Maths for Brain Imaging: Lecture 8

W.D. Penny

Wellcome Department of Imaging Neuroscience,
University College, London WC1N 3BG.

November 29, 2006

1   Contents

- Sinusoidal models

- Fourier transform

- Welch's method

- Multitaper method

- Multivariate spectral analysis

- Source reconstruction of MEG Gamma activity

1.1 Sines and cosines

Sines and cosines can be understood in terms of the vertical and horizontal displacement of a fixed point on a rotating wheel; the wheel has unit length and rotates *anti-clockwise*. The angle round the wheel is measured in degrees or radians $(0 - 2\pi$; for unit radius circles the circumference is 2π, radians tell us how much of the circumference we've got). If we go round the wheel a whole number of times we end up in the same place, eg.$\cos 4\pi = \cos 2\pi = \cos 0 = 1$. Frequency, f, is the number of times round the wheel per second. Therefore, given $x = \cos(2\pi ft)$, $x = 1$ at $t = 1/f, 2/f$ etc. For $x = \cos(2\pi ft + \Phi)$ we get a head start (lead) of Φ radians. Negative frequencies may be viewed as a wheel rotating *clockwise* instead of anti-clockwise.

1.2 Sampling and aliasing

If we assume we have samples of the signal every T_s seconds and in total we have N such samples then T_s is known as the sampling period and $F_s = 1/T_s$ is the sampling frequency in Hertz (Hz) (samples per second). The nth sample occurs at time $t[n] = nT_s = n/F_s$. The cosine of sampled data can be written

$$x[n] = \cos(2\pi ft[n]) \qquad (1)$$

At a sampling frequency F_s the only *unique* frequencies are in

2

the range 0 to $(F_s/2)$Hz. Any frequencies outside this range become *aliases* of one of the unique frequencies.

For example, if we sample at 8Hz then a -6Hz signal becomes indistinguishable from a 2Hz signal. More generally, if f_0 is a unique frequency then its aliases have frequencies given by

$$f = f_0 + kF_s \tag{2}$$

where k is any positive or negative integer, eg. for $f_0 = 2$ and $F_s = 8$ the two lowest frequency aliases, given by $k = -1$ and $k = 1$, are -6Hz and 10Hz.

2   Sinusoidal models

If our time series has a periodic component in it we might think about modelling it with the equation

$$x[n] = R_0 + R cos(2\pi f t[n] + \Phi) + e[n] \tag{3}$$

where R_0 is the offset (eg. mean value of x[n]), R is the amplitude of the sine wave, f is the frequency and Φ is the phase. Because of the trig identity

$$\cos(A + B) = \cos A \cos B - \sin A \sin B \tag{4}$$

the model can be written in an alternative form

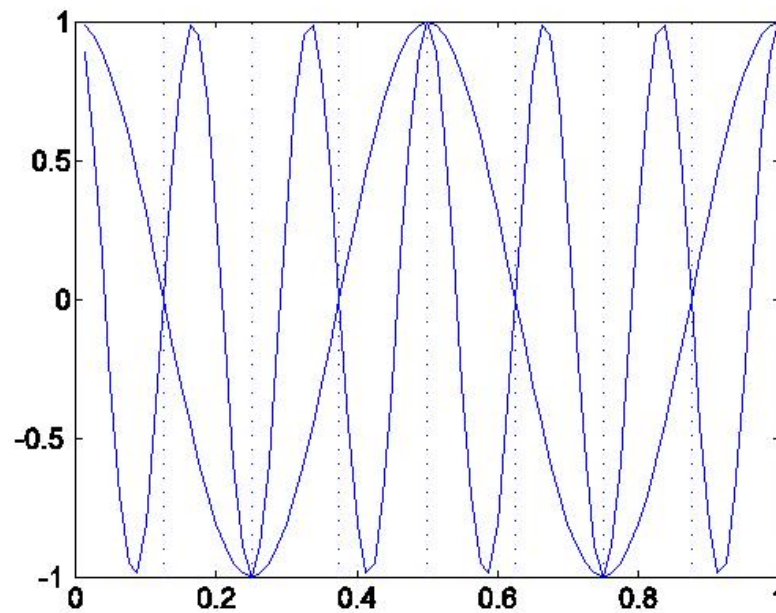$$x[n] = R_0 + a \cos(2\pi f t[n]) + b \sin(2\pi f t[n]) + e[n] \tag{5}$$

Figure 1: **Aliases** *The figure shows a 2Hz cosine wave and a -6Hz cosine wave as solid curves. At sampling times given by the dotted lines, which correspond to a sampling frequency of 8Hz, the $-6Hz$ signal is an alias of the 2Hz signal. Other aliases are given by equation 2.* 4

where $a = R\cos(\Phi)$ and $b = -R\sin(\Phi)$. This is the form we consider for subsequent analysis.

2.1 Fitting the model

If we let $\boldsymbol{x} = [x(1), x(2), ..., x(N)]^T$, $\boldsymbol{w} = [R_0, a, b]^T$, $\boldsymbol{e} = [e_1, e_2, ..., e_N]^T$ and

$$
\boldsymbol{A} = \begin{bmatrix} 1 & cos2\pi ft[1] & sin2\pi ft[1] \\ 1 & cos2\pi ft[2] & sin2\pi ft[2] \\ 1 & cos2\pi ft[3] & sin2\pi ft[3] \\ .. & .. & .. \\ 1 & cos2\pi ft[N] & sin2\pi ft[N] \end{bmatrix} \tag{6}
$$

then the model can be written in the matrix form

$$
\boldsymbol{x} = \boldsymbol{A}\boldsymbol{w} + \boldsymbol{e} \tag{7}
$$

which is in the standard form of a multivariate linear regression problem. The solution is therefore

$$
\boldsymbol{w} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{x} \tag{8}
$$

2.2 But sinewaves are orthogonal

We restrict ourselves to a frequency f_p which is an integer multiple of the *base frequency*

$$
f_p = pF_b \tag{9}
$$

where $p = 1..N/2$ and

$$f_b = \frac{F_s}{N} \tag{10}$$

eg. for $F_s = 100$ and $N = 100$ (1 seconds worth of data), $f_b = 1Hz$ and we can have f_p from 1Hz up to 50Hz[1]. The *orthogonality* of sinewaves is expressed in the following equations

$$\sum_{n=1}^{N} \cos 2\pi f_k t[n] = \sum_{n=1}^{N} \sin 2\pi f_k t[n] = 0 \tag{11}$$

$$\sum_{n=1}^{N} \cos 2\pi f_k t[n] \sin 2\pi f_l t[n] = 0 \tag{12}$$

$$\sum_{n=1}^{N} \cos 2\pi f_k t[n] \cos 2\pi f_l t[n] \quad = \quad \begin{matrix} 0 & k \neq l \\ N/2 & k = l \end{matrix} \tag{13}$$

$$\sum_{n=1}^{N} \sin 2\pi f_k t[n] \sin 2\pi f_l t[n] \quad = \quad \begin{matrix} 0 & k \neq l \\ N/2 & k = l \end{matrix} \tag{14}$$

The results depend on the fact that all frequencies that appear in the above sums are integer multiples of the base frequency; see figure 2.

---

[1]To keep things simple we don't allow f_p where $p = N/2$; if we did allow it we'd get N and 0 in equations 13 and 14 for the case $k = l$. Also we must have N even.
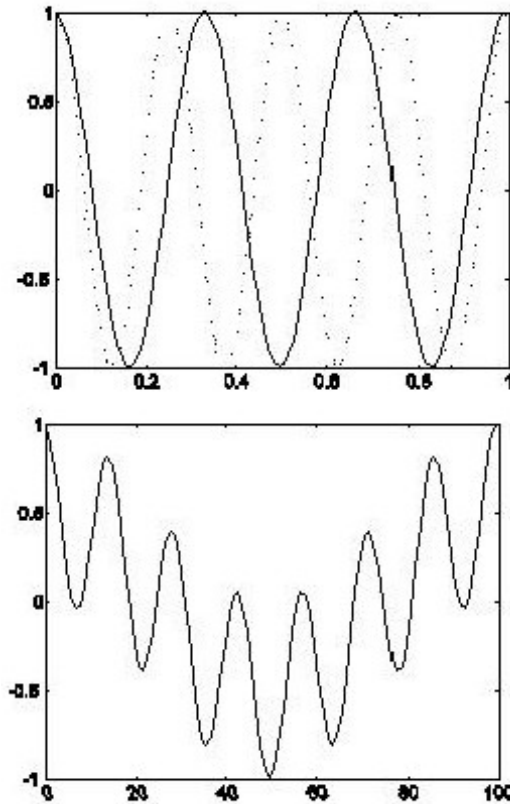
Figure 2: **Orthogonality of sinewaves** *Figure (top) shows* $\cos 2\pi 3 f_b t[n]$ *and* $\cos 2\pi 4 f_b t[n]$, *cosines which are 3 and 4 times the base frequency* $f_b = 1Hz$. *For any two integer multiples* k, l *we get* $\sum_{n=1}^{N} \cos 2\pi f_k t[n] \cos 2\pi f_l t[n] = 0$. *This can be seen from Figure (bottom) which shows the product* $cos2\pi 3 f_b t[n] cos 2\pi 4 f_b t[n]$. *Because of the trig identity* $cos A cos B = 0.5 cos(A+B) + 0.5 cos(A-B)$ *this looks like a 7Hz signal superimposed on a 1Hz signal. The sum of this signal over a whole number of cycles can be seen to be zero; because each cos term sums to zero. If, however, k or l are not integers the product does not sum to zero and the orthogonality breaks down.*

This property of sinewaves leads to the result

$$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{D} \tag{15}$$

where $\boldsymbol{D}$ is a diagonal matrix. The first entry is N (from the inner product of two columns of 1's of length N; the 1's are the coefficients of the constant term R_0) and all the other entries are $N/2$. A matrix $\boldsymbol{Q}$ for which

$$\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{D} \tag{16}$$

is said to be orthogonal. Therefore our $\boldsymbol{A}$ matrix is orthogonal. Hence

$$\boldsymbol{w} = \boldsymbol{D}^{-1}\boldsymbol{A}^T\boldsymbol{x} \tag{17}$$

which is simply a *projection* of the signal onto the basis matrix, with some pre-factor . Given that $\boldsymbol{w} = [a, b, R_0]^T$ we can see that, for example, a is computed by simply projecting the data onto the second column of the matrix $\boldsymbol{A}$, eg.

$$a = \frac{2}{N}\sum_{n=1}^{N}\cos(2\pi ft)x_t \tag{18}$$

Similarly,

$$b = \frac{2}{N}\sum_{n=1}^{N}\sin(2\pi ft)x_t \tag{19}$$

$$R_0 = \frac{1}{N}\sum_{n=1}^{N}x_t \tag{20}$$

We applied the simple sinusoidal model to a 'sunspot data set' as follows. We chose 60 samples between the years 1731 and 1790 (because there was a fairly steady mean level in this period). The sampling rate $F_s = 1\text{Year}$. This gives a base frequency of $f_b = 1/60$. We chose our frequency $f = pf_b$ with p=6; giving a complete cycle once every ten years. This gave rise to the following estimates; $R_0 = 53.64$, $a = 39.69$ and $b = -2.36$.

2.3 Fourier Series

We might consider that our signal consists of lots of periodic components in which case the *multiple sinusoidal model* would be more appropriate

$$x(t) = R_0 + \sum_{k=1}^{p} R_k \cos(2\pi f_k t + \Phi_k) + e_t \tag{21}$$

where there are p sinusoids with different frequencies and phases. In a discrete Fourier series there are $p = N/2$ such sinusoids having frequencies

$$f_k = \frac{kF_s}{N} \tag{22}$$

where $k = 1..N/2$ and F_s is the sampling frequency. Thus the frequencies range from F_s/N up to $F_s/2$. The Fourier series expansion
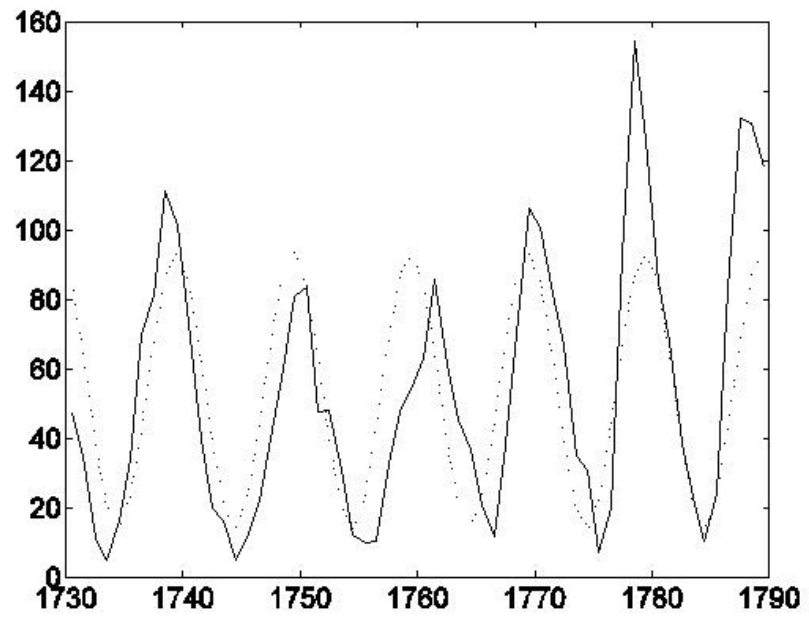
Figure 3: *Sunspot index (solid line) and prediction of it from a simple sinusoidal model (dotted line).*

of the signal $x(t)$ is

$$x(t) = R_0 + \sum_{k=1}^{N/2} R_k \cos(2\pi f_k t + \Phi_k) \qquad (23)$$

Notice that there is no noise term. Because of the trig identity

$$\cos(A + B) = \cos A \cos B - \sin A \sin B \qquad (24)$$

this can be written in the form

$$x(t) = a_0 + \sum_{k=1}^{N/2} a_k \cos(2\pi f_k t) + b_k \sin(2\pi f_k t) \qquad (25)$$

where $a_k = R_k \cos(\Phi_k)$ and $b_k = -R_k \sin(\Phi_k)$. Alternatively, we have $R_k^2 = a_k^2 + b_k^2$ and $\Phi = \tan^{-1}(b_k/a_k)$. Equivalently, we can write the nth sample as

$$x[n] = a_0 + \sum_{k=1}^{N/2} a_k \cos(2\pi f_k t[n]) + b_k \sin(2\pi f_k t[n]) \qquad (26)$$

where $t[n] = nT_s$.

The important things to note about the sinusoids in a Fourier series are (i) the frequencies are *equally* spread out, (ii) there are $N/2$ of them where N is the number of samples, (iii) Given F_s and N the frequencies are *fixed*. Also, note that in the Fourier series 'model' there is *no noise*.

The Fourier coefficients can be computed by a generalisation of the process used to compute the coefficients in the simple sinusoidal model.

$$a_k = \frac{2}{N} \sum_{n=1}^{N} \cos(2\pi f_k t[n]) x[n] \tag{27}$$

Similarly,

$$b_k = \frac{2}{N} \sum_{n=1}^{N} \sin(2\pi f_k t[n]) x[n] \tag{28}$$

$$a_0 = \frac{1}{N} \sum_{n=1}^{N} x[n] \tag{29}$$

These equations can be derived as follows. To find, for example, a_k, multiply both sides of equation 26 by $\cos(2\pi f_k t[n])$ and sum over n. Due to the orthogonality property of sinusoids (which still holds as all frequencies are integer multiples of a base frequency) all terms on the right go to zero except for the one involving a_k. This just leaves $a_k(N/2)$ on the right giving rise to the above formula.
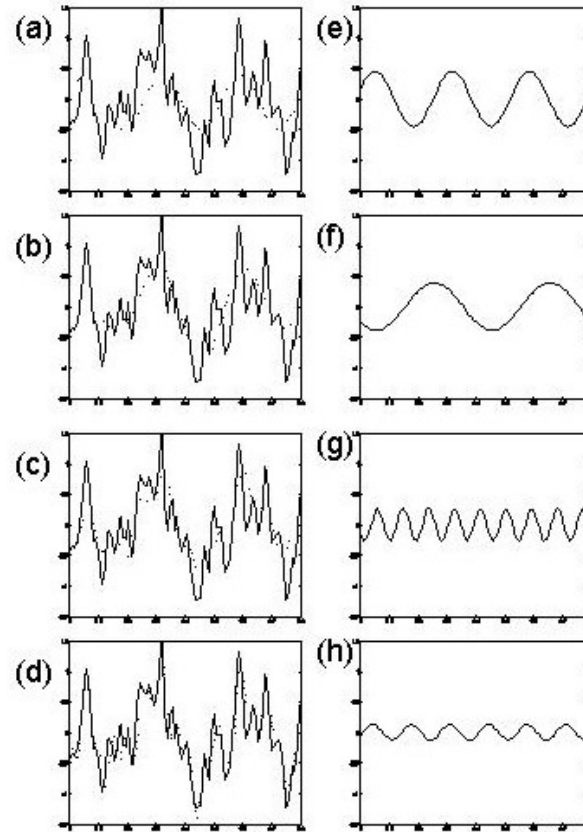
Figure 4: *Signal (solid line) and components of the Fourier series approximation* $\sum_{k=1}^{p} R_k cos(2\pi f_k + \Phi_k)$ *(dotted lines) with (a) $p = 1$, (b) $p = 2$, (c) $p = 3$ and (d) $p = 11$ where we have ordered the components according to amplitude. The corresponding individual terms are (e) $R^2 = 0.205, f = 3.75$ and $\Phi = 0.437$, (f) $R^2 = 0.151$, $f = 2.5$ and $\Phi = 0.743$, (g) $R^2 = 0.069$, $f = 11.25$ and $\Phi = 0.751$ and (h) $R^2 = 0.016$, $f = 7.5$ and $\Phi = -0.350$.*

13

3 Complex numbers

3.1 Power series

A function of a variable x can often be written in terms of a series of powers of x. For the sin function, for example, we have

$$\sin x = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + ... \tag{30}$$

We can find out what the appropriate coefficients are as follows. If we substitite $x = 0$ into the above equation we get $a_0 = 0$ since $sin0 = 0$ and all the other terms disappear. If we now *differentiate* both sides of the equation and substitute $x = 0$ we get $a_1 = 1$ (because $\cos 0 = 1 = a_1$). Differentiating twice and setting $x = 0$ gives $a_2 = 0$. Continuing this process gives

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + ... \tag{31}$$

Similarly, the series representations for $cosx$ and e^x can be found as

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + ... \tag{32}$$

and

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + ... \tag{33}$$

14

More generally, for a function $f(x)$ we get the general result

$$f(x) = f(0) + x f'(0) + \frac{x^2}{2!} f''(0) + \frac{x^3}{3!} f'''(0) + ... \quad (34)$$

where $f'(0)$, $f''(0)$ and $f'''(0)$ are the first, second and third derivatives of $f(x)$ evaluated at $x = 0$. This expansion is called a *Maclaurin series*.

So far, to calculate the coefficients in the series we have differentiated and substituted $x = 0$. If, instead, we substitute $x = a$ we get

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!} f''(a) + \frac{(x-a)^3}{3!} f'''(a) + ... \quad (35)$$

which is called a *Taylor series*.

3.2 Complex numbers

Very often, when we try to find the roots of an equation, we may end up with our solution being the square root of a negative number. For example, the quadratic equation

$$ax^2 + bx + c = 0 \quad (36)$$

has solutions

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (37)$$

If $b^2 - 4ac < 0$ we need the square root of a negative number. To handle this, mathematicians have defined the number

$$i = \sqrt{-1} \tag{38}$$

allowing all square roots of negative numbers to be defined in terms of i, eg. $\sqrt{-9} = \sqrt{9}\sqrt{-1} = 3i$. These numbers are called *imaginary numbers* to differentiate them from *real numbers*.

Finding the roots of equations, eg. the quadratic equation above, requires us to combine imaginary numbers and real numbers. These combinations are called *complex numbers*. For example, the equation

$$x^2 - 2x + 2 = 0 \tag{39}$$

has the solutions $x = 1+i$ and $x = 1-i$ which are complex numbers.

A complex number $z = a + bi$ has two components; a real part and an imaginary part which may be written

$$
\begin{aligned}
a &= Re\{z\} \\
b &= Im\{z\}
\end{aligned}
\tag{40}
$$

The *absolute value* of a complex number is

$$R = Abs\{z\} = \sqrt{a^2 + b^2} \tag{41}$$

and the *argument* is

$$\theta = Arg\{z\} = \tan^{-1}\left(\frac{b}{a}\right) \tag{42}$$

The two numbers $z = a + bi$ and $z^* = a - bi$ are known as *complex conjugates*; one is the complex conjugate of the other. When multiplied together they form a real number. The roots of equations often come in complex conjugate pairs.

3.3 Complex vectors

The transpose, x^T, becomes a 'Hermitian' transpose, x^H, which is the usual transpose but the elements become conjugates. This means that the length of a vector (squared) is now $x^H x$ instead of $x^T x$.

3.4 Complex exponentials

If we take the exponential function of an imaginary number and write it out as a series expansion, we get

$$e^{i\theta} \quad = \quad 1 + \frac{i\theta}{1!} + \frac{i^2\theta^2}{2!} + \frac{i^3\theta^3}{3!} + \dots \tag{43}$$

By noting that $i^2 = -1$ and $i^3 = i^2 i = -i$ and similarly for higher powers of i we get

$$e^{i\theta} \quad = \quad \left[1 - \frac{\theta^2}{2!} + \dots \right] + i \left[\frac{\theta}{1!} - \frac{\theta^3}{3!} + \dots \right] \tag{44}$$

Comparing to the earlier expansions of $\cos\theta$ and $\sin\theta$ we can see that

$$e^{i\theta} = \cos\theta + i\sin\theta \tag{45}$$

which is known as *Euler's formula*. Similar expansions for $e^{-i\theta}$ give the identity

$$e^{-i\theta} = \cos\theta - i\sin\theta \tag{46}$$

We can now express the sine and cosine functions in terms of complex exponentials

$$\begin{aligned} \cos\theta &= \frac{e^{i\theta} + e^{-i\theta}}{2} \\ \sin\theta &= \frac{e^{i\theta} - e^{-i\theta}}{2i} \end{aligned} \tag{47}$$

3.5 DeMoivre's theorem

By using the fact that

$$e^{i\theta} e^{i\theta} = e^{i\theta + i\theta} \tag{48}$$

(a property of the exponential function and exponents in general eg. $5^3 5^3 = 5^6$) or more generally

$$(e^{i\theta})^k = e^{ik\theta} \tag{49}$$

we can write

$$(cos\theta + i\sin\theta)^k = cosk\theta + isink\theta \tag{50}$$

which is known as *DeMoivre's theorem.*

3.6 Argand diagrams

Any complex number can be represented as a complex exponential

$$a + bi = Re^{i\theta} = R(cos\theta + i \sin \theta) \tag{51}$$

and drawn on an *Argand diagram.* Multiplication of complex numbers is equivalent to rotation in the complex plane (due to DeMoivre's Theorem).

$$(a + bi)^2 = R^2 e^{i2\theta} = R^2(cos2\theta + i \sin 2\theta) \tag{52}$$

4 Discrete Fourier Transform

Fourier series can be expressed in terms of complex exponentials. This representation leads to an efficient method for computing the coefficients. We can write the cosine terms as complex exponentials

$$a_k \cos(2\pi f_k t[n]) = a_k \frac{\exp(i2\pi f_k t[n]) + \exp(-i2\pi f_k t[n])}{2} \tag{53}$$

where $i^2 = -1$. Picture this as the addition of two vectors; one above the real axis and one below. Together they make a vector on the real axis which is then halved.

We can also write the sine terms as

$$b_k \sin(2\pi f_k t[n]) = b_k \frac{\exp(i2\pi f_k t[n]) - \exp(-i2\pi f_k t[n])}{2i} \qquad (54)$$

Picture this as one vector above the real axis minus another vector below the real axis. This results in a purely imaginary (and positive) vector. The result is halved and then multiplied by the vector $\exp(3\pi/2)$ ($-i$, from multiplying top and bottom by i) which provides a rotation to the real axis.

Adding them (and moving i to the numerator by multiplying b_k top and bottom by i) gives

$$\frac{1}{2}(a_k - b_k i)\exp(i2\pi f_k t[n]) + \frac{1}{2}(a_k + b_k i)\exp(-i2\pi f_k t[n]) \qquad (55)$$

Note that a single term at frequency k has split into a complex combination (the coefficients are complex numbers) of a positive frequency term and a *negative frequency* term. Substituting the above result into equation 26 and noting that $f_k t[n] = kn/N$ we get

$$x[n] = a_0 + \frac{1}{2}\sum_{k=1}^{N/2}(a_k - b_k i)\exp(i2\pi kn/N) + \frac{1}{2}\sum_{k=1}^{N/2}(a_k + b_k i)\exp(-i2\pi kn/N)$$

$$(56)$$

If we now let

$$\tilde{X}(k) = \frac{N}{2}(a_k - b_k i) \qquad (57)$$

and note that for real signals $\tilde{X}(-k) = \tilde{X}^*(k)$ (negative frequencies are reflections across the real plane, ie. conjugates) then the $(a_k + b_k i)$ terms are equivalent to $\tilde{X}(-k)$. Hence

$$x[n] = a_0 + \frac{1}{2N} \sum_{k=1}^{N/2} \tilde{X}(k) \exp(i2\pi kn/N) + \frac{1}{2N} \sum_{k=1}^{N/2} \tilde{X}(k) \exp(-i2\pi kn/N) \tag{58}$$

Now, because $\tilde{X}(N-k) = \tilde{X}(-k)$ (this can be shown by considering the Fourier transform of a signal $x[n]$ and using the decomposition $\exp(-i2\pi(N-k)n/N) = \exp(-i2\pi N/N) \exp(i2\pi kn/N)$ where the first term on the right is unity) we can write the second summation as

$$x[n] = a_0 + \frac{1}{2N} \sum_{k=1}^{N/2} \tilde{X}(k) \exp(i2\pi kn/N) + \frac{1}{2N} \sum_{k=N/2}^{N-1} \tilde{X}(k) \exp(-i2\pi(N-k)n/N) \tag{59}$$

Using the same exponential decomposition allows us to write

$$x[n] = a_0 + \frac{1}{N} \sum_{k=1}^{N-1} \tilde{X}(k) \exp(i2\pi kn/N) \tag{60}$$

If we now let $X(k+1) = \tilde{X}(k)$ then we can absorb the constant a_0 into the sum giving

$$x[n] = \frac{1}{N} \sum_{k=1}^{N} X(k) \exp(i2\pi(k-1)n/N) \tag{61}$$

which is known as the *Inverse Discrete Fourier Transform* (IDFT). The terms $X(k)$ are the *complex valued Fourier coefficients.* We have the relations

$$
\begin{align}
a_0 &= Re\{X(1)\} \tag{62} \\
a_k &= \frac{2}{N}Re\{X(k+1)\} \\
b_k &= \frac{-2}{N}Im\{X(k+1)\}
\end{align}
$$

The complex valued Fourier coefficients can be computed by first noting the orthogonality relations

$$
\sum_{n=1}^{N} \exp(i2\pi(k-1)n/N) = \begin{array}{ll} N & k=1, \pm(N+1), \pm(N+2) \\ 0 & otherwise \end{array}
$$

If we now multiply equation 61 by $\exp(-i2\pi ln/N)$, sum from 1 to N and re-arrange we get

$$
\begin{align}
X(k) &= \sum_{n=1}^{N} x(n)\exp(-i2\pi(k-1)n/N) \tag{63} \\
&= DFT(x)
\end{align}
$$

which is the Discrete Fourier Transform (DFT).

4.1 Power Spectral Density

The power in a signal is given by

$$P_x = \sum_{n=1}^{N} |x[n]|^2 \tag{64}$$

We now derive an expression for P_x in terms of the Fourier coefficients. If we note that $|x[n]|$ can also be written in its conjugate form (the conjugate form has the same magnitude; the phase is different but this does'nt matter as we're only interested in magnitude)

$$|x[n]| = \frac{1}{N} \sum_{k=1}^{N} X^*(k) \exp(-i2\pi(k-1)n/N) \tag{65}$$

then we can write the power as

$$P_x = \sum_{n=1}^{N} |x[n] \frac{1}{N} \sum_{k=1}^{N} X^*(k) \exp(-i2\pi(k-1)n/N)| \tag{66}$$

If we now change the order of the summations we get

$$P_x = \frac{1}{N} \sum_{k=1}^{N} |X^*(k) \sum_{n=1}^{N} x(n) \exp(-i2\pi(k-1)n/N)| \tag{67}$$

where the sum on the right is now equivalent to $X(k)$. Hence

$$P_x = \frac{1}{N} \sum_{k=1}^{N} |X(k)|^2 \tag{68}$$

23

We therefore have an equivalence between the power in the time domain and the power in the frequency domain which is known as *Parseval's relation.* The quantity

$$P_x(k) = |X(k)|^2 \tag{69}$$

is known as the *Power Spectral Density* (PSD).

4.2 Filtering

The filtering process

$$x[n] = \sum_{l=-\infty}^{\infty} x_1(l)x_2(n-l) \tag{70}$$

is also known as *convolution*

$$x[n] = x_1(n) * x_2(n) \tag{71}$$

We will now see how it is related to frequency domain operations. If we let $w = 2\pi(k-1)/N$, multiply both sides of the above equation by $\exp(-iwn)$ and sum over n the left hand side becomes the Fourier transform

$$X(w) = \sum_{n=-\infty}^{\infty} x[n]\exp(-iwn) \tag{72}$$

and the right hand side (RHS) is

$$\sum_{n=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} x_1(l)x_2(n-l)\exp(-iwn) \tag{73}$$

Now, we can re-write the exponential term as follows

$$\exp(-iwn) = \exp(-iw(n-l))\exp(-iwl) \tag{74}$$

Letting $n' = n - l$, we can write the RHS as

$$\sum_{l=-\infty}^{\infty} x_1(l)\exp(-iwl) \sum_{n'=-\infty}^{\infty} x_2(n')\exp(-iwn') = X_1(w)X_2(w) \tag{75}$$

Hence, the filtering operation is equivalent to

$$X(w) = X_1(w)X_2(w) \tag{76}$$

which means that convolution in the time domain is equivalent to multiplication in the frequency domain. This is known as the *convolution theorem*.

4.3   Autocovariance and Power Spectral Density

The autocovariance of a signal is given by

$$\sigma_{xx}(n) = \sum_{l=-\infty}^{\infty} x(l)x(l-n) \tag{77}$$

Using the same method that we used to prove the convolution theorem, but noting that the term on the right is $x(l-n)$ not $x(n-l)$ we can show that the RHS is equivalent to

$$X(w)X(-w) = |X(w)|^2 \tag{78}$$

which is the Power Spectral Density, $P_x(w)$. Combining this with what we get for the left hand side gives

$$P_x(w) = \sum_{n=-\infty}^{\infty} \sigma_{xx}(n) \exp(-iwn) \tag{79}$$

which means that the PSD is the Fourier Transform of the autocovariance. This is known as the *Wiener-Khintchine Theorem.* This is an important result. It means that the PSD can be estimated from the autocovariance and vice-versa. It also means that the PSD and the autocovariance contain the same information about the signal.

It is also worth noting that since both contain no information about the phase of a signal then the signal cannot be uniquely constructed from either. To do this we need to know the PSD *and* the *Phase spectrum* which is given by

$$\Phi(k) = \tan^{-1}(\frac{b_k}{a_k}) \tag{80}$$

where b_k and a_k are the real Fourier coefficients.

We also note that the Fourier transform of a symmetric function is real. This is because symmetric functions can be represented entirely by cosines, which are themselves symmetric; the sinewaves, which constitute the complex component of a Fourier series, are no longer necessary. Therefore, because the autocovariance is symmetric the PSD is real.

4.4 The Periodogram

The periodogram of a signal x_t is a plot of the normalised power in the kth harmonic versus the frequency, f_k of the kth harmonic. It is calculated as

$$
\begin{aligned}
I(f_k) &= \frac{T_s}{N}(a_k^2 + b_k^2) \\
&= \frac{T_s}{N}|\sum_{n=1}^{N} x(n)\exp(-i2\pi(k-1)n/N)|^2
\end{aligned}
\tag{81}
$$

where a_k and b_k are the Fourier coefficients.

The periodogram is a low bias (actually unbiased) but high variance [2] estimate of the power at a given frequency. This is therefore a problem if the number of data points is small; the estimated spectrum will be very spiky.

To overcome this, a number of algorithms exist to smooth the periodogram ie. to reduce the variance. The Bartlett method, for example, takes an N-point sequence and subdivides it into K nonoverlapping segments and calculates $I(f_k)$ for each. The final periodogram is just the average over the K estimates. This results in a reduction in variance by a factor K at the cost of reduced spectral resolution (by a factor K).

---

[2] It is an *inconsistent* estimator, because the variance does'nt reduce to zero as the number of samples tends to infinity.

4.5 Modified periodograms

Other methods modify the periodogram by using a time domain window, h, also known as a data taper, such that the modified periodogram is given by

$$
\begin{aligned}
I(f_k) &= \frac{T_s}{N}(a_k^2 + b_k^2) \\
&= \frac{T_s}{N}|\sum_{n=1}^{N} h(n)x(n)\exp(-i2\pi(k-1)n/N)|^2
\end{aligned}
\tag{82}
$$

For example, the matlab `psd` function chooses h to be a Hanning window.

Welch's method (`pwelch.m` in matlab) takes overlapping segments, and averages modified periodograms. unreliable.

4.6 Multi-tapering

This uses multiple data tapers. The tapers are orthogonal to each other. Given a length N sequence

- Choose the desired spectral resolution, W, eg. $\pm 2Hz$

- Get corresponding Slepian sequences, h_i, $i = 1..2NW - 1$. These provide an eigenbasis in local frequency space ($\pm W$) for finite length data sequences.
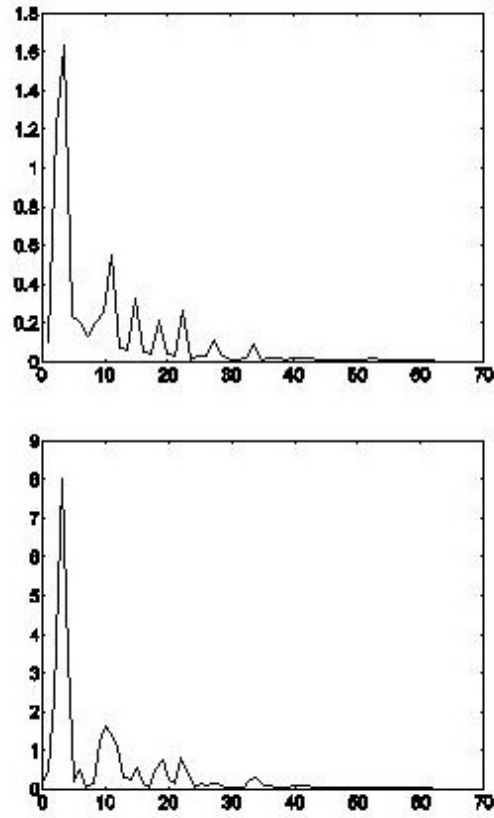
- Sum modified periodograms

Figure 5: *Periodogram (top) and Welch's modified Periodogram (bottom).*

This is implemented in eg. matlab's `pmtm.m` function. For theory see [6], and for application to brain imaging data see [5].

See eg. `sunspot_spectra.m`.

4.7 MEG data

Bauer et al. [1] estimate spectral density of MEG data during a delayed match-to-sample task. For each subject they computed spectra during periods when non-matching stimuli were correctly rejected. Spectra in range 40-180Hz, were computed using a multitaper method with 200ms windows and a spectral resolution of $W = 10Hz$.

They then computed z-scores for each time frequency bin by comparing with a baseline period. These z-scores were then averaged over subject.

5 Multiple time series

5.1 Cross-correlation

Given *two* time series x_t and y_t we can delay x_t by T samples and then calculate the *cross-covariance* between the pair of signals. That is

$$\sigma_{xy}(T) = \frac{1}{N-1} \sum_{t=1}^{N} (x_{t-T} - \mu_x)(y_t - \mu_y) \tag{83}$$

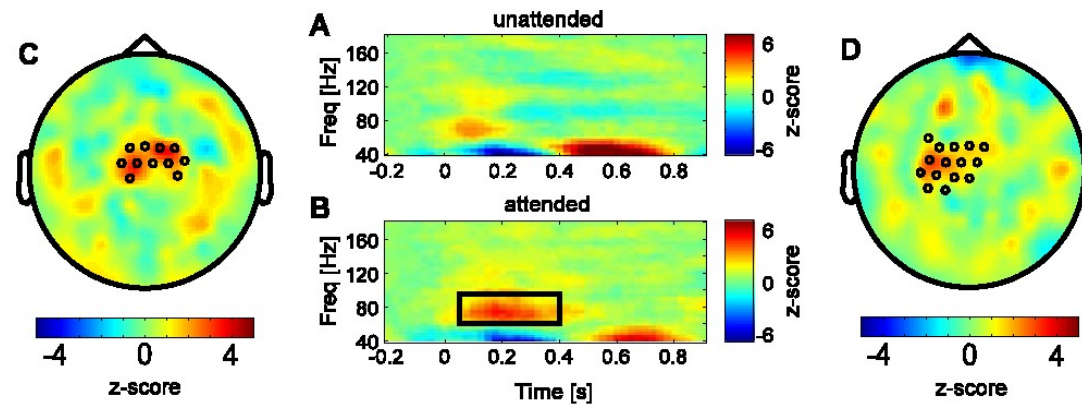Figure 6: Delayed match-to-sample task.

Figure 7: Spectra for attended and unattended side, averaged over electrode set shown in (C) for left finger stimulation and (B) right finger stimulation.

where μ_x and μ_y are the means of each time series and there are N samples in each. The function $\sigma_{xy}(T)$ is the *cross-covariance* function. The *cross-correlation* is a normalised version

$$r_{xy}(T) = \frac{\sigma_{xy}(T)}{\sqrt{\sigma_{xx}(0)\sigma_{yy}(0)}} \tag{84}$$

where we note that $\sigma_{xx}(0) = \sigma_x^2$ and $\sigma_{yy}(0) = \sigma_y^2$ are the variances of each signal. Note that

$$r_{xy}(0) = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{85}$$

which is the correlation between the two variables. Therefore unlike the autocorrelation, r_{xy} is not, generally, equal to 1.

The cross-correlation is a normalised cross-covariance which, assuming zero mean signals, is given by

$$\sigma_{xy}(T) = <x_{t-T}y_t> \tag{86}$$

and for negative lags

$$\sigma_{xy}(-T) = <x_{t+T}y_t> \tag{87}$$

Subtracting T from the time index now gives

$$\sigma_{xy}(-T) = <x_t y_{t-T}> \tag{88}$$

which is different to $\sigma_{xy}(T)$. To see this more clearly we can subtract T once more from the time index to give

$$\sigma_{xy}(-T) = <x_{t-T}y_{t-2T}> \qquad (89)$$

Hence, the cross-covariance, and therefore the cross-correlation, is an *asymmetric* function (the autocorrelation is symettric).

To summarise: moving signal A right (forward in time) and multiplying with signal B is not the same as moving signal A left and multiplying with signal B; unless signal A equals signal B.

5.2   Cross Spectral Density

Just as the Power Spectral Density (PSD) is the Fourier transform of the auto-covariance function we may define the Cross Spectral Density (CSD) as the Fourier transform of the cross-covariance function

$$P_{12}(w) = \sum_{n=-\infty}^{\infty} \sigma_{x_1 x_2}(n) \exp(-iwn) \qquad (90)$$

Note that if $x_1 = x_2$, the CSD reduces to the PSD. Now, the cross-covariance of a signal is given by

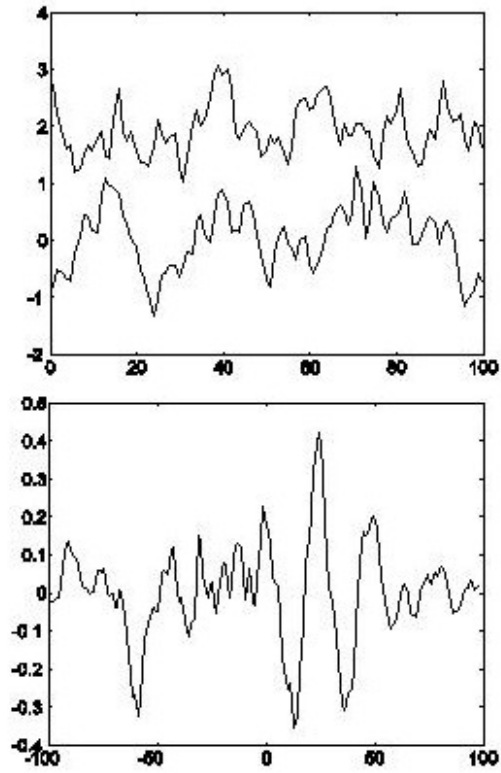$$\sigma_{x_1 x_2}(n) = \sum_{l=-\infty}^{\infty} x_1(l)x_2(l-n) \qquad (91)$$

Figure 8: *Top plot: Signals x_t (top curve) and y_t (bottom curve). Bottom plot: Cross-correlation function $r_{xy}(T)$. A lag of T denotes the top series, x, lagging the bottom series, y. Notice the big positive correlation at a lag of 25.*

Substituting this into the earlier expression gives

$$P_{12}(w) = \sum_{n=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} x_1(l) x_2(l-n) \exp(-iwn) \qquad (92)$$

By noting that

$$\exp(-iwn) = \exp(-iwl) \exp(iwk) \qquad (93)$$

where $k = l - n$ we can see that the CSD splits into the product of two integrals

$$P_{12}(w) = X_1(w) X_2(-w) \qquad (94)$$

where

$$X_1(w) = \sum_{l=-\infty}^{\infty} x_1(l) \exp(-iwl) \qquad (95)$$

$$X_2(-w) = \sum_{k=-\infty}^{\infty} x_2(k) \exp(+iwk)$$

For real signals $X_2^*(w) = X_2(-w)$ where * denotes the complex conjugate. Hence, the cross spectral density is given by

$$P_{12}(w) = X_1(w) X_2^*(w) \qquad (96)$$

This means that the CSD can be evaluated in one of two ways (i) by first estimating the cross-covariance and Fourier transforming or

(ii) by taking the Fourier transforms of each signal and multiplying (after taking the conjugate of one of them). A number of algorithms exist which enhance the spectral estimation ability of each method. These algorithms are basically extensions of the algorithms for PSD estimation, for example, for type (i) methods we can perform Blackman-Tukey windowing of the cross-covariance function and for type (ii) methods we can employ Welch's algorithm for averaging modified periodograms before multiplying the transforms. See Carter [2] for more details.

The CSD is complex because the cross-covariance is asymmetric (the PSD is real because the auto-covariance is symmetric; in this special case the Fourier transorm reduces to a cosine transform).

5.3   PSD matrix

The frequency domain characteristics of a multivariate time-series (eg. two or more) may be summarised by the power spectral density *matrix* (Marple, 1987[4]; page 387). For d time series

$$\boldsymbol{P}(f) = \begin{pmatrix} P_{11}(f) & P_{12}(f) & \cdots & P_{1d}(f) \\ P_{12}(f) & P_{22}(f) & \cdots & P_{2d}(f) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ P_{1d}(f) & P_{2d}(f) & \cdots & P_{dd}(f) \end{pmatrix} \tag{97}$$

where the diagonal elements contain the spectra of individual channels and the off-diagonal elements contain the cross-spectra.

The matrix is called a *Hermitian matrix* because the elements are complex numbers.

5.4 Coherence and Phase

The *complex coherence function* is given by (Marple 1987; p. 390)

$$r_{ij}(f) = \frac{P_{ij}(f)}{\sqrt{P_{ii}(f)}\sqrt{P_{jj}(f)}} \tag{98}$$

The coherence, or *mean squared coherence* (MSC), between two channels is given by

$$r_{ij}^2(f) = \mid r_{ij}(f) \mid^2 \tag{99}$$

The phase spectrum, between two channels is given by

$$\theta_{ij}(f) = tan^{-1}\left[\frac{Im(r_{ij}(f))}{Re(r_{ij}(f))}\right] \tag{100}$$

The MSC measures the linear correlation between two time series at each frequency and is directly analagous to the squared correlation coefficient in linear regression. As such the MSC is intimately related to *linear filtering*, where one signal is viewed as a filtered version of the other. This can be interpreted as a linear regression at each frequency. The optimal regression coefficient, or linear filter, is given by

$$H(f) = \frac{P_{xy}(f)}{P_{xx}(f)} \tag{101}$$

This is analagous to the expression for the regression coefficient $a = \sigma_{xy}/\sigma_{xx}$ (see first lecture). The MSC is related to the optimal filter as follows

$$r_{xy}^2(f) = |H(f)|^2 \frac{P_{xx}(f)}{P_{yy}(f)} \qquad (102)$$

which is analagous to the equivalent expression in linear regression $r^2 = a^2(\sigma_{xx}/\sigma_{yy})$.

At a given frequency, if the phase of one signal is *fixed* relative to the other, then the signals can have a high coherence at that frequency. This holds even if one signal is entirely out of phase with the other (note that this is different from adding up signals which are out of phase; the signals cancel out. We are talking about the coherence *between* the signals).

At a given frequency, if the phase of one signal changes relative to the other then the signals will not be coherent at that frequency. The time over which the phase relationship is constant is known as the *coherence time*. See [**?**], for an example.

5.5 Welch's method for estimating coherence

Algorithms based on Welch's method (such as the cohere function in the matlab system identification toolbox) are widely used [2] [7]. The signal is split up into a number of segments, N, each of length T and the segments may be overlapping. The complex coherence

estimate is then given as

$$\hat{r}_{ij}(f) = \frac{\sum_{n=1}^{N} X_i^n(f)(X_j^n(f))^*}{\sqrt{\sum_{n=1}^{N} X_i^n(f)^2}\sqrt{\sum_{n=1}^{N} X_j^n(f)^2}} \qquad (103)$$

where n sums over the data segments. This equation is exactly the same form as for estimating correlation coefficients (see chapter 1). Note that if we have only $N = 1$ data segment then the estimate of coherence will be 1 regardless of what the true value is (this would be like regression with a single data point). Therefore, we need a number of segments.

Note that this only applies to Welch-type algorithms which compute the CSD from a product of Fourier transforms. We can trade-off good spectral resolution (requiring large T) with low-variance estimates of coherence (requiring large N and therefore small T). We can also overlap segments.

6 Source reconstruction of MEG Gamma activity

Specify frequency-domain model, and reconstruct using maximum likelihood estimator

$$
\begin{aligned}
y_f &= Lr_f + e_f \qquad (104) \\
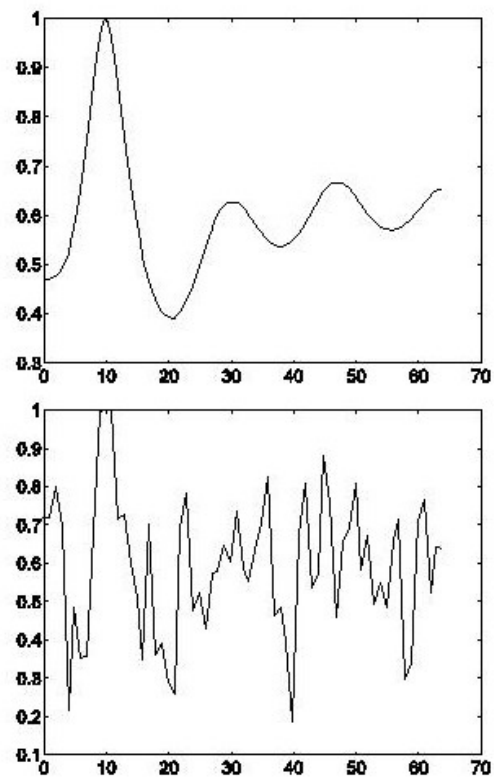\hat{r}_f &= (L^T P_f^{-1} L)^{-1} L^T P_f^{-1} y_f
\end{aligned}
$$

Figure 9: *Coherence estimates using (bottom) Welch's method and (top) MAR model*

41

where L is the lead-field and P_f is the PSD matrix over sensors computed for time-period of interest.

This can be augmented to a Bayesian estimator in the usual way. This leads to the Dynamic Imaging of Coherent sources (DICs) algorithm [3].

We can write the reconstructed activity as

$$\hat{r}_f \;=\; w_{rf} y_f \tag{105}$$

The power is then given by

$$p_{rf} = w_{rf}^* P_f w_{rf} \tag{106}$$

7 Further topics

- AR/MAR - parametric

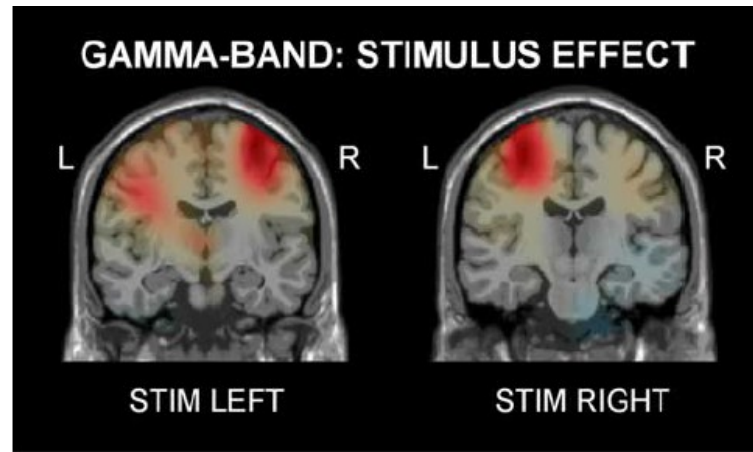- Subspace methods - small number of modes

- Wavelets - nonstationarity

Figure 10: Source analysis of gamma power (60-95Hz) versus baseline

References

[1] M. Bauer, R. Oostenveld, M. Peeters, and P. Fries. Tactile spatial attention enhances gamma-band activity in somatosensory cortex and reduces low-frequency activity in parieto-occipital areas. *The Journal of Neuroscience*, 26(2):490–501, 2006.

[2] G. Clifford Carter. Coherence and time delay estimation. *Proceedings of the IEEE*, 75(2):236–255, 1987.

[3] J. Gross, J. Kujala, M. Hammalainen, L. Timmerman, A. Schnitzler, and R. Salmelin. Dynamic imaging of coherent sources: studying neural interactions in the brain. *Proceedings of the National Academy of Sciences*, 98(2):694–699, 2001.

[4] S.L. Marple. *Digital spectral analysis with applications*. Prentice-Hall, 1987.

[5] P.P. Mitra and B. Pesaran. Analysis of dynamic brain imaging data. *Biophysical Journal*, 76:691–708, 1999.

[6] D.B. Percival and A.T. Walden. *Spectral analysis for physical applications: multitaper and conventional univariate techniques*. Cambridge University Press, UK., 1993.

[7] J.C. Shaw and D. Simpson. "eeg coherence: Caution and cognition". *British Psychological Society Quaterly*, 30/31, 1997.