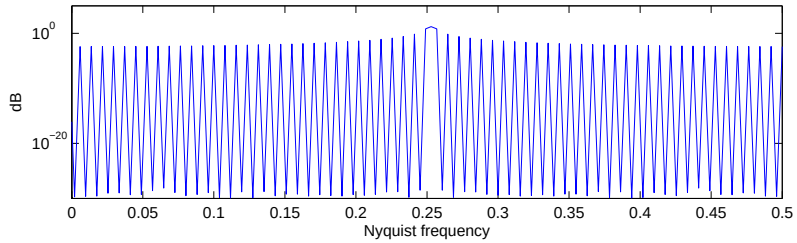


Figure 2.5: The power spectral density of sinusoidal signal with frequency 0.25 Hz. One can see the main lobe at this frequency but in this log scale one can also see the heavy spectral leakage.

where $w(n)$ is a so-called window or taper. A desirable window has a spectrum with a narrow main lobe and low power side lobes. The main lobe needs to be narrow for good resolution of peaks. For example, if there are two adjacent frequencies current in a signal, a broad main lobe may cause them to be perceived as one single component.

There are many different windows proposed that may be advantageous in different applications. Some of the more common ones are presented in figure 2.6 together with their spectra.

2.3.4 Cross spectrum and Coherence

When analyzing two signals and want to know how similar they are and how they are related, the cross spectrum and the coherence are two valuable tools.

Definition 3. The cross spectrum is a spectrum showing simultaneous frequencies in the signals $x(t)$ and $y(t)$ and can be found as

$$S_{xy}(f) = X(f)^* Y(f), \tag{2.8}$$

where $X(f) = \mathscr{F}(x(t))$ and $Y(f) = \mathscr{F}(y(t))$.

Note that, where the auto spectrum is always real valued as $X(f)^* X(f) = |X(f)|^2$, this is not the case of the cross spectrum. When analyzing the cross spectrum one may therefore look at the absolute value of the cross spectrum.

The coherence C_{xy} is the the same as the Cross spectrum but normalized with the spectrums of the two signals. The Cauchy-Schwartz inequality then gives coherence the property that $0 \leq C_{xy}(f) \leq 1$.

Definition 4. The Magnitude squared Coherence is defined as

$$C_{xy}^2(f) = \frac{|S_{xy}(f)|^2}{S_x(f) S_y(f)} \tag{2.9}$$

where $S_{xy}(f)$ is the cross spectrum of the two signals $x(t)$ and $y(t)$. $S_x(f)$ and $S_y(f)$ are their respective auto-spectrums.

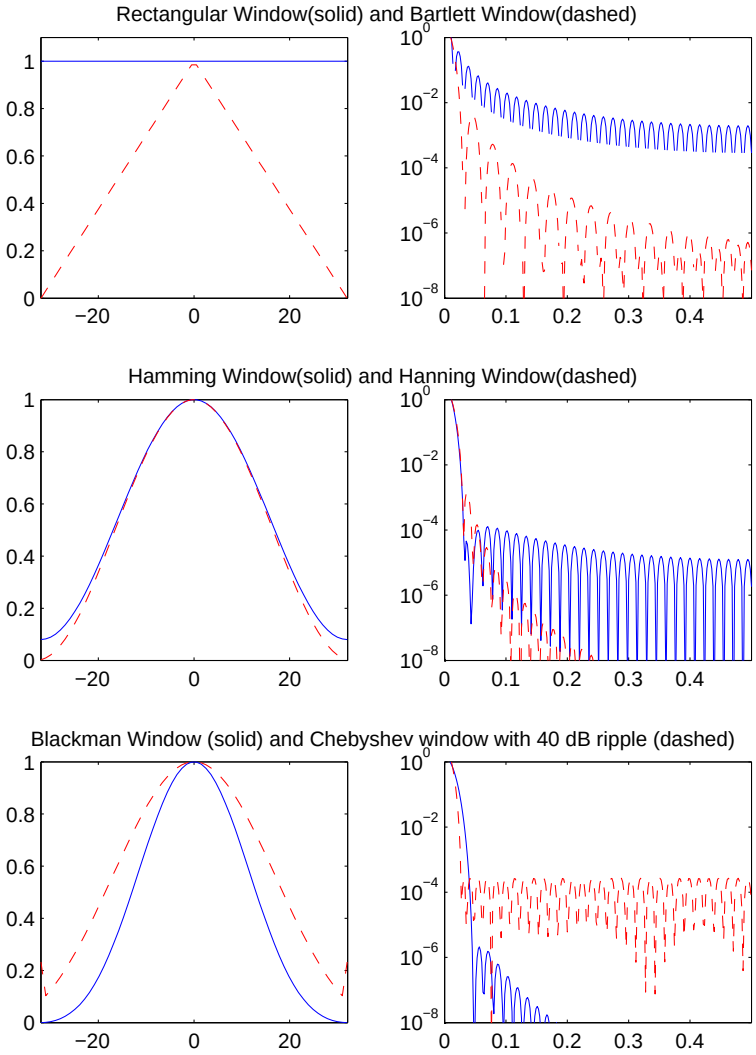


Figure 2.6: Six commonly used windows in the right column plots with their respective spectrums in the left column. As mentioned, a good window has a narrow spectral main lobe and low power side lobes.

Whereas the cross spectrum shows where both signals have high power the coherence will show how frequencies are related between two signals. If a frequency has the same power in the two analyzed signals, this frequency will have a coherence of 1, regardless of if the power is high or low.

One can see an example of coherence in figure 2.7. The left column represents three different filters. By passing a white noise signal through one of these filters and then computing the coherence between the original white noise and the filtered noise one gets an idea of how the filter works. The coherence is estimated using the Thomson multitapers (see section 2.5.3).

2.3.5 Dual frequency coherence

The dual frequency coherence is closely related to the cross spectrum but will show more clearly how two frequencies in two signals are correlated in amplitude and phase. As the cross spectrum only gives the absolute amplitudes it will be hard to detect correlations in less powerful frequencies. Therefore it is a good idea to normalize the cross spectrum with the spectra for each signal, and hence estimate the coherence instead. This method was originally presented by Mellors, R.J. et al. [7].

To estimate the dual-frequency coherence, the Thomson spectral estimates will be used here (see section 2.5.3). When a bandwidth NW is chosen I pick the $K = 2NW - 1$, first windows to estimate the spectrum. As the eigenvalues of the last few windows may differ a bit from one, each window will be weighted with its eigenvalue.

Begin by calculating the discrete Fourier transform of each windowed signal

$$y_k(f) = \sum_{n=0}^{N-1} h_k(n)x(n)e^{-i2\pi fn}, \tag{2.10}$$

where h_k is the k:th Thomson window.

Then estimate the cross-spectra between each frequency of the two signals y^i and y^j are defined as

$$\hat{S}_{ij}(f_1, f_2) = \frac{W}{K} \sum_{k=1}^{K} \lambda_k y_k^i(f_1)^* y_k^j(f_2), \tag{2.11}$$

where λ_k is the eigenvalue corresponding to the $k : th$ window and W is the sum of inverted eigenvalues to normalize the effect of the weights.

$$W = \sum_{k=1}^{K} \frac{1}{\lambda_k}. \tag{2.12}$$

Calculate the dual frequency coherence by normalizing with the cross-spectra for the two signals. The cross-spectra is calculated as

$$\hat{S}_{ij}(f) = \frac{W}{K} \sum_{k=1}^{K} \lambda_k y_k^i(f)^* y_k^j(f), \tag{2.13}$$

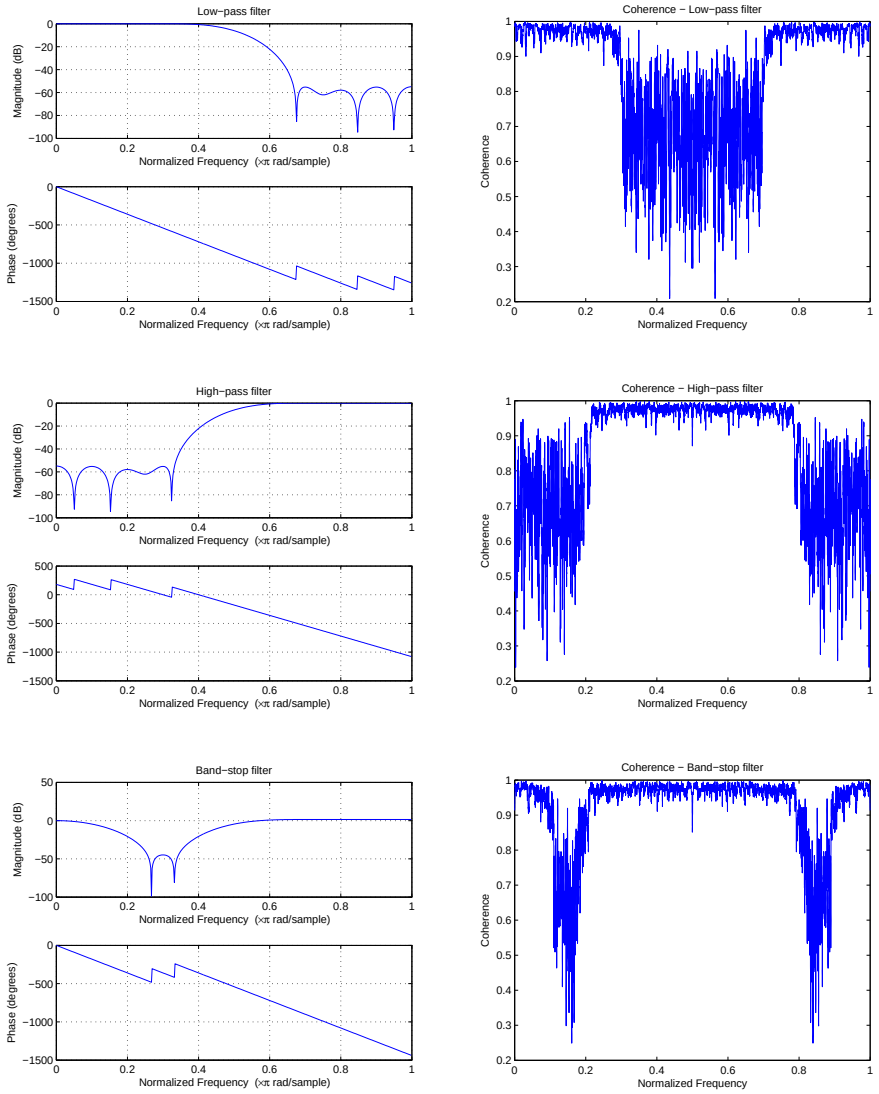


Figure 2.7: White noise signals are filtered through 3 filters, one low-pass, one high-pass and one band-stop filter. The coherence is then calculated between the original signals and the one passed through a filter. One can see that the coherence is closely related to the corresponding filter magnitude.

i.e. the same as the cross-spectra between frequencies as in eq. 2.11, but here the cross-spectra is only calculated between each frequency and it self.

The cross-spectra may be complex valued and one therefore looks at the squared absolute values of the cross-spectra.

$$\gamma_{ij}(f_1, f_2) = \sqrt{\frac{|Sij(f_1, f_2)|^2}{S_{ii}(f_1)S_{jj}(f_2)}} \tag{2.14}$$

The normalization with the cross-spectra for the two signals will show us correlations between frequencies even if they are of low power which may be advantageous. However if the power of a frequency is approaching zero, high values will be seen in the dual frequency coherency plot even though the frequency is non-existing. One may therefore want to add white noise to a signal so that there are some power at each frequency to avoid zero divisions.

This method has proven valuable when one wants to find dispersive frequencies that are coherent in time. For example, the method is very effective when looking at two chirp signals with different accelerations.

$$\begin{aligned} x(t) &= 100sin(2\pi t^2 \cdot 0.075/600) + e_1(t) \quad t \in [0, 600], \\ y(t) &= 100sin(2\pi t^2 \cdot 0.050/600) + e_2(t) \quad t \in [0, 600], \end{aligned}$$

where $e_i(t)$ is white noise with standard deviation $\sigma = 0.1$. The dual-frequency coherence can be seen in figure 2.8, where the frequencies of the two signals are on the two axis. An off-centered line can be seen for the frequencies that are coherent. As the SNR is very high, the noise makes no apparent effect on the frequencies represented in the signal. However; as soon as the coherence is estimated for frequencies that are non-existing in the signal, the estimate is much more noisy.

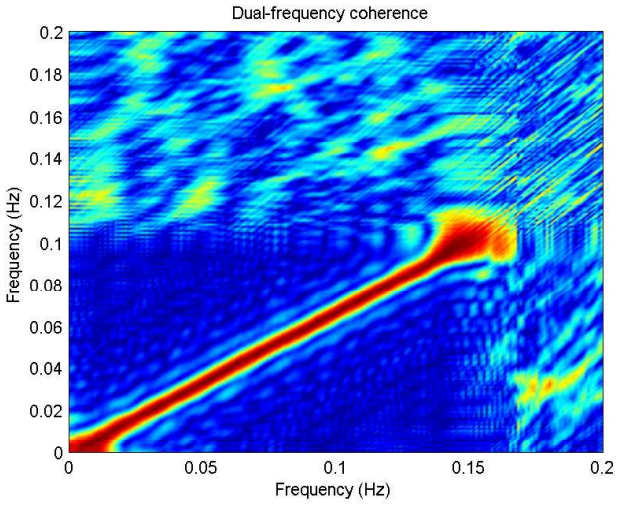


Figure 2.8: Dual-frequency coherence for two chirp signals with different acceleration. The off-center line marks the frequencies that are coherent between the two signals. As one chirp accelerates faster than the other the line has an angle lower than $\pi/4$.

An additional approach may be to divide the signal into short time sequences and calculate the dual-frequency coherence for each section. This however adds a third dimension which makes the result hard to present and interpret.

2.4 Time-Frequency analysis

When the signal one wants to analyze is non-stationary the spectrum is of little use. One therefore have to add a time-dimension and estimate the frequencies in the signal at each time.

2.4.1 Spectrogram

The Spectrogram is used to estimate frequencies in a signal that is non-stationary by introducing a time dimension in the frequency analysis. The methods divides the signal into short time segments and assumes that the signal is stationary within each segment. One then computes the periodogram for each segment. The Fourier transform of short time segments is called the Short Time Fourier Transform (STFT). This can be found as

$$S_x(t,f) = \int_{-\infty}^{\infty} x(t_1)h^*(t_1 - t)e^{-i2\pi f t_1} dt_1, \tag{2.15}$$

where $h(t)$ is a window in time. Similarly to the periodogram case, one then finds the spectrogram as $S_x(t,f) = |X(t,f)|^2$.

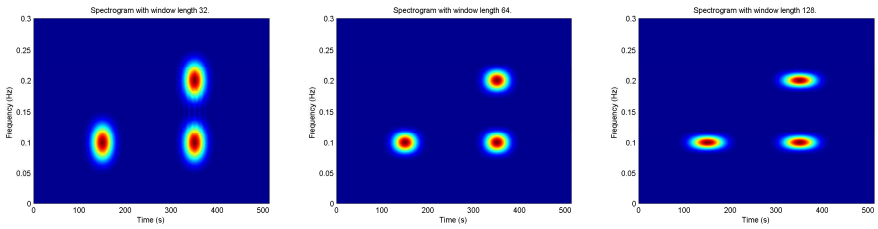


Figure 2.9: Spectrograms of a signal containing three Gaussian components. In the left most spectrogram a time window of length 32 has been used, window length 64 in the middle and in the right spectrogram a window of length 128. One can see that when a longer window is used a better resolution in frequency is received but at the same time a worse resolution in time.

As usual, the signals investigated are sampled and one therefore needs the discrete version of the STFT.

$$\hat{S}_x(t,f) = \left| \sum_{t_1=0}^{N-1} x(t_1) h^*(t_1 - t) e^{-i2\pi f t_1} \right|^2 . \tag{2.16}$$

The main problem with this method is the tradeoff in resolution between time and frequency. To achieve a good resolution in time one needs to choose a time window, $h(t)$, that is narrow. This however will return a poor frequency resolution. Vice versa, a long time window will give a good resolution in frequency and poor resolution in time (see figure 2.9). This is a problem, for example, if a signal containing two components with very similar frequency which may then be perceived as one single component.

2.4.2 The Quadratic class

The Quadratic class is a set of time-frequency estimators that can be written on the form

$$Q_z(t,f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A_z(\nu,\tau) \Phi(\nu,\tau) e^{i2\pi(\nu t - \tau f)} d\nu d\tau, \tag{2.17}$$

where A_z is the Ambiguity function (see section 2.4.5) of the analytic signal z and Φ is a kernel .

2.4.3 Analytic signals and the Hilbert transform

Definition 5. A signal is analytic if $X(f) = 0$ for $f < 0$ where $X = \mathscr{F}(x(t))$. In other words, the signal only contains positive frequencies.

When looking at real valued signals the spectral density of the negative frequencies is a mirror image of the spectral density of the positive frequencies,

$X(f) = X(-f)$. The analytic signal can then be found by taking the Hilbert transform of the signal.

Definition 6. The Hilbert transform of a signal is defined as

$$\mathscr{H}(x(t)) = \mathscr{F}^{-1}\left\{(-i \cdot \text{sign}(f))\,\mathscr{F}(x(t))\right\}, \tag{2.18}$$

where the sign function is defined as

$$\text{sign(f)} = \begin{cases} 1 & \text{if } f > 0 \\ 0 & \text{if } f = 0 \\ -1 & \text{if } f < 0 \end{cases}.$$

2.4.4 The Wigner-Ville distribution

As previously mentioned, the main problem with the STFT is the bad resolution. To get better resolution than the spectrogram one can use the Wigner-Ville distribution, which is a member of the quadratic class.

Definition 7. The Wigner-Ville distribution is defined as

$$W_z(t,f) = \int_{-\infty}^{\infty} z(t+\frac{\tau}{2})z^*(t-\frac{\tau}{2})e^{-i2\pi f\tau}d\tau \tag{2.19}$$

where $z(t)$ is the analytic signal acquired by Hilbert transforming the signal.

For a deterministic signal with a single component the Wigner distribution will give the instantaneous frequency. However there is one huge downside with the Wigner-Ville distribution. Assume that one has a two component signal, $z(t) = z_1(t) + z(t)$, and then calculate the Wigner-Ville distribution.

$$\begin{aligned} W_z(t,f) &= \int_{-\infty}^{\infty} \left(z_1(t+\frac{\tau}{2}) + z_2(t+\frac{\tau}{2})\right)\left(z_1(t-\frac{\tau}{2}) + z_2(t-\frac{\tau}{2})\right)^* e^{-i2\pi f\tau}d\tau \\ &= \int_{-\infty}^{\infty} z_1(t+\frac{\tau}{2})z_1^*(t-\frac{\tau}{2})e^{-i2\pi f\tau}d\tau + \int_{-\infty}^{\infty} z_2(t+\frac{\tau}{2})z_2^*(t-\frac{\tau}{2})e^{-i2\pi f\tau}d\tau \\ &\quad + \int_{-\infty}^{\infty} z_1(t+\frac{\tau}{2})z_2^*(t-\frac{\tau}{2})e^{-i2\pi f\tau}d\tau + \int_{-\infty}^{\infty} z_2(t+\frac{\tau}{2})z_1^*(t-\frac{\tau}{2})e^{-i2\pi f\tau}d\tau \\ &= W_{x_1}(t,f) + W_{x_2}(t,f) + 2\cdot\Re\left(W_{x_1,x_2}(t,f)\right). \end{aligned}$$

The first two terms in the last row are called the auto terms and the last term is called the cross term. This cross term will appear in the middle between every pair of components in the signal. The cross terms will also oscillate perpendicular to the line connecting the two true terms. A signal containing many components or a low SNR is then hard to estimate due to the high number of cross terms.

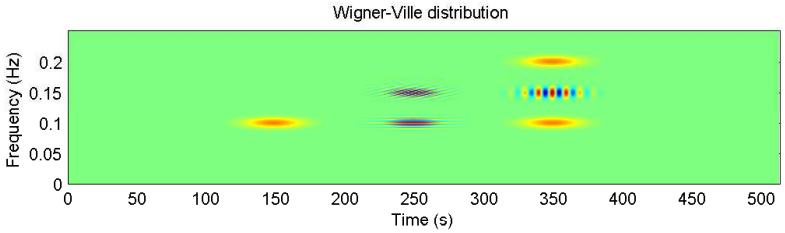


Figure 2.10: Wigner-Ville distribution of signal containing three Gaussian components $(t, f) = (150\,,\,0.1)$, $(t, f) = (350\,,\,0.1)$ and $(t, f) = (350\,,\,0.2)$. The oscillating cross terms can be seen in the middle between every pair of components.

There are a few advantages of using the analytic signal when estimating the time-frequency spectrum. By removing all negative frequencies one gets rid of all cross term that would otherwise emerge between every mirror image couple of $X(f) = X(-f)$.

When estimating the Wigner-Ville distribution of a discrete-time signal one can not get values that are shorter apart than the unit distance. To get a value of $z(t + \frac{\tau}{2})$ in eq. (2.19) one therefore has to down sample the signal by a factor 2. One can then only estimate frequencies in the range $f \in [-\frac{1}{4}\,,\,\frac{1}{4}]$. When estimating the frequencies higher than $\frac{1}{4}$ an aliasing effect will occur. For example, calculating the WVD for a signal containing the frequency $0.3\,Hz$, an estimation will appear at frequency $-0.2\,Hz$. But knowing the signal is analytic one can then sample up after the estimation and fold the negative frequencies back into the positive plane.

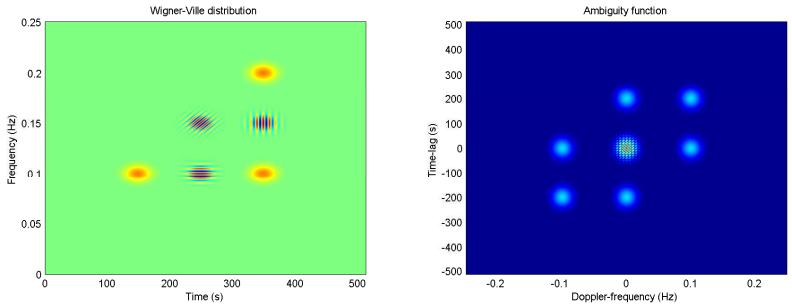
2.4.5 The ambiguity domain

The ambiguity domain is another member of the quadratic class, related to the Wigner-Ville distribution by a 2-dimensional Fourier transform, $\mathscr{F}^{-1}_{f\to\tau}$ and $\mathscr{F}_{t\to\nu}$. The variable ν can be interpreted as frequency-lag (called Doppler-frequency) and τ as time-lag. The ambiguity function can also be found by

$$A_z(\nu, \tau) = \int_{-\infty}^{\infty} z(t + \frac{\tau}{2}) z^*(t - \frac{\tau}{2}) e^{-i2\pi\nu t} dt. \tag{2.20}$$

The ambiguity function still suffers from cross terms, although these show up in a favorable way. As there is no time- or frequency lag between a component and itself in a signal, all true terms will end up together in origo. Cross terms however will end up peripherally. Utilizing this feature of the ambiguity domain has been shown to be a good way of reducing cross terms. By applying an ambiguity kernel that only preserves the components close to origo and then transform back the time-frequency domain one erases much of the cross terms.

When looking at real life signals however, they seldom contain true instantaneous sinusoids. There will then be some time and frequency lag within each component. These will then stretch outside origo. One must therefore find a



(a) (b)

Figure 2.11: (a) Cross-terms show up between every pair of components in the Wigner-Ville distribution
(b) In the ambiguity domain the cross terms show up outside origo. Each cross term from the Wigner-Ville distribution show up twice in the ambiguity domain. Once for positive lags and once for negative lags.

kernel which will reduce cross terms but still retain the signal. The size and shape of an optimal Ambiguity kernel will therefore differ between every signal. The kernel can either be multiplied to the Ambiguity function or applied by convolving the 2-dimenional Fourier transformed kernel in the time-frequency domain.

2.4.6 Some noteworthy distributions and kernels

One of the most commonly used kernels is the **Choi-Williams** kernel.

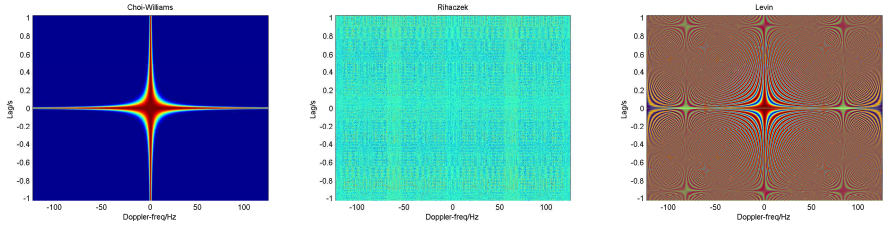
$$\Phi_{CW}(\nu,\tau) = e^{-\alpha(\nu\tau)^2}. \tag{2.21}$$

It has the feature to pass any cross terms which have only time-lag OR frequency lag. This means that sinusoids or Dirac-functions would still remain intact, something eg. a Gaussian kernel would filter out. The Choi-Williams kernel can be seen in figure 2.12a, where $\alpha = 1$.

Another important distribution is **the Rihaczek distribution** (RD) which is defined as

$$R_z(t,f) = z(t)Z^*(f)e^{-i2\pi ft}. \tag{2.22}$$

The distribution can be interpreted as the complex-valued energy density of the signal. RD do suffer from some cross term as well. Terms will show up whenever both the time representation and frequency representation differ from zero. Assume there are Gaussian components at $(t,f) = (10,1)$ and $(t,f) = (20,-1)$. RD will then both show these two components, but also two additional cross terms at $(t,f) = (10,-1)$ and $(t,f) = (20,1)$ as both signal itself and the frequency representation of the signal differ from zero at these points. The



(a) The Choi-Williams Kernel (b) The Rihaczek kernel (absolute value) (c) The Levin kernel

Figure 2.12: Three of the more commonly used ambiguity kernels.

Ambiguity kernel can be seen in figure 2.12b, but is hard to interpret visually as it is complex valued.

One large downside with RD is that it is complex valued. Something the **Levin** distribution takes care of by simply taking the real part of RD

$$L_z(t,f) = \mathrm{Re}\left\{z(t)Z^*(f)e^{i2\pi ft}\right\}. \tag{2.23}$$

The Levin kernel is seen in figure 2.12c.

2.4.7 Time-frequency coherence

The time-frequency coherence is analogous to the 2 dimensional case of frequency coherence. As the cross-spectra may be complex valued one usually looks at the Magnitude squared coherence.

$$C_{xy}^2(t,f) = \frac{|S_{xy}(t,f)|^2}{S_x(t,f)\cdot S_y(t,f)}. \tag{2.24}$$

There are many papers handling the case of time-frequency coherence analysis of EEG signals, among others [5, 14]. One may use any of the above mentioned methods to estimate the cross-spectra and auto-spectra in equation 2.24. Using STFT or Welch method are two possibilities. Using a member of quadratic class however may give better resolution in both time and frequency. An upside with STFT and Welch is that they ensure the spectra to always be positive; a property the quadratic class unfortunately lacks. Where negative values turn up, numerical errors may occur that causes the time-frequency coherence to attain values outside the range $0 \leq C_{xy}^2(t,f) \leq 1$. One must therefore confine the coherence estimate to the time-frequency region where $S_x(t,f) > 0$ and $S_y(t,f) > 0$.

2.5 Multitaper estimators

2.5.1 Welch's Method

Welch introduced a method to reduce variance and bias. Instead of estimating the power spectral density of the entire set of samples, one divides the signal into smaller subsets. The subsets may overlap and a common choice is that each subset overlaps 50% with the previous one. One can then estimate the PSD on each of the segments and then get the final estimation of the spectrum as the mean of all subsets. This can be viewed as a window, moving in time. One may also use windows when estimating the PSD for each segment to reduce variance even more.

Welch's method may also be used to estimate time-frequency spectras. This is done by choosing a window length N and then dividing this time-window into smaller sub-windows of length dN. One can then estimate the spectra of these sub-windows. The average of the sub-spectras are then used as the estimate of the spectrum of the larger time-window.

Welch's method is still commonly used when estimating spectra and I will therefore use it as a reference method.

2.5.2 Multitapering theory

The main idea of multitapers is that by using several periodograms and averaging over these, a reduction in variance is obtained. To obtain this reduction in variance, the periodograms have to be uncorrelated. In the Welch method the same window is used but these are applied on time-shifted versions of the data, which results in uncorrelated periodograms. In 1982, David Thomson wrote a paper on a method using the whole data sequence for all periodograms [12]. By using orthogonal windows the periodograms could still be uncorrelated. It has been shown that the Thomson method outperforms the Welch method in terms of leakage, resolution and variance for many different spectra. When producing multitaper estimates of a spectrum from N samples of a discrete-time random process $x(n)$ one averages over K spectrograms.

$$\hat{S}(f) = \sum_{k=1}^{K} \alpha_k \hat{S}_k(f), \tag{2.25}$$

where α_k is a weighting factor and

$$\hat{S}_k(f) = |\sum_{n=0}^{N-1} x(n) h_k(n) e^{-i2\pi fn}|^2 \tag{2.26}$$

Equation 2.26 is a windowed periodogram obtained by using the data window $\mathbf{h}_k = [h_k(0)...h_k(N-1)]^T$. The multitaper estimate in eq. 2.25 is then obtained by weighting over K periodograms.

If one considers the multitaper estimation method as a filtering procedure in a filter bank of FIR-filters, the impulse responses of the sub filters are $\mathbf{h}_k$

and the corresponding frequency functions are $H_k(f) = \mathbf{h}_k^T \cdot a(f)$ where a is the discrete Fourier vector. Given the input signal $x(n)$, the power of the output signal is

$$\begin{aligned} P_B &= \int_{-W}^{W} |H_k(f)|^2 S(f) df \\ &= \mathbf{h}_k^T \int_{-W}^{W} a(f) S_x(f) a^H(f) df \mathbf{h}_k \\ &= \mathbf{h}_k^T \int_{-1/2}^{1/2} a(f) S_W(f) a^H(f) df \mathbf{h}_k \\ &= \mathbf{h}_k^T R_W \mathbf{h}_k, \end{aligned} \tag{2.27}$$

where $S_W(f)$ is equal to $S_x(f)$ in the band $(-W\,,\,W)$ and zero for all other frequencies. The covariance matrix R_W has a Toeplitz structure with elements

$$r_W(l) = r_x(l) \star 2W sinc(2\pi W \cdot l), \quad 0 \leq |l| \leq N-1, \tag{2.28}$$

where $r_x(l)$ is the covariance function of $x(n)$, $sinc(n) = \frac{sin(n)}{n}$ and $\star$ denotes convolution. The sinc-function is the Fourier transform of the box function in the frequency domain that sets the band $(-W\,,\,W)$.

One wants to choose K window functions which maximize the output power P_B. This optimization is performed with total power of a window equal to one, i.e.,

$$P_{tot} = \int_{-1/2}^{1/2} |H_k(f)|^2 df = \mathbf{h}_k^T \mathbf{h}_k = 1 \tag{2.29}$$

The solution with respect to $\mathbf{h}_k$ is the set of eigenvectors of the eigenvalue problem

$$R_W q_k = \lambda_k q_k, \quad k = 1...N. \tag{2.30}$$

These eigenvectors are then used as windows, or multitapers, to estimate a spectrum.

A few multitapers that are expected to be useful in this thesis are presented below.

2.5.3 The Thomson multitapers

The Thomson multitapers assumes the analyzed spectrum is white, i.e.

$$r_x(l) = \begin{cases} 1 & l = 0 \\ 0 & l \neq 0 \end{cases},$$

in eq. (2.28). The windows generated as a solution to eq. 2.30 are the Thomson multitapers but they are also known as Discrete Prolate Spheroidal

Sequences (DPSS) or Slepian windows. The eigenvectors corresponding to eigenvalues close to one have amplitude close to zero at the edges of the window and are appropriate to use as windows. These are the windows used for Thomson estimation. The Thomson multitaper can be found in `MATLAB` using the function `[h,V]=dpss(N,NW)`, which gives the $2 \cdot NW$ first windows of sample length N in h and their respective concentrations (in V) in the frequency band $|\omega| \leq 2\pi W$ where W is the half-bandwidth and ω is in radians/sample.

Realizations of Slepian windows can be seen in figure 2.13.

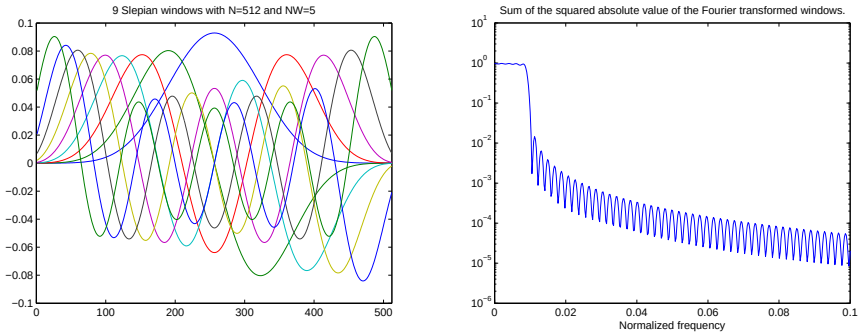


Figure 2.13: The nine first slepian windows when $N = 512$ and $NW = 5$ in the left plot and the sum of the squared absolute value of their Fourier transforms in the right plot.

2.5.4 The Hermitian multitapers

The Hermitian multitapers are somewhat related to the Thomson windows in that they both assume the analyzed spectrum will be white. The difference is that the Thomson multitaper finds the eigenvectors in the rectangle area $f \in [-W\,,\,W]$, where the Hermitian multitapers instead will find the optimal multitapers, in the mean square error sense, for a white spectrum in the area $t^2 + f^2 \leq R^2$, ie. a circle with radius R. This is proved by Daubechies in 1988 [1]. The $k:th$ order Hermite functions is defined by

$$h_k(t) = \pi^{-1/4}\sqrt{2^k(k-1)!}\left(t - \frac{d}{dt}\right)^{k-1} e^{-t^2/2} \qquad k = 0, 1, 2... \tag{2.31}$$

This form however is unpractical due to the nested derivatives. The recursive form of the functions is therefore easier to use.

$$\begin{aligned} h_0(t) &= e^{t^2/2}e^{-t^2} \\ h_1(t) &= 2te^{t^2/2}e^{-t^2} \\ h_k(t) &= 2t \cdot h_{k-1} - 2(k-2) \cdot h_{k-2} \qquad k = 2, 3, 4, ... \end{aligned} \tag{2.32}$$

These windows then have to be normalized to have unit energy. Depending on the radius R, the eigenvalues of these tapers can be found as

$$\lambda_k(R) = 1 - e^{R^2/2} \sum_{p=0}^{k} \frac{1}{p!} 2^{-p} R^{2p}. \tag{2.33}$$

As with the Thomson multitapers, one wants to use the tapers with eigenvalues as close to 1 as possible.

It has been shown that the Hermitian gives the best performance for white noise spectra in the time-frequency domain in terms of locality and orthonormality of the windows in the mean square error sense..

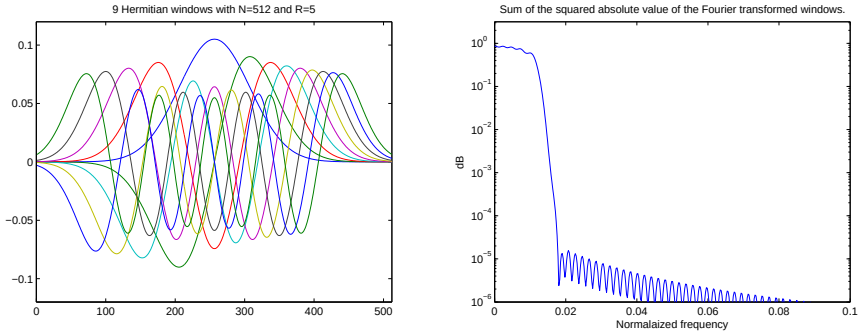


Figure 2.14: The nine first Hermitian windows when $N = 512$ and $R = 5$ in the left plot and the sum of the squared absolute value of their Fourier transforms in the right plot.

2.5.5 Peak matched multitapers

When estimating a PSD where peaks exists the Peak matched multiple windows (PMMW) have been shown to perform well [3]. The main idea of the PMMW is a combination of the Thomson windows and peak matching. Where the Thomson multitapers assumes a white spectrum ($\phi_x(f) = 1, \quad f \in [-B/2\,, B/2]$), the PMMW adds a peak shape to expected spectrum.

$$S_x(f) = \begin{cases} e^{\frac{-2C|f|}{10B \cdot log_{10}(e)}} & |f| \leq W \\ 0 & |f| > W \end{cases}. \tag{2.34}$$

A penalty function is also added to suppress leakage,

$$S_G(f) = \begin{cases} G & |f| \leq W \\ 1 & |f| > W \end{cases}, \tag{2.35}$$

where G is chosen to get a satisfactory result. Setting $G = 1000$, giving a penalty function of 30 dB has proven to be a good choice. The corresponding

Toeplitz covariance matrix is named R_G and will be added to the eigenvalue problem in eq (2.30) as

$$R_B q_k = \lambda_k R_G q_k, \quad k = 1...N. \tag{2.36}$$

Solving this generalized eigenvalue problem will generate the PMMW.

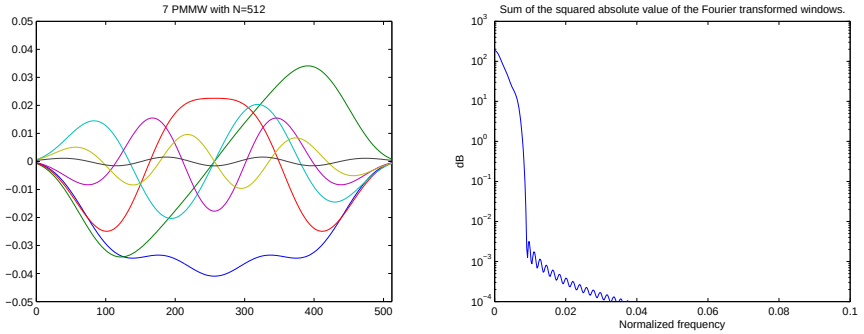


Figure 2.15: The 7 first PMMW with N=512 and a penalty function of 30 dB.

2.5.6 LSP optimal windows

Knowing that the spectrum one wants to find belongs to a locally stationary process, one can find the optimal windows in the mean square error sense. These are derived in [8] and [13]. This is done by setting up the integration of expected error of the spectrum in the ambiguity domain

$$J(\phi) = \int\int E\left\{\hat{A}_z(\nu,\tau)\phi(\nu,\tau) - E\left\{\hat{A}_z(\nu,\tau)\right\}\right\}^2 d\nu d\tau. \tag{2.37}$$

Minimization of the error function $J(\phi)$ then gives the optimal ambiguity kernel as

$$\phi_{opt}(\nu,\tau) = \frac{E\left\{\hat{A}_z(\nu,\tau)\right\}^2}{E\left\{\hat{A}_z^2(\nu,\tau)\right\}}. \tag{2.38}$$

Assuming we know the parameters of the LSP one can then find the optimal ambiguity kernel which can be transformed into desired quadratic domain.

The resulting windows turn out to be closely related to the Hermite multitapers but where each window has an optimized weighting factor that represent how much of the frequencies represented by each window is expected to be seen in the signal.

2.5.7 Time-frequency multitapering

The windows that are derived above are eigenvectors and hence one-dimensional. One can therefore use them, as they are, when estimating spectra or spectrograms. However one can also use them as kernels in the quadratic class by rotating them into a 2-dimensional structure.

3 Evaluation of methods on simulated signals

3.1 LSP modelling

To evaluate the scheme to estimate the parameters of an LSP, presented in section 2.2.1, I generated 20 realizations with $c = 9$, scaling frequency $F1 = 50$, an oscillating frequency of 0.125 and a signal length $N = 802$. White noise was added with variance $\sigma^2 = 1$. The 20 realizations war used to estimate the covariance matrix. Estimations are then made of the parameters c, $F1$, $f0$, a and σ^2.

This was then repeated 100 times to see how accurate the scheme is. The results are evaluated using root mean square error (RMSE). For example the RMSE of the estimation $\hat{c}$ of c is calculated as

$$RMSE(\hat{c}) = \sqrt{\frac{1}{100}\sum_{k=1}^{100}(\hat{c_k} - c)^2}.$$

The results can be seen in table 2. As can be seen in the same table the estimations of $f0$ and σ^2 are quite accurate. However, the estimations of the parameters c and $F1$ have large variations and seem to be biased. This is most likely due too the fact that numerical optimizations of multiple parameters are made simultaneously which may give rise to large errors. The parameter a was estimated too low due to the same reason.

Parameter	True value	Mean of estimations	RMSE
c	9	9.9642	3.5057
$F1$	50	60.9887	12.1036
$f0$	0.125	0.1249	$8.14 \cdot 10^{-4}$
a	1	0.4974	0.5097
σ^2	1	0.9997	0.0100

Table 2: RMSE of the estimated parameters in the LSP model.

When one has an estimation of the LSP parameters one can then find the optimal multitapers to estimate the spectrum of the process. However, as these estimates seem to have a relatively high variation, one may want to investigate how sensitive the optimal multitapers are to variations of these parameters. The estimations of the parameters maybe so bad that the corresponding optimal multitapers may perform worse than a more general method.

3.2 Time-frequency estimations

To evaluate the methods presented in section 2, simulations will be made using LSP. Realizations will be generated using specified parameters c and $F0$. As the covariance function of LSP is known one can find the true spectrum using equation 2.4.

$$\begin{aligned} S(t,f) &= \int_{-\infty}^{\infty} r(t,\tau)e^{-i2\pi f\tau}d\tau \\ &= \int_{-\infty}^{\infty} e^{-t^2/4}e^{-c/4\cdot\tau^2}e^{-i2\pi f\tau}d\tau \\ &= e^{-t^2/4}\sqrt{\frac{4\pi}{c}}\cdot e^{-(\pi f)^2 4/c}. \end{aligned}$$

Knowing this is the true spectrum, one can now estimate the spectrum using the different methods and measure the root mean square error (RMSE). I will here use the formula

$$RMSE = \sqrt{\frac{1}{N}\frac{1}{M}\sum_t\sum_f(\hat{S}(t,f) - S(t,f))^2} \tag{3.1}$$

where N is the number of samples in the signal and M is the number of points onto which the discrete Fourier transform is calculated.

Evaluations will be made on the Thomson multitapers, the Hermitian multitapers, Peak Matched multiple windows and the LSP optimal multitapers but also the Welch method as it is a commonly used method within the field and therefore a good reference result to compare the other methods with.

The main idea was to model the EEG signals as a LSP and estimate the parameters. As the optimization scheme used in this thesis did not find any good results (see section 4.3) I manually choose parameters that seemed to coincide with the EEG spectra. The parameters that where chosen were $c = 5$ and $F1 = 12$.

Each method will estimate the spectra of 100 realizations. The RMSE will then be calculated on all the 100 spectra and then averaged but I will also calculate the RMSE of the mean spectrum. When evaluating the different methods I manually found parameters that gave low RMSE, by simply testing values of the parameters and choosing those that gave the lowest RMSE. Optimization could of course in some sense be made but as the model itself needs improvement I felt this was not necessary. Instead trial and error methods were used to see what parameters generated good results.

Evaluations were made of the LSP optimal windows using both the true parameters and estimated parameters using the scheme from section 2.2.1. The estimation of the parameters was done using all 100 realizations at once to estimate the covariance matrix.

The parameters that where chosen for the different windows were:

- LSP - $c = 5$ and $F1 = 12$ as these were the parameters in the model.
- LSP est - $c = 6.0877$ and 13.5617
- Thomson MT - $NW = 4$, $N = 120$ and $K = 3$.
- Hermite - $R = 16$, $N = 110$ and $K = 4$.

- PMMW - $N = 100$ and $K = 9$.
- Welch - $N = 100$, $dN = 50$ and 50% overlap.

The frequency will only be estimated up to $Fs/4$; this could have been chosen even lower as the frequencies are concentrated around 0 Hz.

As the methods are not optimized to estimate the frequencies of a LSP they may estimate the power wrong. Assuming this power is the same in the entire time-frequency plane a scaling factor is also estimated for every method. This scaling factor is optimized using mean square error.

$$\min_a \sum_t \sum_f (a \cdot \hat{S}(t,f) - S(t,f))^2$$

The RMSE results are presented in table 3 and the estimated spectra can be seen in figure 3.1. One can there see that Thomson and the Hermite multitapers generate similar results. The Thomson tapers perform slightly worse though and the expected square shape of the Thomson spectrum fits the the true LSP spectrum worse as well. I will therefore not use these tapers in further analysis of the EEG signals. The LSP optimal tapers naturally gave the best RMSE as they are optimal in shape and weighting in the mean square error sense.

Method	Mean RMSE of 100 spectra	RMSE of mean spectrum
LSP optimal	0.0557	0.0105
LSP est	0.0753	0.0234
Thomson MT	0.0731	0.0429
Hermite	0.0675	0.0263
PMMW	0.0970	0.0691
Welch	0.0892	0.0539

Table 3: RMSE of spectra estimated with different methods.

4 Results

4.1 Pre-processing of data

The signals where initially stored in EEGlab files. Here, a low pass filter at 200 Hz and a high pass filter at 0.03 Hz where applied and the signals where EOG artifact corrected. After converting them to `MATLAB` files a script was constructed that retrieved the required set of responses. The different responses where sorted in channels and visual field sets i.e. the responses from different locations of the scalp depending on where the string of letters where shown. The possible channels are shown in figure (1.1) and the possible visual fields are the left visual field (lvf), the center visual field (cvf) and the right visual field (rvf).

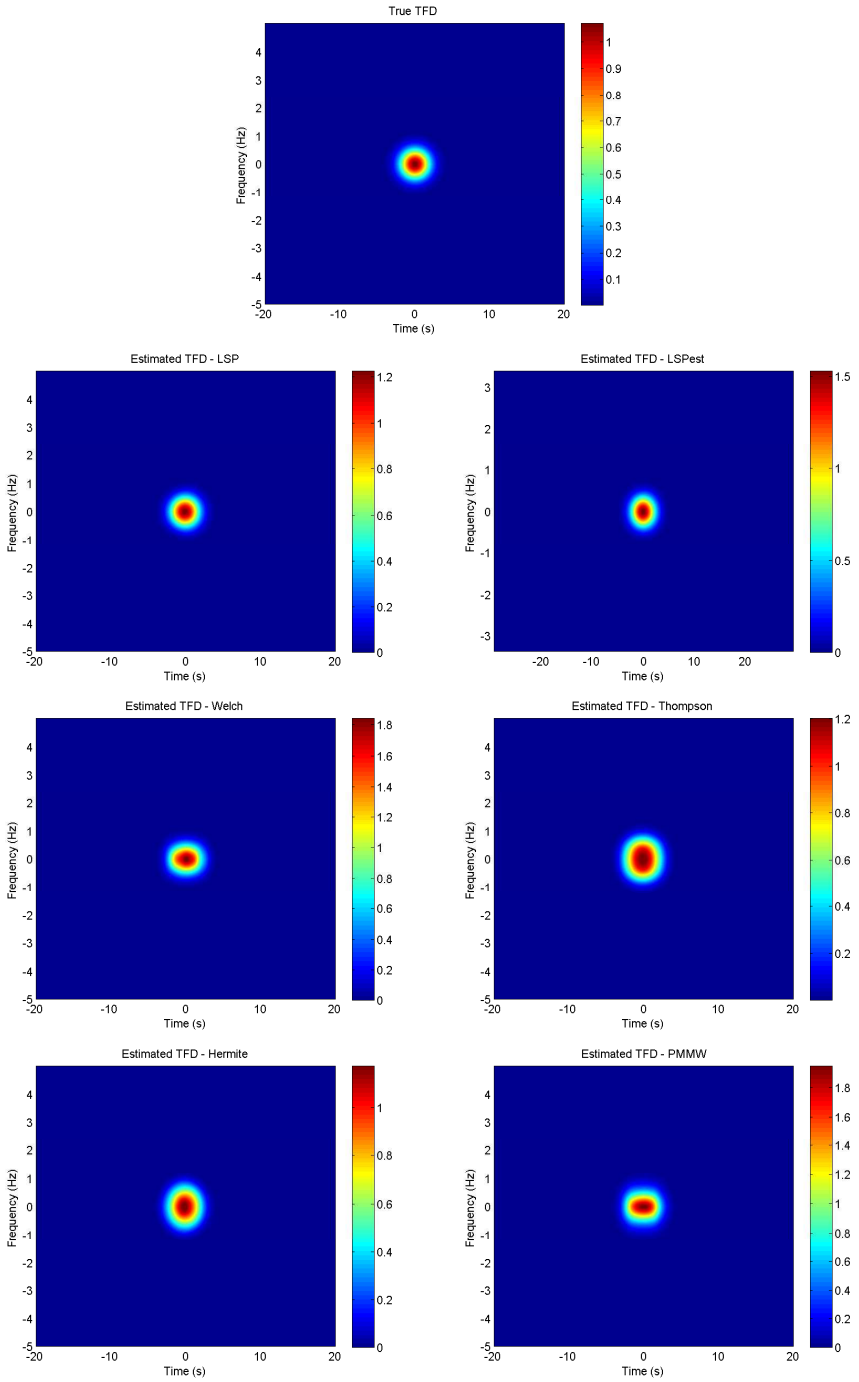


Figure 3.1: Estimated spectra of the LSPs using four different methods. The true spectrum calculated analytically can be seen in the top plot.

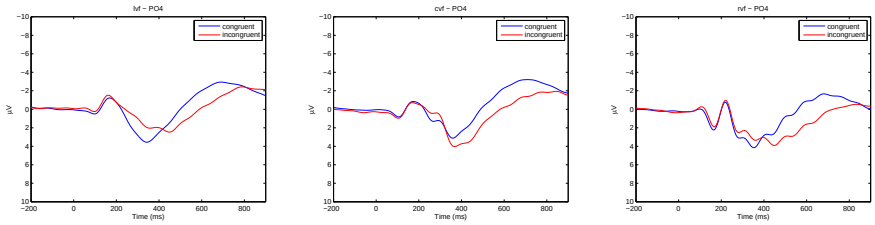


Figure 4.1: Grand averages from the $PO4$ channel when the strings of letters where shown in the three different visual fields. (Note that grand averages are usually plotted with inverted y-axis)

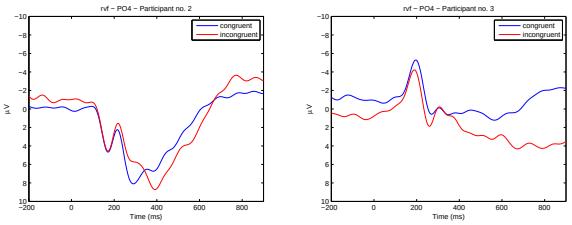


Figure 4.2: Average responses from two different participants taken from the same channel in response of stimuli from the right visual field.

The signals contained noise and oscillations that are of no interests in this thesis, eg. the classic 50 Hz main electricity frequency. An additional low-pass FIR filter with a 20 Hz cut off frequency was therefore applied. As the set of signals was so large, this filter was applied on the grand averages. The signals were also Hilbert-transformed.

4.2 Grand Average and difference curves

Computing the grand average of the signals when sorted into congruent and incongruent sets and then into visual-field-subsets gives some immediate visual hints of the final results. As mentioned in section 1.3 the main interest lies in the $N2$ event which here seems to occur around 300 ms. In figure 4.1 grand-averages from the $PO4$ channel can be seen. As $PO4$ is a channel placed on the posterior right part of the scalp, and the biggest difference is then expected to be seen when the string is shown in *lvf* and less difference when shown in *cvf* or *rvf*.

The variance of the averages between each participant is pretty high. One should therefore note that it is a very strong assumption to make; that every persons brain response is the same. As an example see figure 4.2 where the average of responses from participant 2 and 3 are shown separately. The averages are formed from the same channel and in response to the same stimuli but one can still see that their EEG responses are very different.

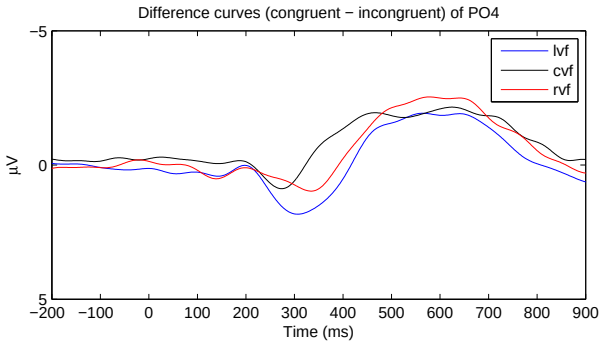


Figure 4.3: Difference curves from channel PO4. The biggest difference is seen when strings are shown in the left visual field which coincides with the hypothesis as PO4 is a channel on posterior right part of the scalp.

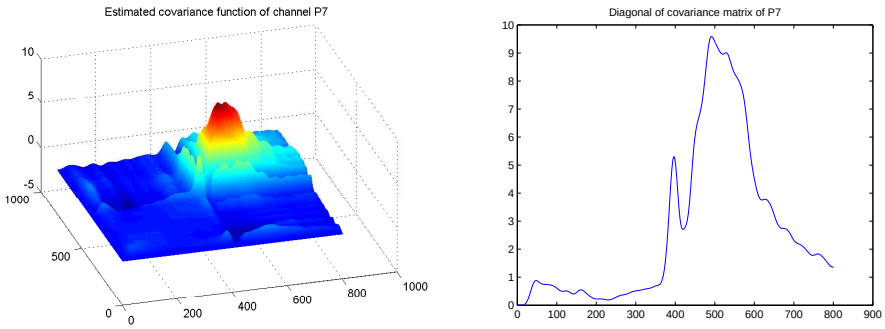
An example of the difference curves can be seen in figure 4.3, where the three curves are the difference in responses of the three visual fields measured at the $PO4$ channel.

4.3 LSP modelling of EEG

The scheme was not so successful on real EEG data. Primarily the brain responses were much too complex to be modeled as a single LSP. An example of this is seen in figure 4.4. As the diagonal of the covariance matrix seen in figure 4.4b contains multiple components, trying to fit one single Gaussian function will then return bad results.

Some of the channels had potential to be modeled well as a LSP process, An example of this is shown in figure 4.5. However the scheme still performed bad and the LSP model that is fitted in the figure is therefore fitted manually. Again, that the scheme fails was probably due to the fact that multiple parameters where being optimized simultaneously in separate steps. Another problem was that the solution found often resulted in that the parameter $\hat{c}$ was set to a value less than 1. As $c \geq 1$ was a requirement, a simple function was added to the anti-diagonal that penalized c-values under one. This however only resulted in $\hat{c} = 1$.

The solutions I used for this problem was to manually fit the parameters to fit the $N1$ and $N2$ events as it was here that the point of interest lies. The parameters I ended up using was $F1 = 12$ and $c = 5$.



(a) (b)

Figure 4.4: (a) Cross covariance matrix of channel P7, where to stimuli is congruent and in the lvf.
(b) Diagonal of the covariance matrix to the left.

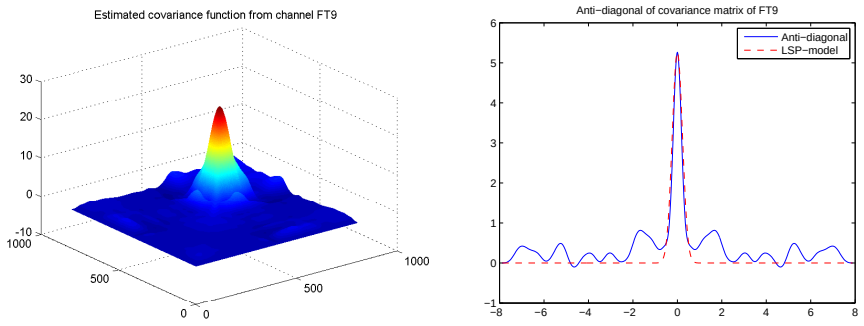


Figure 4.5: Cross covariance matrix of channel FT9, where to stimuli is congruent and in the lvf.

4.4 Time-frequency analysis

As there are 36 channels to be analyzed, three possible visual fields where the strings can be shown and a set of different analysis methods to be used, this result section consists of a large set of plots. I have chosen four of the channels to present in the report: $PO3$ (left posterior channel), $PO4$ (right posterior channel), CZ (middle center) and FZ (center anterior). In this section however only the plots for the channel $PO4$ will be shown. The rest of the plots can be found in Appendix A. All plots are zoomed in to show frequencies between 0 and 20 Hz and the time span -100 and 900 ms in comparison to when the letters show up on the screen. The color scaling of the plots is the same for each pair of channel and method.

The parameters used in the different methods are the ones presented in section 2, where the parameters are set to fit the LSP model.

4.4.1 Welch's method

Even though the Welch's method performed well on the LSP it did not perform as well on the EEG-signals. Even though a shorter time-window was used compered to the other multitaper methods a worse frequency resolution is acquired. In figure 4.6 one can see the spectra of channel PO4 for congruent and incongruent stimuli from the left visual field. Each time window was chosen to $N = 100$ samples and the inner windows to $Nd = 50$ samples where each sub-window overlaps 50 %.

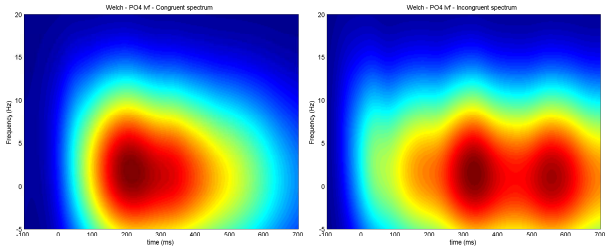
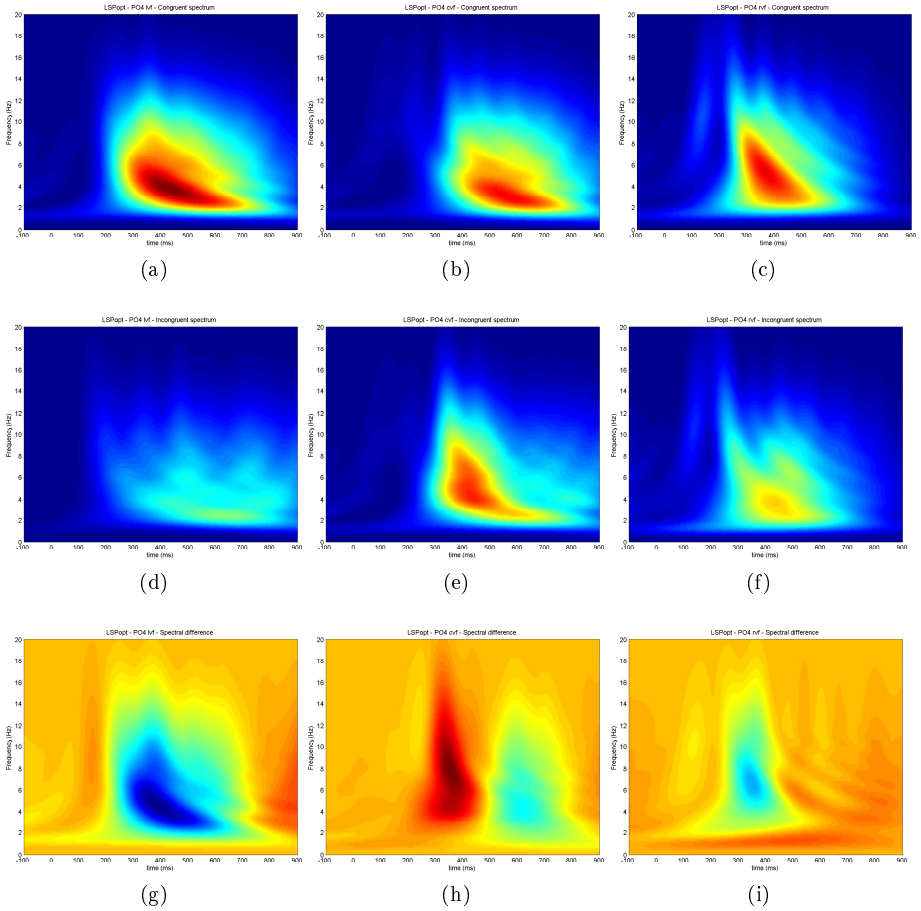


Figure 4.6: Welch's method used to estimate the spectra of channel PO4 for stimuli from the left visual field.

4.4.2 LSP optimal multitapers

The LSP optimal multitapers give descent results around time $0 - 400\,ms$, but there seem to be some oscillating effects after that. During the time span $0 - 400$ ms this method returns the best results in terms of locality.



(a) (b) (c)

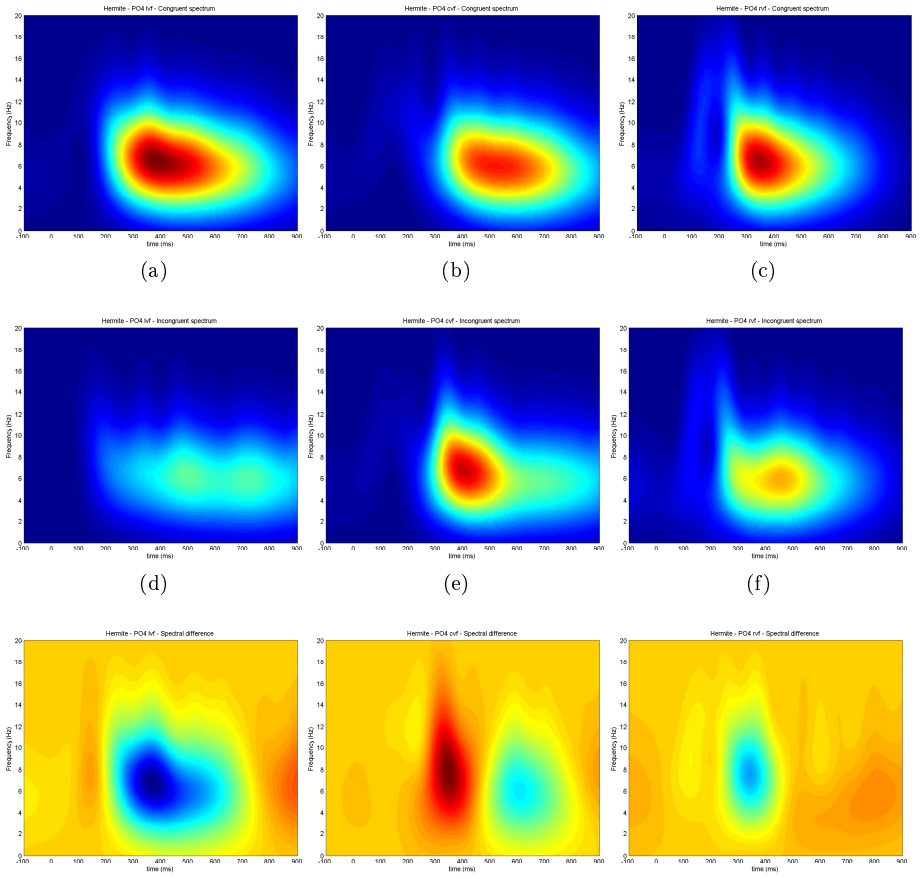
(d) (e) (f)

(g) (h) (i)

Figure 4.7: PO4 - LSPopt

4.4.3 The Hermite windows

Of the more general methods the Hermite windows returns the best results in terms of concentration.



(a) (b) (c)

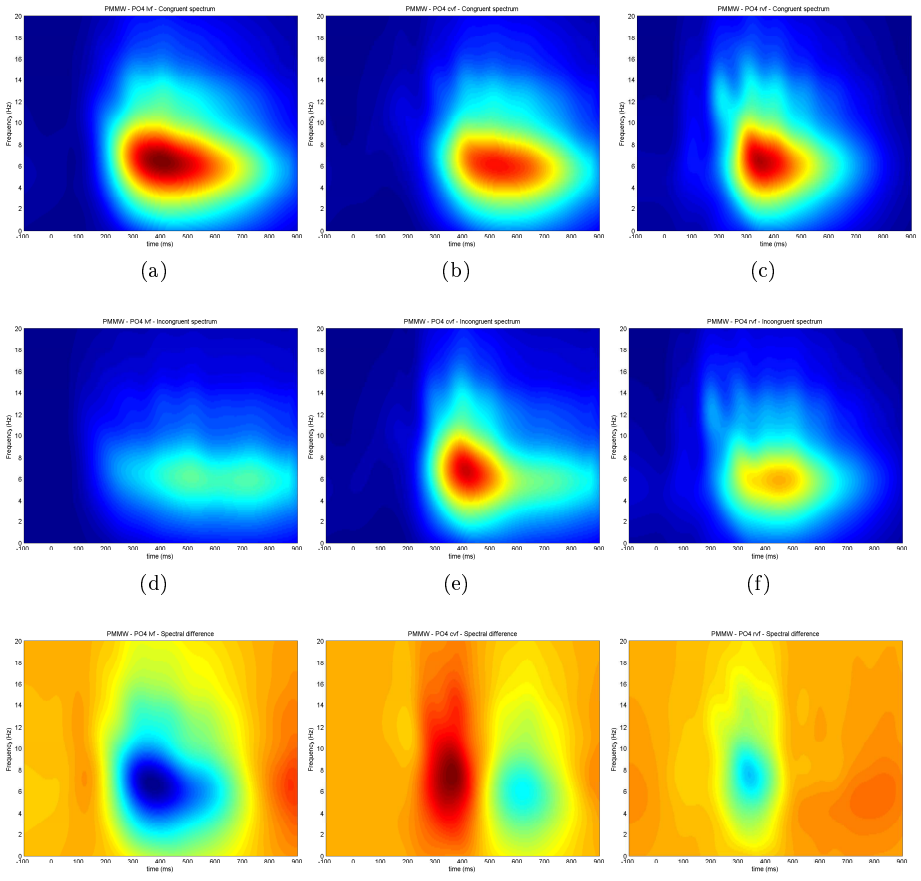
(d) (e) (f)

(g) (h) (i)

Figure 4.8: PO4 - Hermite

4.4.4 Peak matched multitapers

The PMMW returns very similar results compared to the Hermite windows.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 4.9: PO4 - PMMW

4.4.5 Interpretation of time-frequency representations

Interpreting the result plots from a psychological point of view may be better done by someone trained in the field.

However just looking at the plots for $PO3$ and $PO4$, both seem to follow the hypothesis.

The spectral difference plots of $PO3$ (appendix A.0.1) shows a bigger difference at the right side of the scalp. The difference seem to be the lowest on the right part of the brain at time 300 ms, where as clear frequency differences are seen for lvf and cvf.

Comparing the spectra for $PO4$ (appendix A.0.2) visually would support the hypothesis, where a more powerful response is seen for congruent lvf than incongruent. A much stronger difference is seen with a mass center around 350 ms.

Looking at the $N2$ events of FZ the results are quite inconclusive. A stronger response is seen for congruent stimuli on lvf. A stronger response for *incongruent* stimuli on cvf and about the same response for both stimuli on rvf.

CZ (appendix A.0.3 and A.0.4) seem to get a more bilateral difference between congruent and incongruent stimuli, where the same response was given whether the stimuli was from lvf or rvf. A stronger response is seen for the incongruent stimuli at both these visual fields.

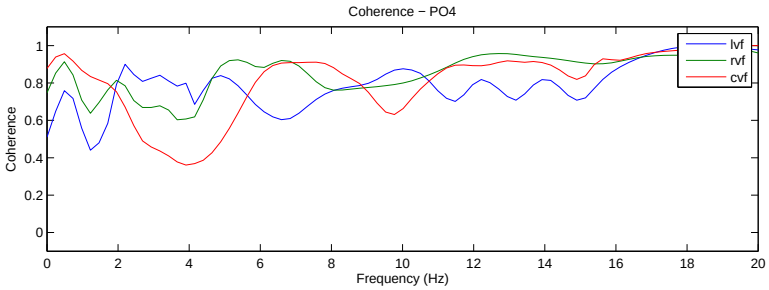


Figure 4.10: Coherence between brain responses to congruent stimuli and incongruent stimuli.

4.5 Coherence

4.5.1 Frequency coherence

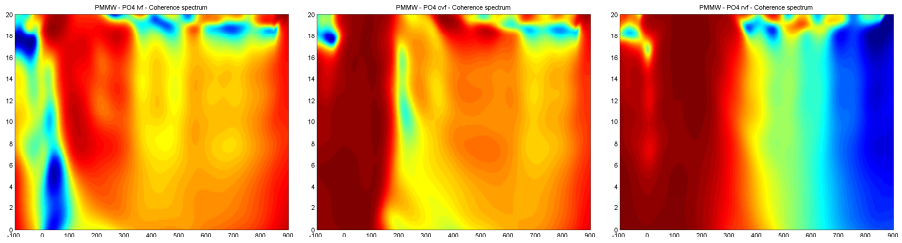
An attempt at frequency coherence was made to analyze the signals without using a time dimension. This however gave no good results as the signals are not stationary enough. When looking at the whole signals they seem to have roughly the same frequency content. The coherence between congruent stimuli and incongruent stimuli from channel $PO4$ can be seen in figure 4.10. One can notice that the biggest dips are seen are seen roughly at the same frequencies for all visual fields, ie. ~ 1 Hz and ~ 4 Hz. As they are same for the three visual fields they can most likely be derived from the later brain responses ($400 - 1000$ ms) where the responses always seem to differ, more or less, between congruent and incongruent stimuli.

4.5.2 Time-frequency coherence

When calculating the time-frequency coherence, I used the same multitapering methods as for the time-frequency spectra. However some of the methods generated negative values within the region of analysis and one could therefore not compute the coherence. The peak-matched multitapers proved to be must robust in this part and I therefore only present the results using this method. Again I did the analysis on the channels $PO3$, $PO4$, FZ and CZ with only $PO4$ presented here and the rest in Appendix B.

When looking at the coherence plots for $PO3$ and $PO4$ they coincide with the hypothesis. Lower coherence is seen around 300 ms contra-lateral to the stimuli.

The results for CZ and FZ are rather inconclusive. CZ have high coherence at all frequencies and for all visual fields. FZ have lower coherence for lvf stimuli and roughly the same for cvf and rvf.



(a) (b) (c)

Figure 4.11: Time-frequency coherence of PO4.

4.5.3 Dual-Frequency coherence

The plots in figure 4.12 show the dual-frequency-plots for the two electrodes $P3$ and $P4$, that are positioned at the back of the head . The strings are shown in the right visual field and we can therefore see clear differences between the congruent and the incongruent auto-dual-frequency plots at $P3$, i.e. the left side of the brain, which was expected. At $P4$ one can see that the plots are similar. However it is very hard to interpret the dual-frequency coherence between the signals as all plots are very noisy, and the signals are built up of multiple components where many of them seem to be coherent and it is therefore unclear what one is really looking for. The signals are clearly similar with much coherence in the frequency band < 15 Hz.

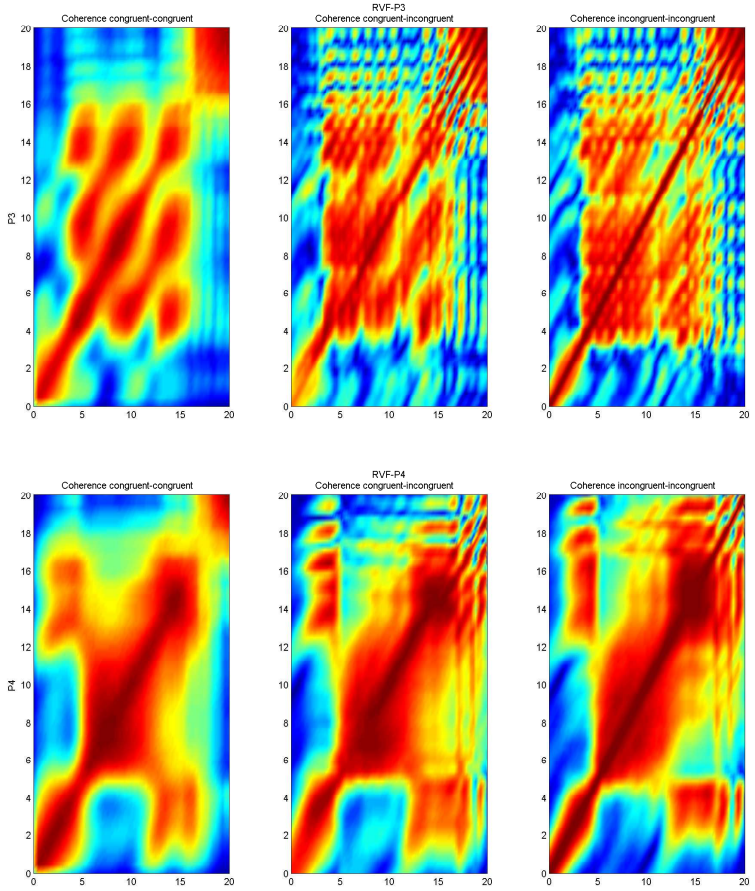


Figure 4.12: Dual-frequency coherence of EEG-signals. The top three plots a measured from the left posterior part of the brain and the bottom three plots from the right posterior of the brain. In this case the letters were shown in the right visual field and one can therefore expect differences in the left posterior part.

The time-dual-frequency coherence was also implemented and tested but generated no satisfactory results as they where too noisy and hard to interpret due to the four dimensionality. Therefore these results will not be presented here.

5 Discussion and comments

5.1 Pre-processing of data

This is a section where much more could have been done. Assumptions were now made that all subjects reacted the same every time to a certain stimuli. One could look at how responses differ between subjects. How long was the mean reaction time of each participant. Do participant react the same to stimuli throughout the experiment or do they respond differently when they are more familiar with the possible stimuli.

5.2 LSP modelling of EEG

I must first mention that LSP modelling worked quite well when used on channels that gave less powerful responses on the stimuli. The LSP model did not turn out very well when looking at more active EEG signals. One main problem is that a LSP only contains one component, centered around time 0. As most EEG evoked potentials consist of multiple component, that all occur after time 0, the model has to be altered. One solution to this might be to consider the signals as sums of LSPs with different center times. The question is then how complicated one should make the model? As the brain may respond very differently to other performed task, should one try to fit a model to each task and channel or are the responses similar enough to make one descent model?

Another problem is the optimization of parameters in the model; the scheme used in his thesis has to be improved. Best case scenario would be to find an analytic expression for this estimation, instead of the now used numerical solution. If an analytic expression can not be found, one would at least like to find the parameters in one step, instead of four.

5.3 Time-frequency analysis

Multiple methods where used and evaluated on the model. To evaluate the methods I tested them on simulated LSP. One might want to try them out on many different LSPs to give a more wide result instead of just a single one as in this thesis.

Choosing a way to present the results of the analysis on the EEG signals was also a problem. Having multiple channels and analysis methods generated an abundance of plots that would more than double the number of pages in this thesis if presented. As the main concern of this thesis was to implement and evaluate the multitaper methods I felt that I wanted to keep the number of plots low enough to be able to take them all in but high enough to get an idea of the results.

Something that was also tested in this thesis was to estimate a spectrum of every individual's result and then average over all subjects. As every persons brain response differed this only gave a quite smudged result that was hard to interpret.

References

[1] Daubechies, I., "Time-Frequency Localization Operators: A Geometric Phase Space Approach". *IEEE Trans. on Inf. Theory*, vol. 34, No. 4, pp 605-612, 1988.

[2] Eriksen, B. A. & Eriksen, C. W., "Effects of noise letters upon the identification of a target letter in a nonsearch task". *Perception & Psychophysics*, vol. 16, No. 1, pp 143-149, 1974

[3] Hansson, M. & Salomonsson, G. "A Multiple Window Method for Estimation of Peaked Spectra". *IEEE Trans. of Signal Processing*, vol 45, No. 3, pp. 778-781, 1997.

[4] He, P., Wilson, G., Russel, C. "Removal of ocular artifacts from electro-ecephalogram by adaptive filtering". *Med. Biol. Eng. Comput.*, No. 42, pp 407-412, 2004.

[5] Liu, C., Gaetz, W. & Zhu, H. "Estimation of Time-Varying Coherence and Its Application in Understanding Brain Functional Connectivity". *EURASIP Journal on Advances in Signal Processing*, Vol. 2010, Article ID 390910, 11 pages, (2010)

[6] Matz, G., & Hlawatsch, F., "Time-Frequency Coherence analysis of Nonstationary Random Processes". *Proc. 10th IEEE-SP Workshop on Statistical Signal and Array Processing*, Pocono Manor (PA), pp. 554-558, Aug. 2000.

[7] Mellors, R. J., Vernon F. L. & Thomson, D. J. "Detection of dispersive signals using multitaper dual-frequency coherence". *Geophysical Journal International*, 135, pp. 146-154, 1998.

[8] Sayeed, A. M., & Jones, D. L. "Optimal Kernels for Nonstationary Spectral Estimation". *IEEE Trans. of Signal Processing*, vol 43, No. 2, pp. 478-491, 1995.

[9] Silverman, R. A., "Locally Stationary Random Processes", *IRE Trans. Inf. Theory*, vol. 3, pp. 182-187, 1957.

[10] The Swartz Foundation. "EEGlab Homepage". Available: *http://sccn.ucsd.edu/eeglab/*. Last accessed 4th June 2012.

[11] Sörnmo, L. & Laguna, P., "Bioelectrical Signal Processing in Cardiac and Neurological applications." *Elsevier Academic Press*, 2005.

[12] Thomson, D. J., "Spectrum Estimation and Harmonic Analysis". *Proceedings of the IEEE*, vol. 70, No. 9, pp 1055-1096, 1982.

[13] Wahlberg, P. & Hansson, M., "Kernels and Multiple Windows for Estimation of the Wigner-Ville Spectrum of Gaussian Locally Stationary Processes", *IEEE Trans. of Signal Processing*, vol. 55, No. 1, pp. 73-84, 2007.

[14] Xu, Y., Haykin, S. & Racine, R. J., "Multiple Window Time-Frequency Distribution and Coherence of EEG Using Slepian Sequences and Hermite Functions", *IEEE Trans on Biomedical Engineering*, vol. 46, No. 7, pp 861-866, 1999.

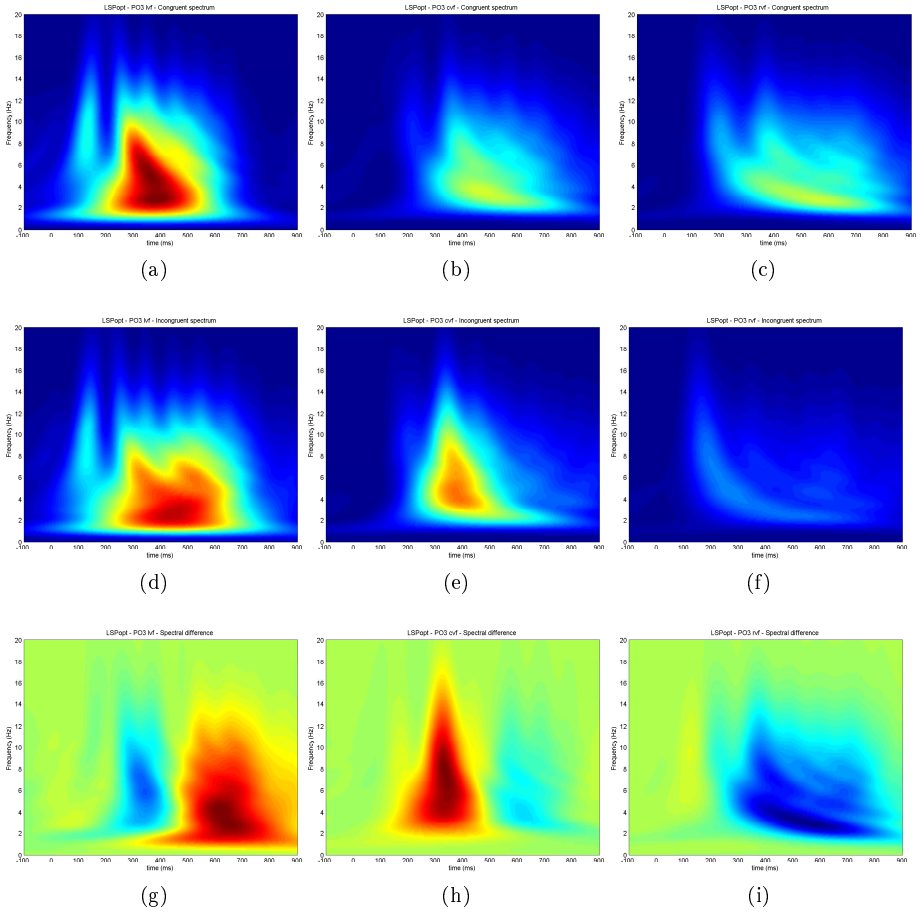
A Time Frequency plots

Here are the spectrum plots for the channels $PO3$, $PO4$, FZ and CZ. As there are multiple plot there are no captions on the images. Instead they all follow the same pattern as presented in table 4. The different channels are presented in their own subsection and each method of analysis on its own page. As mentioned previously the color scaling is the same for each pair of channel and methods. Hence the same color scale is for plots on the same page.

Time-frequency Congruent Left Visual Field	Time-frequency Congruent Central Visual Field	Time-frequency Congruent Right Visual Field
Time-frequency Incongruent Left Visual Field	Time-frequency Incongruent Central Visual Field	Time-frequency Incongruent Right Visual Field
Time-frequency Difference in frequency power Left Visual Field	Time-frequency Difference in frequency power Central Visual Field	Time-frequency Difference in frequency power Right Visual Field

Table 4: Table showing how plot are presented.

A.0.1 PO3

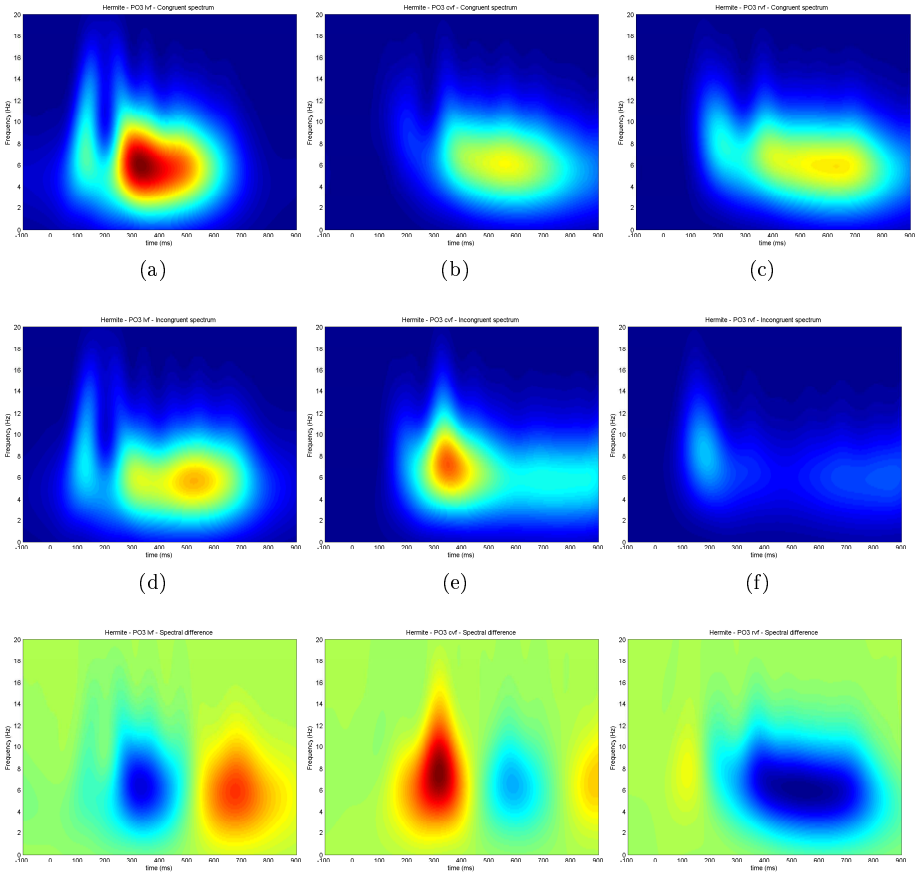


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure A.1: PO3 - LSPopt

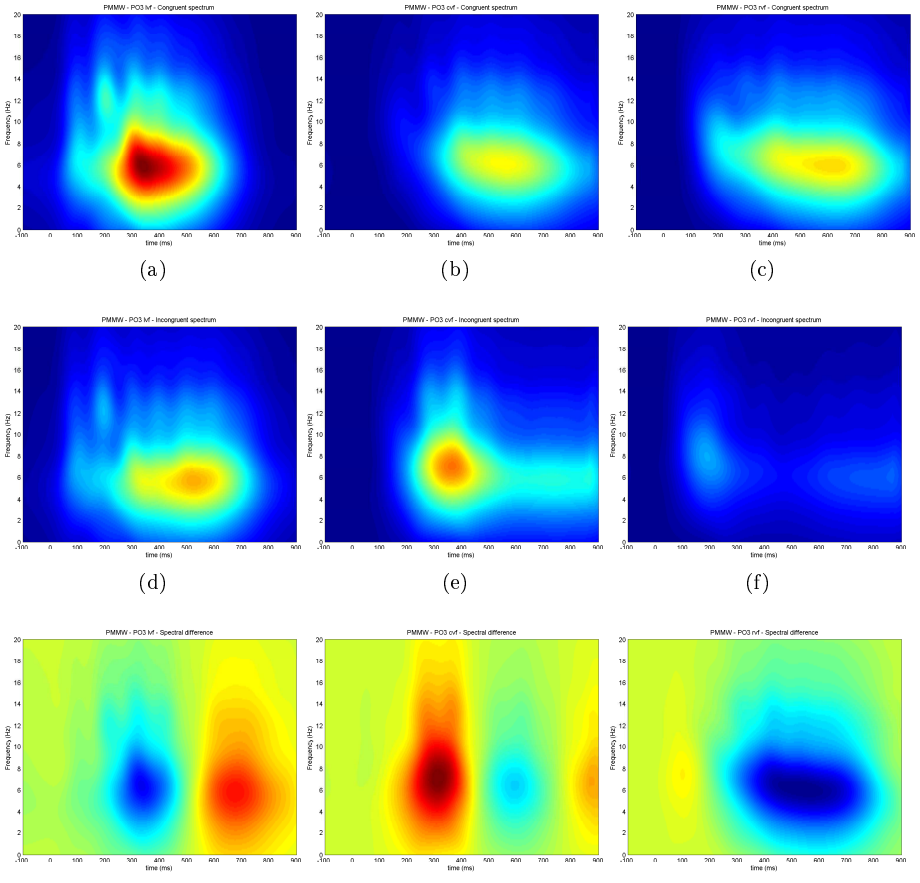


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure A.2: PO3 - Hermite



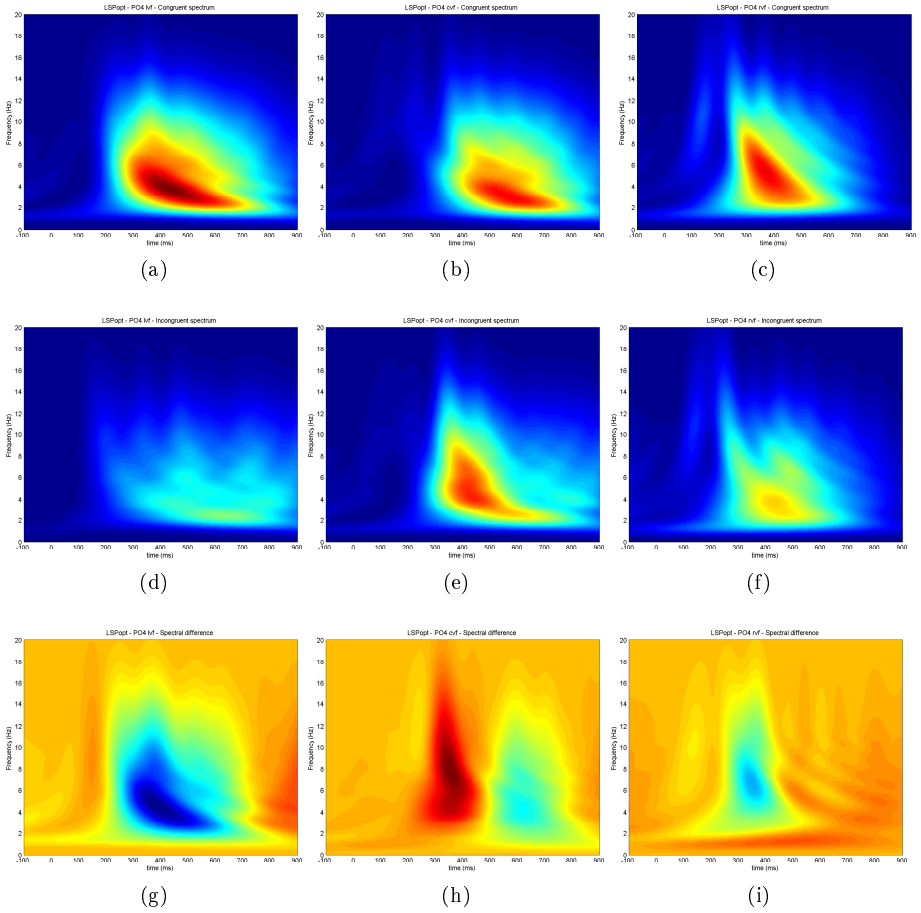
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure A.3: PO3 - PMMW

A.0.2 PO4

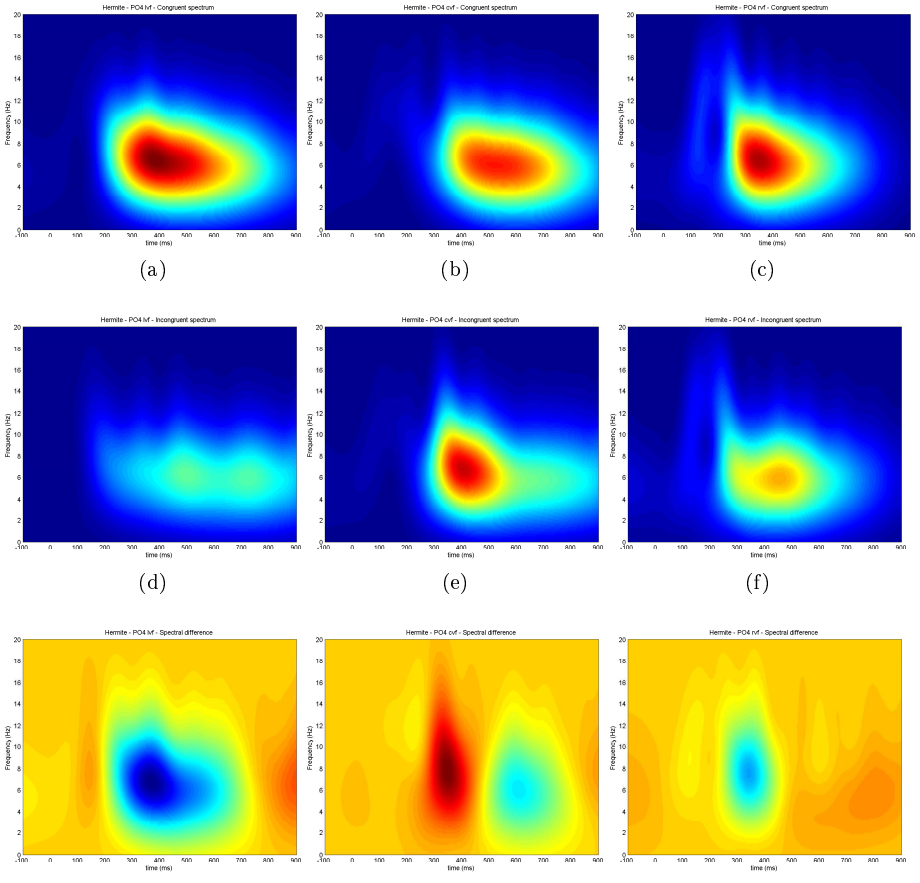


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure A.4: PO4 - LSPopt

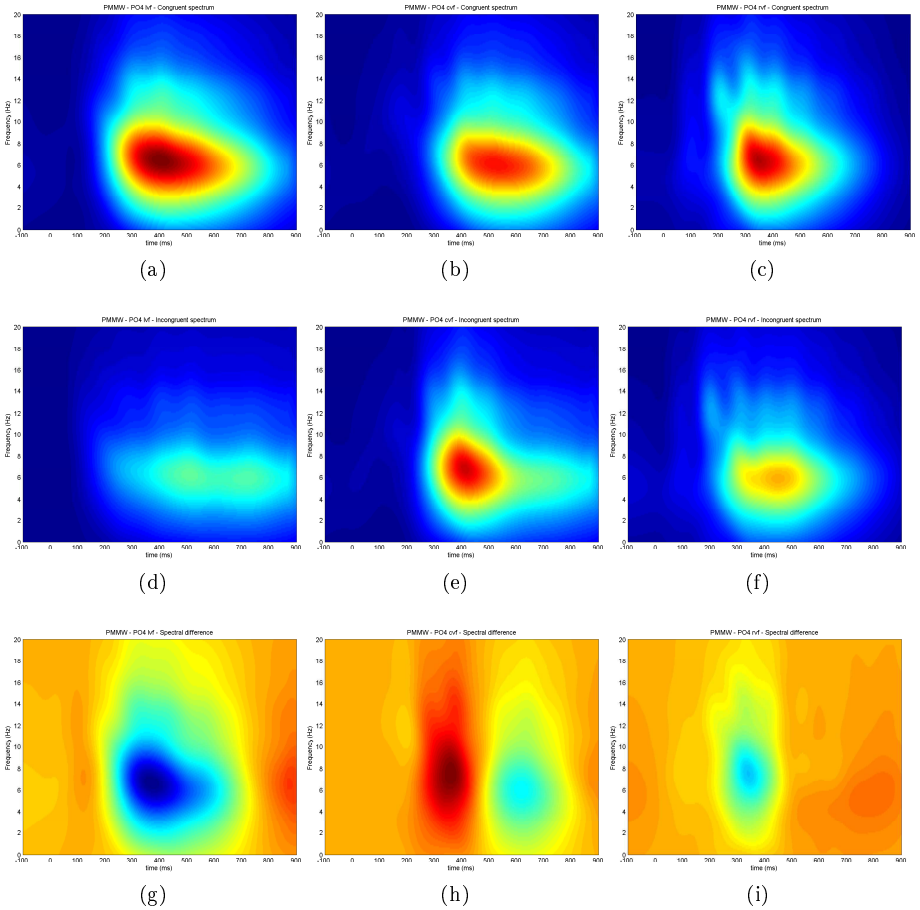


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure A.5: PO4 - Hermite



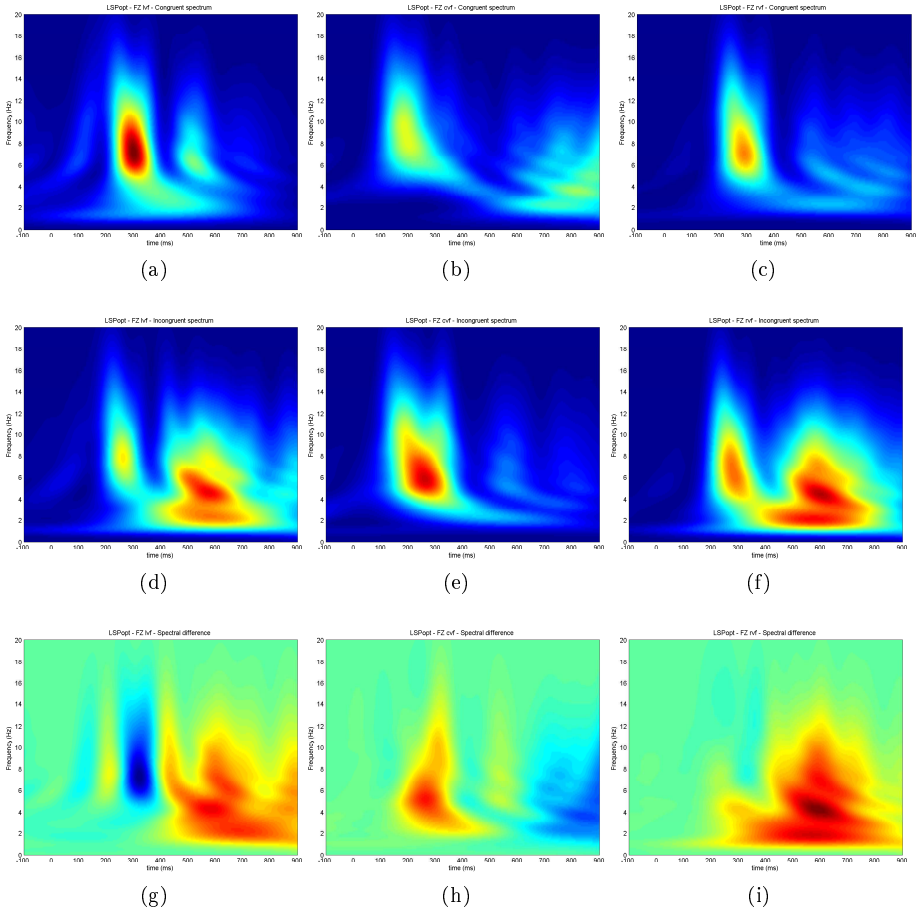
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure A.6: PO4 - PMMW

A.0.3 FZ

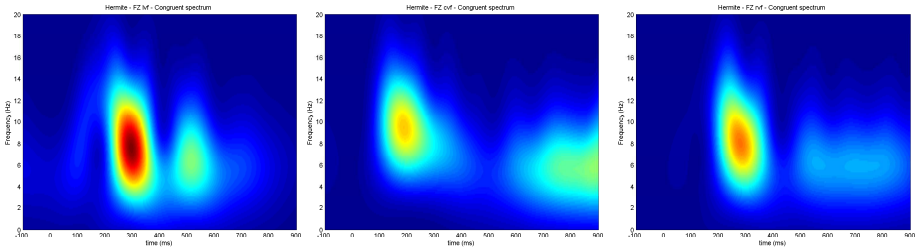


(a) (b) (c)

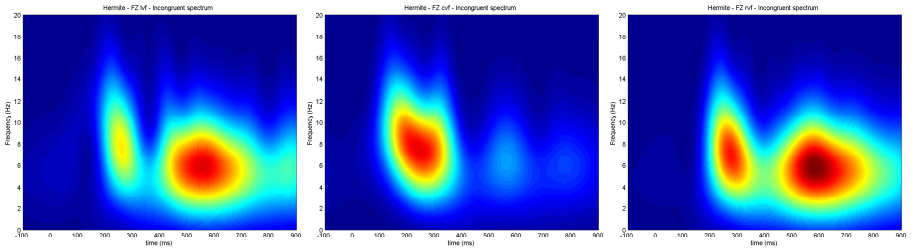
(d) (e) (f)

(g) (h) (i)

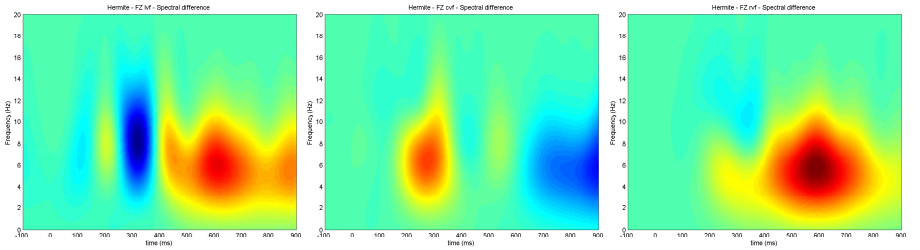
Figure A.7: FZ - LSPopt



(a) (b) (c)

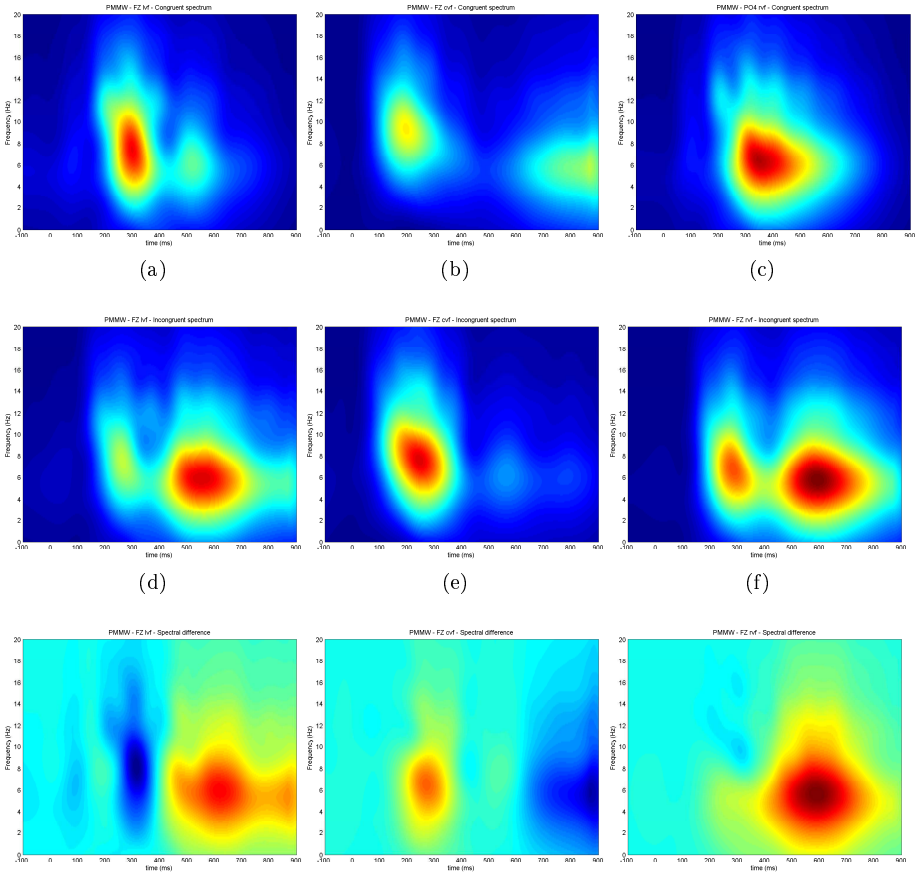


(d) (e) (f)



(g) (h) (i)

Figure A.8: FZ - Hermite



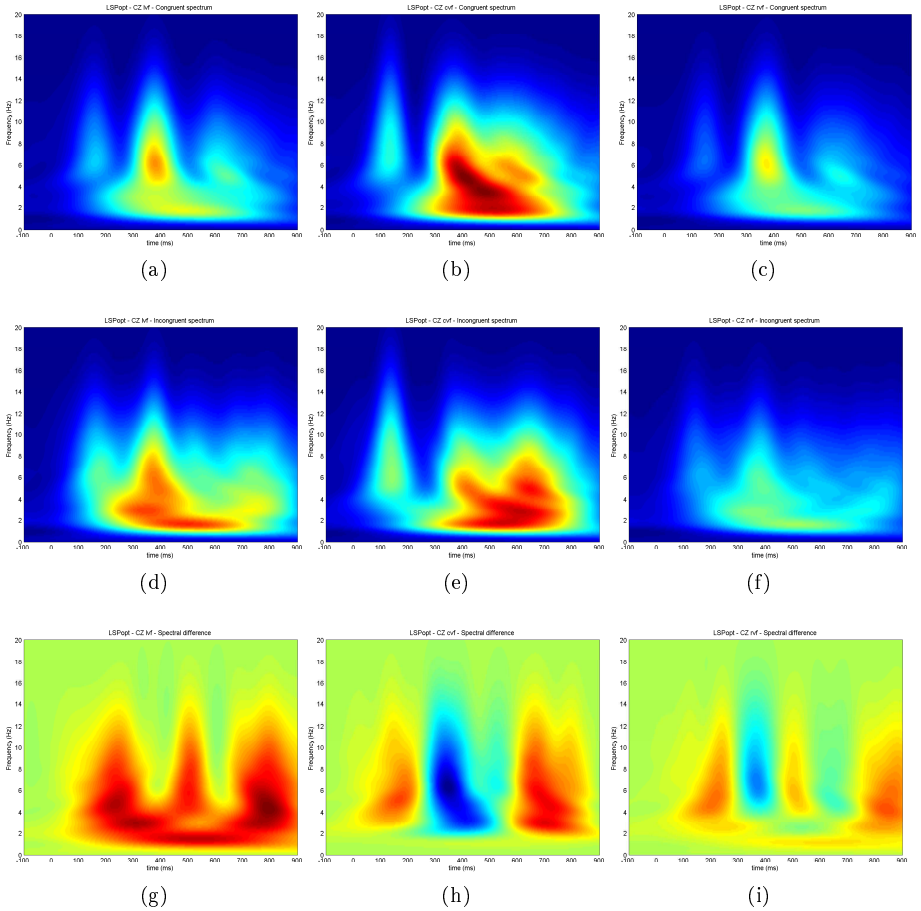
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure A.9: FZ - PMMW

A.0.4 CZ

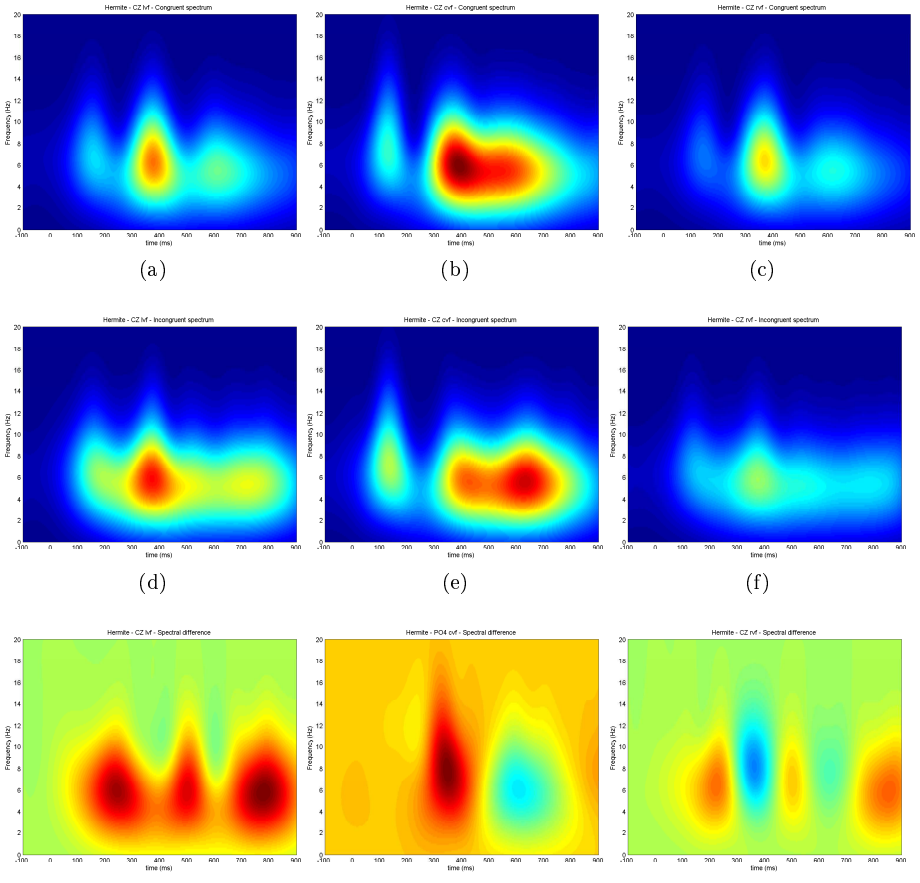


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure A.10: CZ - LSPopt

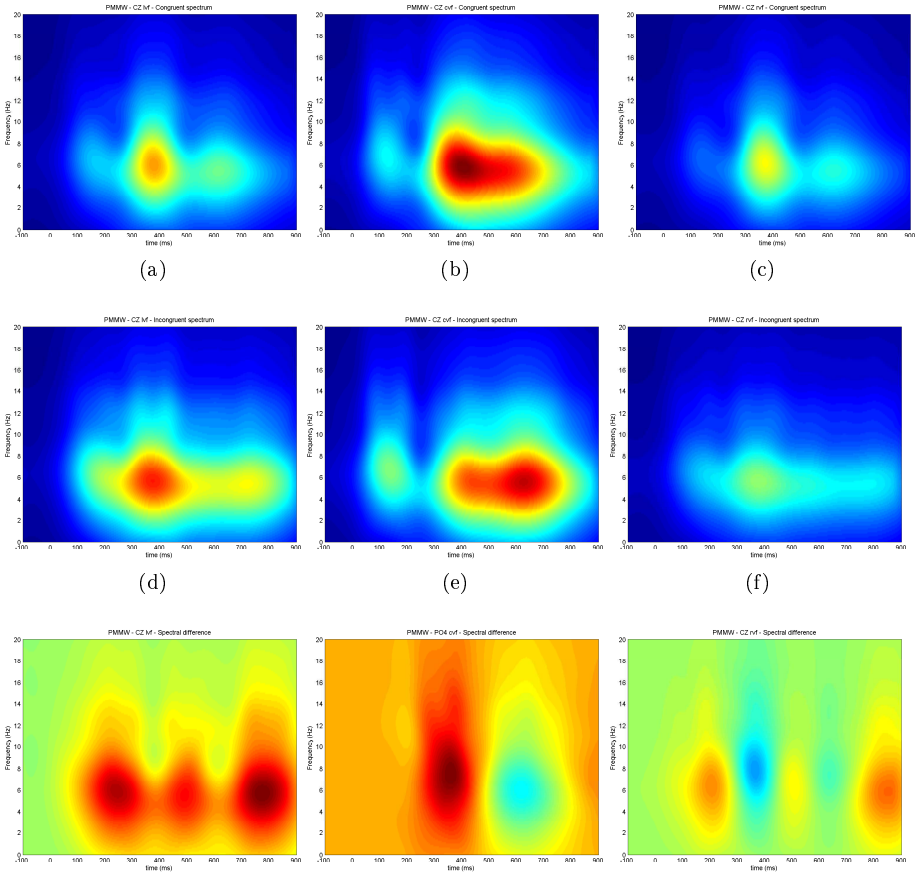


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure A.11: CZ - Hermite



(a) (b) (c)

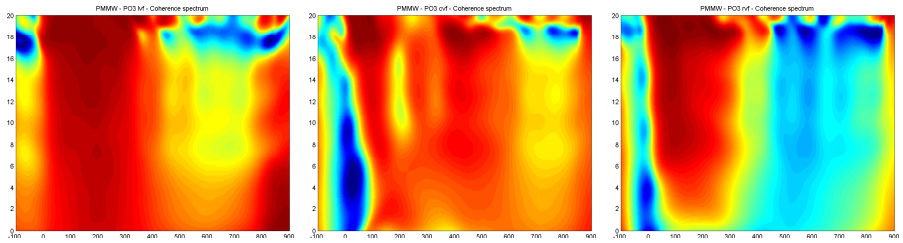
(d) (e) (f)

(g) (h) (i)

Figure A.12: CZ - PMMW

B Time Coherence plots

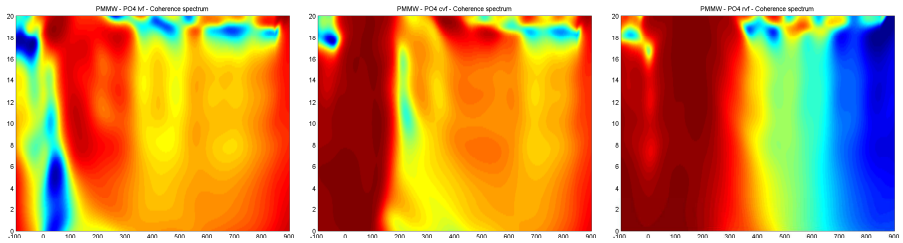
B.0.5 PO3



(a) (b) (c)

Figure B.1: Time-frequency coherence of PO3.

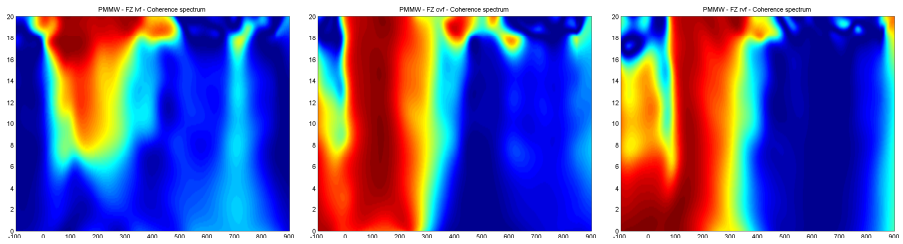
B.0.6 PO4



(a) (b) (c)

Figure B.2: Time-frequency coherence of PO4.

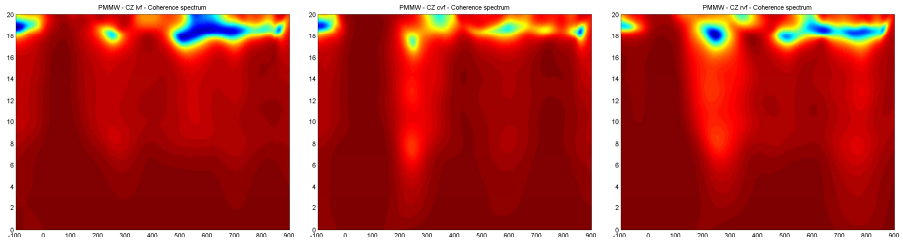
B.0.7 FZ



(a) (b) (c)

Figure B.3: Time-frequency coherence of FZ.

B.0.8 CZ



(a) (b) (c)

Figure B.4: Time-frequency coherence of CZ.