CHAPTER 1

Problem Description

1.1. A Presentation

This paper seeks to explain statistical methods by which Electroencephalography (EEG) data from mismatched speech patterns may be interpreted.

EEG signals are recorded as potential differences between surface electrodes on the scalp. Hence the signal obtained depends upon the positions of the individual electrodes. A common method to determine these positions is the international 10–20 system of electrode placement. This involves placing the electrodes at 10% to 20% of the distance between anatomical sites on the head. Each electrode site has a letter identifying its sub-cranial lobe (i.e. FP-Frontopolar lobe, F-Frontal lobe, T-Temporal lobe, C-Central lobe, P-Parietal lobe, O-Occipital lobe) and a number or another letter identifying its hemispherical location, odd numbers referring to the left hemisphere, even numbers the right. Figure 1.1 is a representation of a plan view of the scalp and the approximate electrode positions.

The EEG recorded in response to a defined auditory stimulus is termed an evoked related potential (ERP). However the low amplitude of the ERP makes it difficult to differentiate from background brain activity (EEG). However as it is assumed that repetitive applications of such a stimulus will most likely activate similar pathways within the brain, several such ERP responses can be averaged to improve the signal to noise ratio, where the background EEG is the "noise". The waveform that results from averaging a number of individual EEG samples is commonly referred to as an average-evoked potential (AEP) [**Callaway, 1975**].

1.1.1. Definitions. *As we will frequently be relying upon acronyms and abbreviations we present them here.*

Definitions for acronyms and abbreviations used in this paper

ADSL	=	*asymmetric digitial subscriber line*
AEP	=	*average-evoked potential*
AIC	=	*Akaike's information criterion*
BIC	=	*Bayesian information criterion*
CWT	=	*continuous wavelet transform*
DIC	=	*deviance information criterion*
DWT	=	*discrete wavelet transform*
ECG	=	*electrocardiography*
EEG	=	*electroencephalography*
ERP	=	*evoked related potential*
GP	=	*spatial gaussian process*
ML	=	*maximum likelihood*
MMN	=	*mismatch negativity*
MRA	=	*multiresolution analysis*
REML	=	*restricted maximum likelihood*
STFT	=	*short-time Fourier transform*

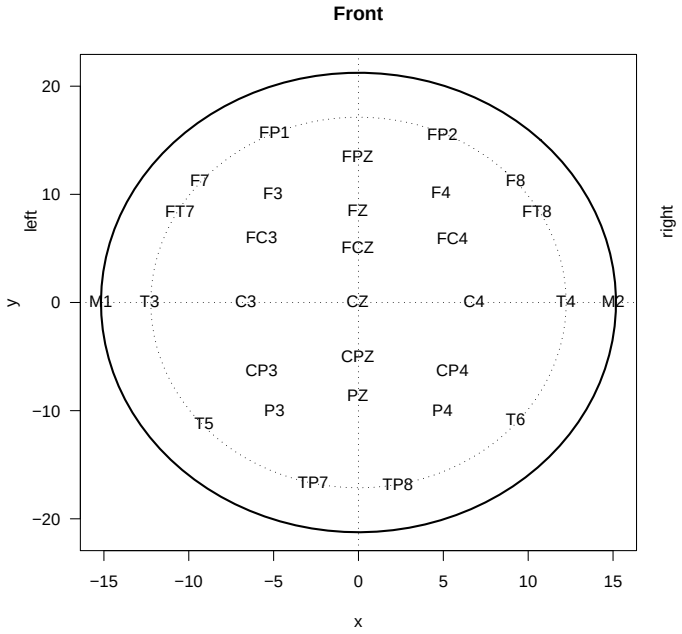
1.1.2. Our Situation and Aim. *The data used were collected in a trial to establish the zone of the cortex where different speech mismatches are recognised* [**Pettigrew, 2004**]. *The original trial data consisted of 10 subjects who were tested for 12 patterns of mismatched syllables (termed Mismatch Negativity, MMN), we however only had data from two subjects. MMN is elicited by a discriminable change in a repetitive background of auditory stimulation. The data used was the response of the brain recorded using the 10–20 international electrode placement system.*

With the aim to develop a model in the plane, each electrode position was given an estimated x and y co-ordinate (Table 1.1). The data used was from 2 subjects undertaking

TABLE 1.1. Electrode Coordinates

Electrode	x	y
FP1	-4.95	*15.73*
FP2	*4.94*	*15.59*
.	.	.
.	.	.
.	.	.
TP8	*2.32*	-16.84

FIGURE 1.1. Position of Electrodes for the 10–20 system



MMN trials, where the recognition of the change in a defined syllable to another: i.e. de → day, was the main focus.

The syllable changes available are: da, day, ge, and gay; therefore the change from any syllable to another could be used for such analysis. In this presentation we shall use:

$$(a) \qquad day \rightarrow gay$$

$$(b) \qquad de \rightarrow ge$$

$$(c) \qquad de \rightleftharpoons day.$$

Where day → gay, a real word deviant, was chosen to compare with de → ge, a non real word deviant, with the aim of attempting to determine which of these "fine acoustic speech contrasts" may elicit a larger MMN response. Similarly de ⇌ day was included to discover if any difference in the latency of the MMN response depended upon the direction of transition, i.e. de → day as opposed to day → de.

A basic problem inherent in MMN research is the very low signal to noise ratios as experienced when using the usual EEG recording techniques. This MMN component, whose amplitude may typically be in the range of low μV, is embedded in the EEG activity that may have peak events of over one hundred μV.

1.2. An Initial Approach

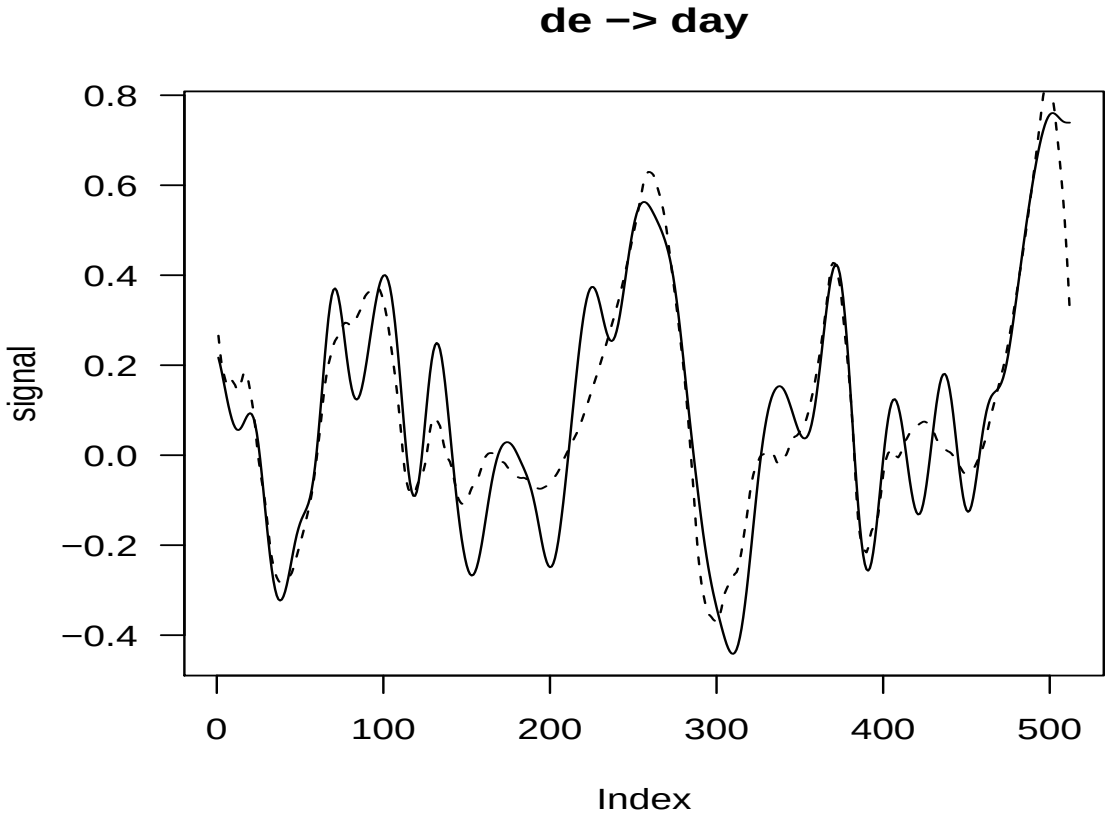
For some years now several studies have shown the usefulness of the technique of wavelets in biomedical signal processing, across a wide range of applications for wavelet transforms. They can be seen in different fields from signal processing to biology and we can be sure the list continues to grow. A commonly mentioned example is in the FBI fingerprint compression standard. Here wavelet transforms are used to compress the finger print pictures for data storage. The previously used methodology did not perform well at high compression ratios, it made it quite difficult to follow the ridge lines in the fingerprint's reconstruction. This did not happen with the Wavelet transform due to its reconstruction property.

*The wavelet transform has been found to have substantial relevance in biomedical engineering. One major problem with biomedical signals is their variability, and often one has little knowledge of what is really a relevant signal and at which scale it is located. However of considerable importance within biomedical signals is that the information of interest is often a mixture of specifics that might be localised in time or spatially, e.g. the spikes and changes in EEG, and some the are more widely spread e.g. slow eye movement during light sleep. So we require analysis methods that can handle such events at opposite end points in regards of their time frequency localisation. Thus, the range of applications of the wavelet transform with its multiresolution analysis has been quite large. For a good overview of some major biomedical applications of wavelets refer to Aldroubi, (**1996**) where this text refers to the theoretical and practical foundations of wavelet methods, together with applications to biomedical signal processing and mathematical models in biology.*

In our situation the approach is based on wavelets to detect the presence of MMN activity in each of the single event related potentials. In essence the problem is based upon classification, using a neural network, of wavelet features representing the time-window of the MMN component.

Developed algorithms in R [**R Development team, 2008**] *were implemented, being the* **waveslim, 2007** *and* **wavesthresh, 2008** *packages. From each electrode for a given syllable change it was possible to plot the signal, deconstruct it using a function that performs the decomposition stage of Mallat's pyramid algorithm* [**Mallat, 1989**]*, i.e., the discrete wavelet transform. Then simply reconstruct using the the discrete inverse wavelet transform. We could then simply plot such for quick view of our accuracy, see Figure 1.2. Here the vertical axis represents the amplitude as recorded at the electrode over time.*

FIGURE 1.2. Plot of signal (solid line) with reconstructed from thresholded wavelet coefficients (dotted line)



This was done for each electrode, highlighting the approximation of the reconstructed signal. And similarly we could extract the wavelet coefficients at a set frequency level. Hence by determining the frequency level, i.e. 1, 2 , 3 and 4, we could build an energy array table where it was possible to examine the energy in the signal, which is equal to the sum of the energy coefficients squared[1]*, at each level respectively.*

This enables us to determine the breakdown of the signal's energy at each resolution level, at each electrode, Table 1.2[2]*. It is also possible to set up a table with the energy coefficients at each respective level, where the energy from each time location in a frequency level is termed an atom and the sum of the energies over a frequency is termed a crystal, (Table 1.3).*

[1]This results from "Parsevals's" relation, see Section 2.3

[2]The percentages across Table 1.2 do not sum to 100% as not all resolution levels that contain signal energy have been included.

TABLE 1.2. Percentage of energy across resolution levels for a given electrode

	Level				
Electrode	*0*	*1*	*2*	*3*	*4*
1	*2.35*	*13.81*	*64.86*	*7.35*	*7.6*
2	*0.0926*	*18.65*	*57.67*	*11.68*	*6.94*
3	*0.1834*	*27.73*	*64.53*	*4.15*	*1.95*
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
29	*0.4046*	*17.46*	*66.75*	*6.57*	*5.39*
30	*6.0473*	*13.67*	*64.00*	*6.44*	*7.93*

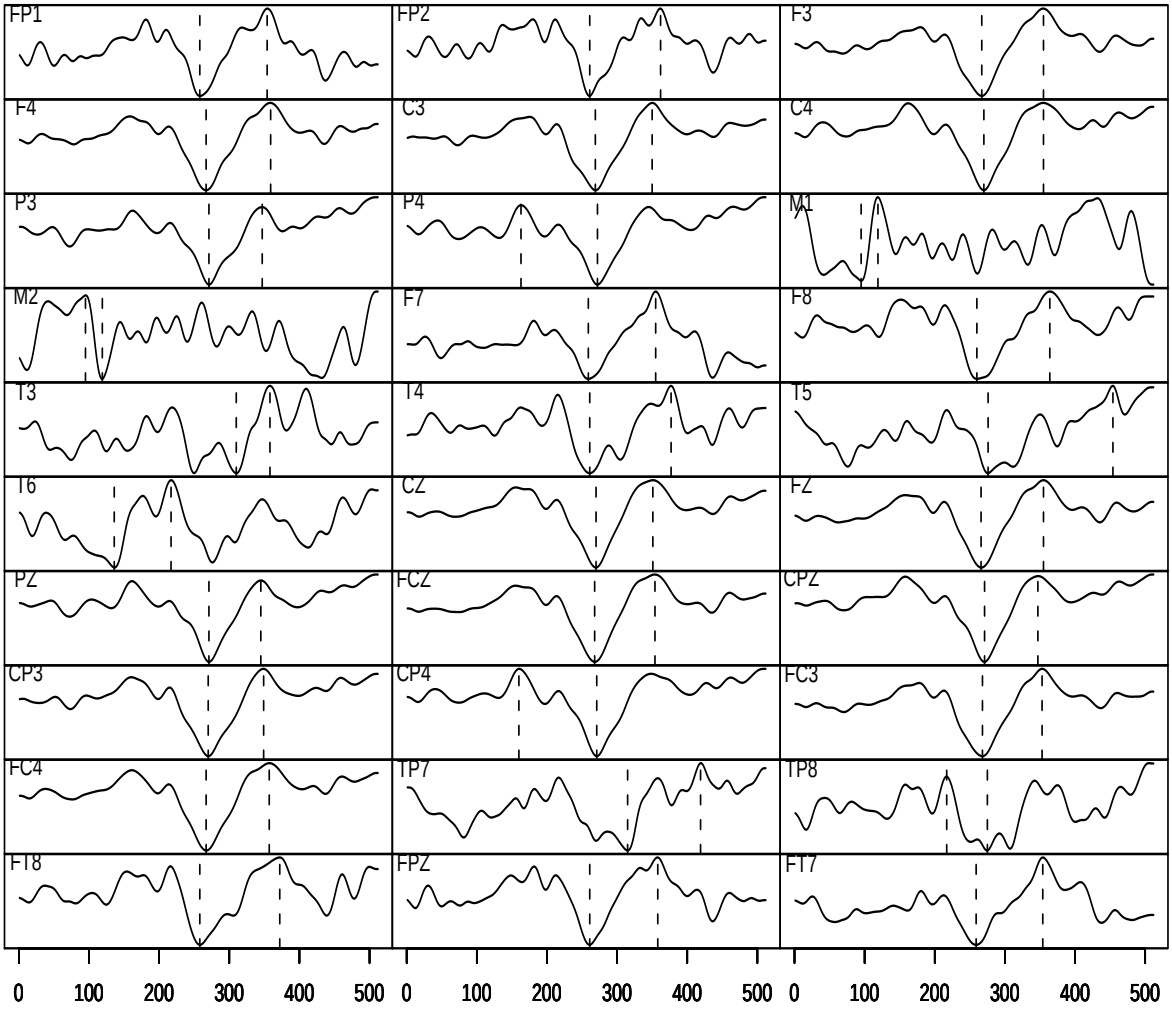
TABLE 1.3. Energy across time location and electrodes

	Electrodes					
Crystal/Atom	*FP1*	*FP2*	*...*	*TP7*	*TP8*	*FPZ*
1.1	*0.9919*	*1.1871*	*...*	*2.3658*	*0.58*	*0.9073*
1.2	*-3.4937*	*-4.3115*	*...*	*0.8979*	*-0.161*	*-3.1302*
2.1	*0.8413*	*1.2212*	*...*	*-0.727*	*0.2405*	*1.1851*
2.2	*1.9233*	*0.7736*	*...*	*-0.1824*	*-0.4931*	*0.6229*
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
4.13	*1.0354*	*0*	*...*	*-0.707*	*0*	*0*
4.14	*0*	*0*	*...*	*-0.5769*	*0*	*0*
4.15	*0*	*0*	*...*	*0*	*-0.6423*	*-1.2278*
4.16	*0*	*0*	*...*	*0*	*0.7622*	*0*

1.2.1. Reconstruction. *Using the various wavelet reconstruction routines in R it was possible to show where each signal had its maximum and minimum level or transition as well as how the signal passes through an abrupt change, (Figure 1.3). These time-locked samples highlight the variability of latency of these maximum and minimum deflections within the signal.*

This together with plan maps of how the signal moved across the scalp, provided us with a methodology on how to trace such a path.

FIGURE 1.3. Plot of Signals with transitions

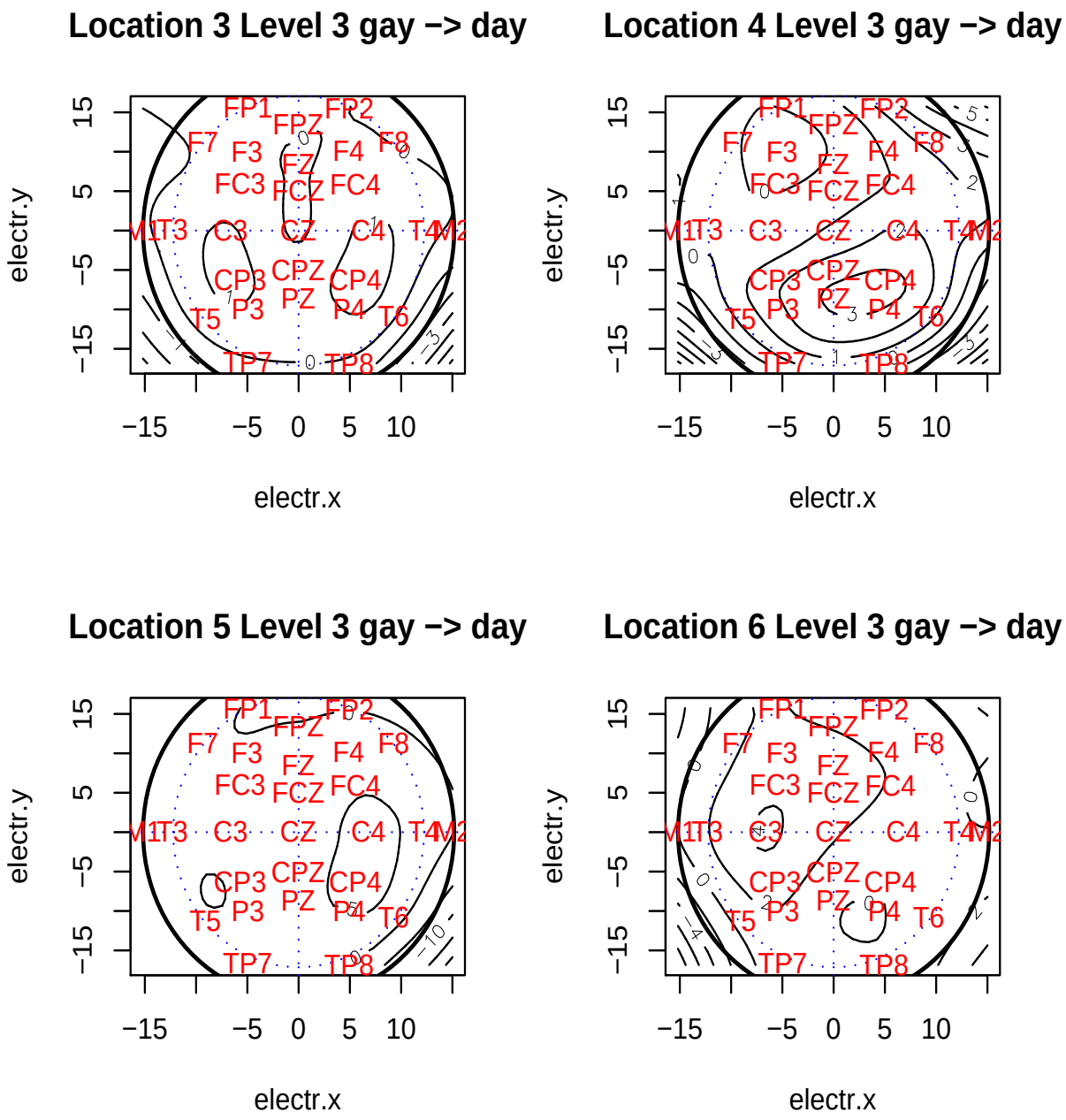


With such information it was possible to develop a contour map of the energy moving across the electrodes using the locations in time that wavelet resolution provides. i.e.; 4 locations at level 2 , 8 locations at level 3. Hence these wavelet energy coefficients at defined locations in time enable us develop a plot at points in time of the energy levels across the scalp, (Figure 1.4).

Initially we considered the time locations (atoms) with the most activity, usually those close to the transition, where we were planning to follow the signals moving across the electrodes. However, after plotting such maps for the different syllable changes at different levels (crystals), no suitably defined movement of the energy levels across the electrodes could be readily observed.

1.2.1.1. *Averaging. Now the result at this stage is possibly due to the lack of resolution within our methodology or as a result of the averaging of the signal. As the EEG*

FIGURE 1.4. Contour plot of Energy at time location



transitions appear well defined and (apart from the desired event related potential) there appears to be a lot of non-event related ongoing brain activity. In an attempt to visualise these event related potentials and to eliminate the spurious activity, the data collected may have been subject to averaging. Here the data provided was most likely averaged over latency and strength of the MMN component in the subjects understudy. Whatever the reason, we can find no clear well-defined path of energy across the electrodes. However

what is very noticeable is that the MMN usually demonstrates a clear positive spike at the frontal electrodes $F3, Fz, F4$ and the central electrodes $C3, Cz, C4$, while opposite deflections when comparing the mastoids $M1$ and $M2$ to each other.

Hence, while we may not be able to easily observe and follow the path of significant deflection in the MMN as a result of this averaging, nonetheless we may not conclude that a MMN has not been elicited in that subject. It has been noted that brain activity could vary widely between the subjects and trials, in regards to both time and distribution across the scalp. Similarly, to allow for the existence of spatial correlation between the electrodes, in turn leads us to consider another methodology, geostatistical analysis of spatial data [refer to Section 4.1].

Development. *Before developing this additional approach using geostatatistical analysis we first attempt to highlight why wavelets and the wavelet transform are relevant to this exercise, an overview of spatial-temporal point processes followed by the spatial data analysis using different models and an outline of the results.*

CHAPTER 2

Wavelet Analysis

2.1. Time and frequency representation

The first and most natural description of a signal is usually obtained by recording variations with time. The representation obtained by the Fourier transform is

$$X(f) = \int_{-\infty}^{+\infty} x(t)e^{-j2\pi ft}dt$$

and the reconstruction or inverse is given by

$$x(t) = \int_{-\infty}^{+\infty} X(f)e^{j2\pi ft}df$$

These are a useful way to describe a signal or decompose a function $f(t)$ in terms of its frequency components. We can say that $X(f)$ may be obtained by expanding the signal $x(t)$ into a set of infinite waves, $e^{-j2\pi ft}$. For example, a frequency burst in $x(t)$ will show as a large component in the Fourier transform $X(f)$. Therefore we can localise or detect frequencies with some accuracy in $x(t)$. Such is called the frequency localization property of the Fourier tranform. One problem here is there is no way to deduce the start or end of the frequency burst by an inspection of $X(f)$. This is why the Fourier transform is not suitable if the signal has time varying frequency spectrum, ie. the signal is non-stationary. Moreover this lack of localization in time makes the Fourier transform unsuitable for data processing systems of non-stationary signal or events, and this type of signal is of relevance in the biomedical field. Indeed large amount of information carried by signal like the ECG or EEG is found in transient short duration changes in the background activity.

2.2. Time frequency analysis

We have noticed that the Fourier transform is not at all well adapted to this analysis of a non - stationary signal. Of course one could consider functions of both time and frequency. Here we look at such time-frequency representations as the short time Fourier transform (STFT) and the wavelet transform.

2.2.1. The Short Time Fourier Transform. *To introduce a time relationship into the Fourier transform, we may pre-window the signal $x(t)$ near a particular time t, and calculate its Fourier transform. If we do such for each time t then the result is called the short-time Fourier transform.*

$$STFT_x(t, f; \omega) = \int_{-\infty}^{+\infty} x(\tau)\omega^*(\tau - t)e^{-j2\pi f\tau}d\tau$$

where $\omega(t)$ is the short time analysis window around $t = 0$ and $f = 0$. Provided that this short-time window is of finite energy, then the STFT is invertible as

$$x(t) = \frac{1}{E_h}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} STFT_x(\tau, \xi; \omega)\omega(t - \tau)e^{-j2\pi t\xi}d\tau d\xi.$$

This relationship highlights that the total signal may be decomposed into a sum of elementary waveforms

$$\omega_{t,f}(\tau) = h(\tau - t)e^{j2\pi j\tau}.$$

This can be expressed as building blocks or atoms. In turn each atom can be obtained from the window $\omega(t)$ by a translation in both time and frequency.

Considering the STFT from the frequency aspect we may state the STFT as

$$STFT_x(t, f; \omega) = \int_{-\infty}^{+\infty} X(\xi)W^*(\xi - f)e^{-j2\pi(\xi - f)\tau}d\xi$$

where X and W, are respectively, the Fourier Transforms of x and ω

This STFT could be thought of as the result of passing the signal through a band-pass filter and whose frequency response is $W^*(\xi - f)$ and is deduced from the mother filter $W(\xi)$ by a translation of f. Hence we could consider that the STFT is similar to a bank of band-pass filters with constant bandwidth.

The STFT provides us a time-frequency representation of the signal but it has resolution problems relating to the Heisenberg Uncertainty Principle. This states that one cannot know the exact time-frequency representation of a signal and that what we can only know are the time intervals in which a certain band of frequencies exist [**Polikar, 1999**][1].

[1]Polikar's online tutorial provides an insight to the limitations of the Fourier based transforms as well as to why the wavelet transform may provide solutions

Hence we have a trade-off between time and frequency resolutions, a good frequency resolution requires a narrow band filter, i.e. a long window $h(t)$, while a good time resolution requires a short window with short time support.

This fixed time-frequency resolution of STFT puts forward a constraint for many applications ie. from perfect time resolution but no frequency resolution to perfect frequency resolution but no time resolution.

The wavelet transform attempts to solve this problem. While a transform of a signal is just another form of representing the signal, it does change the information content already present in the signal. The wavelet transform provides a frequency representation of the signal and was developed to overcome the inadequacies of the STFT. Indeed as the STFT provides us with a constant resolution at all frequencies, the wavelet transform uses a multi-resolution technique, where different frequencies are analysed by different resolutions.

2.2.2. The Continuous Wavelet Transform. *The time and frequency width of the STFT function do not depend upon the location in the time frequency domain. However, in a number of practical situations, high frequency components are usually only present for short intervals and meanwhile low frequency components are present for long durations. The Continuous Wavelet Transform (CWT) can take advantage of these situations by use of good time resolution and poor frequency at high frequencies and conversely, good frequency resolution and poor time resolution at low frequencies. We may define the CWT of a signal x in respect to the analysing wavelet ψ by:*

$$CWT_x(\tau, s, \Psi) = \int_{-\infty}^{+\infty} x(t)\Psi^*_{\tau,s}(t)dt$$

where

$$\Psi_{\tau,s}(t) = \frac{1}{\sqrt{|s|}}\Psi\left(\frac{t-\tau}{s}\right)$$

*Ψ is a continuous function in both the time domain and frequency domain called the mother wavelet and * represents the operation of a complex conjugate. This transformed signal is a function of two variables s and τ, s being called the scale parameter. $|s| < 1$ compresses Ψ and $|s| > 1$ dilates it. τ is referred to as the translation parameter, as it shifts the position of the mother wavelet on the time axis.*

The wavelet transform may not be considered as a time frequency representation of a signal but rather a time scale representation. Here we may note the main difference between the wavelet transform and short time Fourier transform is related to the scale

parameter s. Now when this scale factor is changed, the duration and bandwidth of the wavelet is altered, however its profile remains the same. As low scales ($|s| > 1$) are related to low frequencies and high scales ($|s| < 1$) to high frequencies, we note the CWT uses short windows at high frequencies and conversely long windows at low frequencies. The STFT uses a single window for all frequencies, while the CWT provides a suitable time frequency representation of a signal with requiring the specification of window length. Hence, the CWT partially overcomes the limitation of resolution of the STFT. Furthermore, the bandwidth B of an analysis window is proportional to the frequency f or

$$\frac{B}{f} = K \equiv constant.$$

Therefore the CWT could be considered as a constant - K analysis, or also be seen as a filter bank analysis made up of band pass filters with constant relative bandwidth. The wavelet series is obtained by taking discrete parts of the CWT, hence computation via computers is obtained by sampling the time scale. This sampling can be altered in relation to scale change without violating the Nyquist rate.

Definition 2.1. The Nyqusit rate. Maximum rate of transmitting pules through a system. If B is the effective bandwidth in Hertz, then $2B$ is the maximum number of code elements per second which can be received with certainty. The Nyquist rate effects the minimum sampling rate allowable for accurate reconstruction of a signal. If f is the maximum frequency occurring in the signal, then a minimum sampling rate of $2f$ is demanded.

Remark 2.2. As the scale goes higher hence lower frequencies, the sampling rate can be decreased, there by reducing the required number of computations.

2.2.3. Multiresolution Analysis. *In multiresolution analysis (MRA) a function is considered at various levels of resolutions or approximations. The idea was developed by Meyer* [**Meyer, 1993**] *and Mallet* [**Mallat, 1989**]*. By application of MRA we may divide a complex function into many simple smaller functions and examine them individually. The monograph by Meyer* [**Meyer, 1993**] *provides in-depth mathematical development of MRA and describes the construction and properties of the various wavelet "types". In*

essence a mini encyclopedia of the functions available and the formulas required for using them in specific problems.

Here I will rather highlight the relation that exists between MRA and the Discrete Wavelet Transform.

2.2.4. The Discrete Wavelet Transform. *As the CWT permits us to perform a multiresolution of a signal, it nevertheless has some consequences that would make it nearly unusable for signal processing applications* [**Valens, 1999**]*:*

- *The CWT is highly redundant in regards to signal reconstruction. This follows as a continuously scalable set of wavelet functions is not an orthonormal basis.*[2]
- *The CWT of a signal is a continuous signal and therefore it is composed of an infinite number of terms.*
- *With most functions the CWT will have no analytical solution and such can only be calculated numerically.*

To overcome these problems, the Discrete Wavelet Transform (DWT) was introduced. The DWT is somewhat similar in its working to the CWT but using only discrete positions and scales.

In CWT, the signal is analyzed using a set of basis functions. In DWT, a time scale representation of the digital signal is obtained via digital filtering methods. In effect the signal to be analysed is passed through a number of filters with varied cut off frequencies at different scales.

2.2.4.1. *Multiresolution using filters. As filters are widely used in signal processing, wavelets may be considered as an iteration of filters with rescaling. The resolution of a signal, which could be seen as a measure of the information contained in the signal, can be determined by filtering operations, and the scale can be determined by various sampling methodologies ie. subsampling.*

The DWT can be computed by successive lowpass and highpass filtering of the discrete time-domain signal. This is called the Mallat algorithm.

Now at each level of filtering, having both a high pass and low pass filter, the high pass produces detail information while the low pass can be thought as associated with the

[2]The CWT is a "quasi-orthonormal" transform; it permits reconstruction although the wavelet transform is not orthogonal.

scaling function producing coarse approximations. At each such level these filters produce signals spanning only half the frequency band. Now if, as in agreement with Nyqusit's limitation, the signal in question has a highest frequency of f, which would require a sampling frequency of $2f$ radians, then it would now have a highest frequency of $f/2$ radians. Hence it may be sampled at a frequency of f radians, therefore removing half the samples with no loss of information. This decimation by 2 halves the time resolution as the signal now is represented by half the number of samples. Furthermore, as the half band low pass filtering removes half the the resolution, the decimation by 2 doubles the scale.

Using this approach, the time resolution improves relatively at high frequencies and conversely the frequency resolution improves at low frequencies. This filtering and decimation procedure may be continued until the required level of resolution is achieved. The maximum number of levels depends upon the length of the signal. The DWT of the original signal is obtained by concatenating the detail information from the high pass filter and course approximations from the low pass, starting from the last level of decomposition.

2.2.4.2. *Perfect reconstruction. In most wavelet transform applications, we require that the original signal be synthesized from the wavelet coefficients. Now to be able to perform perfect reconstruction the analysis and the filters need to satisfy the following:*

Let $L_0(z)$ and $L_1(z)$ be low pass analysis and synthesis filters,

$H_0(z)$ and $H_1(z)$ be high pass analysis and synthesis filters.

Then the filters need to satisfy the conditions as presented in Aldroubi, ***(1996)****.*

$$\begin{aligned} L_0(-z)L_1(z) + H_0(-z)H_1(z) &= 0; \\ L_0(z)L_1(z) + H_0(z)H_1(z) &= 2z^{-d}. \end{aligned}$$

REMARK 2.3. The first condition implies that the reconstruction is aliasing-free while the second implies that the amplitude distortion has an amplitude of one.

OBSERVATION 2.4. The perfect reconstruction condition does not change if we switch the analysis and synthesis filters.

We may classify wavelets into two classes:

(a) *orthogonal*

(b) *biorthogonal*

and further discussion of such and associated formulae may be found in Aldroubi, ***(1996)****.*

2.3. Wavelet families and applications

*There are a number of basis functions that can be used as the mother wavelet for wavelet transformation. As this mother wavelet produces all the wavelet functions that may be used within the transformation through translation and scaling, it therefore determines the characteristics of the resulting wavelet transform. As such, details of the application should at least be considered with the mother wavelet chosen in order to use the Wavelet transform effectively. We may say that the wavelet is a function used to to segment a function or a continuous signal into different components. These wavelets are scaled and translated copies (*daughter wavelets*) of a finite length oscillating waveform (*mother wavelet*). From the mother (*or prototype*) wavelet* Ψ *we may generate our wavelet:*

$$\Psi_{j,k}(t) = 2^{j/2}\Psi(2^j t - k)$$

where j, k *are integers.*

This wavelet system can then be used to construct or represent a function or signal, where a linear expansion of the function would be:

$$f(t) = \sum_{k=-\infty}^{+\infty} c_k \varphi(t-k) + \sum_{j=0}^{+\infty} \sum_{k=-\infty}^{+\infty} d_{j,k} \psi(2^j t - k)$$

The energy in a signal is given in the terms of the wavelet transform coefficients by Parsaval's [3] *relation as:*

$$\int |f(t)|^2 \, dt = \sum_{i=-\infty}^{+\infty} |c_i|^2 + \sum_{j=0}^{+\infty} \sum_{k=-\infty}^{+\infty} |d_{jk}|^2 .$$

[3]The equation which states that the square of the length of a vector in an inner product space is equal to the sum of the squares of the inner products of the vector with each member of a complete orthonormal base for the space. Also known as Parseval's identity : *the sum (or integral) of the square of a function is equal to the sum (or integral) of the square of its transform.*

2.3.1. Applications. *Wavelet applications are varied and apparently increasing; from signal processing and filtering, moving to face recognition, dam wall crack detection, data compression, data mining, even monitoring of ADSL interference. Furthermore there is even a specific journal: Journal of Wavelet Theory and Applications - which covers topics including wavelet methods for time series analysis, biomedical imaging, data analysis as well as wavelet applications in topography, nonlinear optics ...*

> *The future of wavelets lies in the as-yet uncharted territory of applications. Wavelet techniques have not been thoroughly worked out in applications such as practical data analysis, where for example discretely sampled time-series data might need to be analyzed. Such applications offer exciting avenues for exploration.* [**Graps, 1995**]

The Haar wavelet is one of the oldest and simplest wavelet, formed by all the translations and dilations of the Haar function ψ^H:

$$\psi^H = \begin{cases} 1 & : \quad x \in \left[0, \frac{1}{2}\right) \\ -1 & : \quad x \in \left[\frac{1}{2}, 1\right) \\ 0 & : \quad otherwise \end{cases}$$

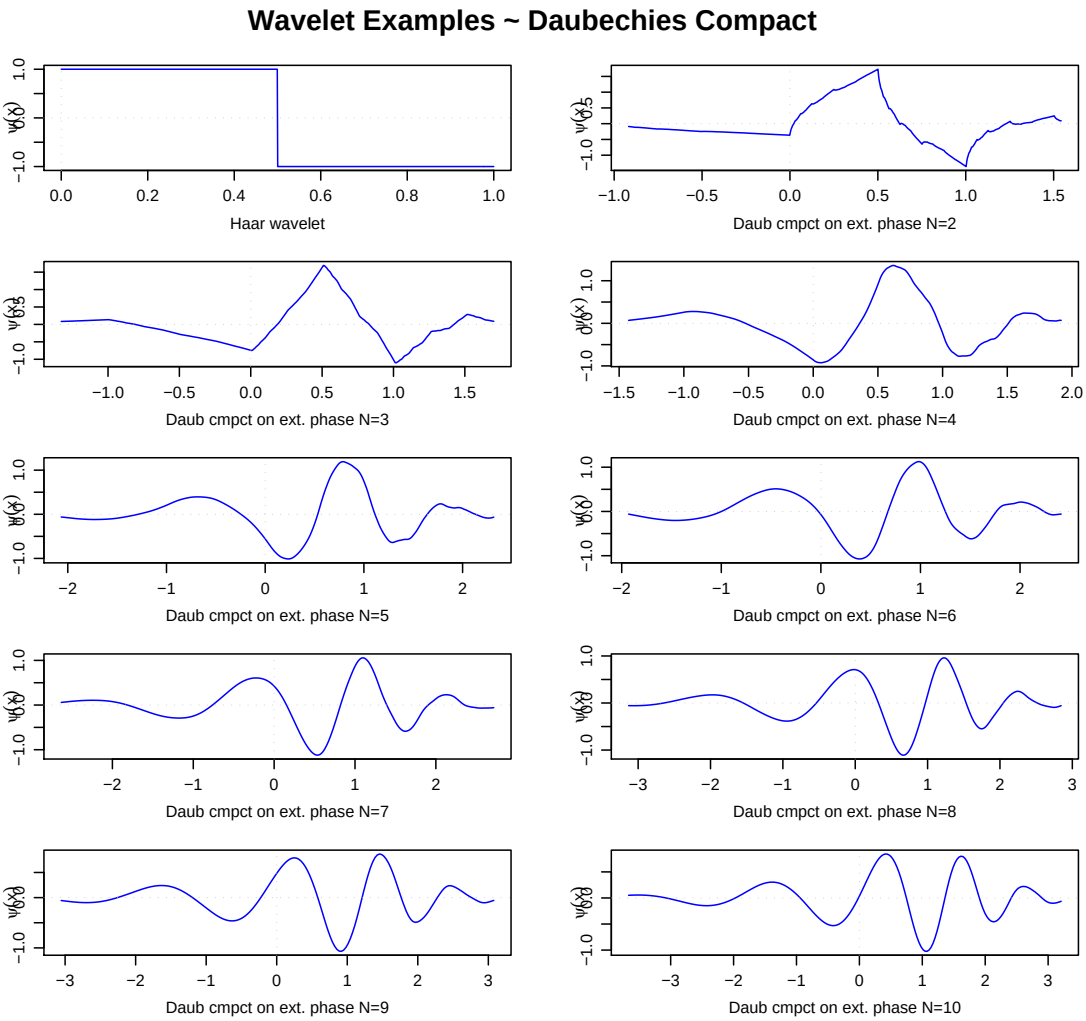
Indeed almost any discussion of wavelets starts with the Haar wavelet. Another is the Daubechies wavelets, which seem very popular as they are both orthogonal and are compactly supported[4] [**Aldroubi, 1996**]*. They are the foundations of wavelet signal processing and are inherent in numerous applications.*

The Haar, Daubechies, Symlets and coiflets are all compactly supported and orthogonal wavelets. These wavelets along with the Meyer wavelets are capable of perfect reconstruction. The Meyer, Morlet and Mexican Hat wavelets are symmetric in shape. Wavelets are chosen on their shapes and ability to analyse the signal within the application in question.

Figure 2.1 demonstrates some wavelets, these from the Daubechies family.

[4]A function has compact support if it is zero outside of a compact set.

FIGURE 2.1. Wavelets



For a concise overview of wavelet methods and uses as well as in depth discussion of the wavelet transform together with a lucid explanation of the many types of wavelets, with their vanishing moments[5]*, (N = x, in Figure 2.1), boundary conditions, packet transforms, etc . . . refer to Nason,* ***(2008)****.*

Nason also steps through some of the key features that wavelet methods provide:

- *Sparsity of representation for a wide range of functions.*
- *The ability to zoom in to analyse functions at number of scales.*
- *Ability to detect and represent localised features.*
- *Efficiency in terms of computational speed and storage.*

[5]Wavelets may possess a number of vanishing momemts: a function $\psi \in L^2(\mathbb{R})$ is said to have n vanishing moments if $\int x^\ell \psi(x)dx = 0$, for $\ell = 0, \ldots, n-1$. If a wavelet has n vanishing moments, then all wavelet coefficients of any polynominal of degree n or less will be equal to zero.

2.3.1.1. *Wavelet properties in relation to Biomedical Applications.*

- ***Wavelets as filter banks*** *In has been noted that the wavelet transform can be seen as a type of spectral analyser. The features that can be extracted from this type of arrangement are energy estimates in the different frequency bands. This approach has been used to to discriminate between differing physiological states. Of note is this feature extraction can only be justified when the signal in question is assumed to be stationary. Also similar results could also be obtained using Fourier techniques.*

- ***Wavelets and time frequency localisation*** *Most biomedical signals are a combination of sharp peak like events, i.e. spikes and transitions, as well as more evenly distributed changes i.e. heart rhythms and EEG waveforms. As the STFT will deliver suitable performance for the latter type of events, it is nevertheless much less suited for analysis of short duration pulses. Similarly the fixed resolution of the STFT does not permit the searching of both of these types of events with good resolution in both time and frequency. The Wavelet transform provides a compromise solution to this resolution problem. It has been shown this analysis is appropriate for analysis of ECG signals, the analysis of EEGs as well as a variety of other physiological signals. Many examples of such maybe seen in Aldroubi,* ***(1996)****.*

- ***Electroencephalography Applications*** *Electroencephalography waveforms such as EEG and ERP recordings from multiple electrodes vary their frequency over time and spatially across sites on the scalp. Hence both EEG and ERP are non-stationary. Of interest to the relevant researchers are three specific events within these data sets, tending to be transient(localised in time), determinable over specific scalp areas (localised in space) and the restriction to set ranges of temporal and spatial regions (localised in scale). As a result of these, wavelets are well-adapted for the analysis of EEG and ERP. Indeed wavelets may be found in many domains of processing neuro-electric signals, such as:*

  - *Noise Filtering*
  - *Waveform compression*
  - *Spike, transient and event detection*
  - *Analysis over time of EEG waveforms.*

CHAPTER 3

Some Theory

3.1. Distributions

This chapter describes some of the theory necessary to highlight this wavelet approach, as well as define the spatial-temporal point processes.

3.1.1. A Point Process. *A spatio-temporal point process is a random collection of points where each point represents the time and the location of an event. Examples of such events include incidence of disease, sightings or births of species, or the occurrences of fires, earthquakes, tsunamis, or volcanic eruptions. The points of a point process are generally assumed to be indistinguishable besides their different times and locations. There is though often additional information available to be stored with the information on time and location. The dataset could, for instance, contain information about several different strains of the same disease or members of different species. Such processes are called marked spatio-temporal point processes, ie. a random collection of points in time and space where each point has associated with it one or more further random variables describing the additional information. Much of the theory of spatio-temporal point processes is based on the theory for spatial point processes* [**Diggle, 2003**].

A spatio-temporal point process $\phi = \{[t_i, x_i]\}$ is defined as a locally finite set of points in a region $\mathbb{R} \times \psi$ of time-space. The set ψ is usually a subset of $\mathbb{R}^2$ or $\mathbb{R}^3$. Hence $\phi(A)$ is a number of points $[t_i, x_i]$ in A, where $A \in \beta(\mathbb{R} \times \psi)$, the Borel σ-algebra on $(\mathbb{R} \times \psi)$.

A Borel set is an element of a Borel σ-algebra. Borel sets are the sets that can be constructed from open or closed sets by repeatedly taking countable unions and intersections. Stated formally:

DEFINITION 3.1. The class B of Borel sets in Euclidean $\mathbb{R}^n$ is the smallest collection of sets that includes the open and closed sets such that if $E, E_1, E_2, \ldots$ are in B then so are $\bigcup_{i=1}^{\infty} E_i$, $\bigcap_{i=1}^{\infty} E_i$, and $\mathbb{R}^n \backslash E$, where $F \backslash E$ is a set difference[1].

[1]The set difference $A \backslash B$ is defined by $A \backslash B = \{x : x \in A \,\text{and}\, x \notin B\}$.

OBSERVATION 3.2. The set of rational numbers is a Borel set, as is the Cantor set.

The simplest and the most important random point patterns are the Poisson point processes. A Poisson process is a point process which satisfies two conditions:

- *The number of events in any bounded set $A \in \beta(\mathbb{R} \times \psi)$ follows a Poisson distribution with mean $\lambda\nu^k(A)$ where ν^k is the k-dimensional Lebesgue measure, the constant λ is the intensity, or mean number of events per unit area.*
- *The number of events in disjoint bounded Borel sets are independent.*

It follows that, conditional on the number of events in any bounded Borel set A, the locations of the events form an independent random sample for the uniform distribution on A.

We consider our EEG signal data as a set of point process datasets, but we do not observe the starting times and spatial origins of the activation directly. Instead we observe the event-related potential caused activation in the brain, which appears extremely complex. However an energy spike at location x at time u will contribute to the signal intensity at x at possibly a later time $t > u$. One possibility might be to consider stimuli at random time points such that we could have stochastic process $\{F_t\}$ such that

$$F_t = \sum g(t - t_i),$$

where $\{t_i\}$ is a Poisson Point Process and g is a form of a dynamic response function.

3.2. Poisson Point Process and Gamma Densities

The probability of finding k points, $k \geq 0$, in an interval of time t is given by

$$P_k(t) = e^{-\lambda_d t} \frac{(\lambda_d t)^k}{k!}$$

where $\lambda_d > 0$ is the average number of points per unit length of time. If we consider the distribution function of, say $D(t)$, of waiting times until the k^{th} Poisson event given a Poisson distribution with a rate of change λ

$$\begin{aligned} D(t) &= P(X \leq t) = 1 - P(X > t) \\ &= 1 - \sum_{h=0}^{k-1} \frac{(\lambda t)^h e^{-\lambda t}}{h!} \\ &= 1 - e^{-\lambda t} \sum_{h=0}^{k-1} \frac{(\lambda t)^h}{h!} \\ &= 1 - \frac{\Gamma(k, \lambda t)}{\Gamma(k)} \end{aligned}$$

for $t \in [0, \infty)$, where $\Gamma(t)$ is a complete gamma function and $\Gamma(a, t)$ an incomplete gamma function [2]*. With k an integer, this distribution is known as the Erlang disribution. The corresponding probability function $P(t)$ of waiting times until the k^{th} poisson event, is obtained by differentiating $D(t)$,*

$$\begin{aligned} P(t) &= D'(t) \\ &= \lambda e^{-\lambda t} \sum_{h=0}^{k-1} \frac{(\lambda t)^h}{h!} - e^{\lambda t} \sum_{h=0}^{k-1} \frac{h(\lambda t)^{h-1}\lambda}{h!} \\ &= \lambda e^{-\lambda t} + \lambda e^{-\lambda t} \sum_{h=1}^{k-1} \frac{(\lambda t)^h}{h!} - e^{\lambda t} \sum_{h=1}^{k-1} \frac{h(\lambda t)^{h-1}\lambda}{h!} \\ &= \lambda e^{-\lambda t} - \lambda e^{-\lambda t} \sum_{h=1}^{k-1} \left[\frac{h(\lambda t)^{h-1}}{h!} - \frac{(\lambda t)^h}{h!} \right] \\ &= \lambda e^{-\lambda t} \left\{ 1 - \sum_{h=1}^{k-1} \left[\frac{h(\lambda t)^{h-1}}{(h-1)!} - \frac{(\lambda t)^h}{h!} \right] \right\} \\ &= \lambda e^{-\lambda t} \left\{ 1 - \left[1 - \frac{(\lambda t)^{k-1}}{(k-1)!} \right] \right\} \\ &= \frac{\lambda(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t}. \end{aligned}$$

Let $\alpha \equiv k$ and define $\theta = 1/\lambda$ to be the time between changes, then the above can be written as $P(t) = \frac{t^{\alpha-1}e^{t/\theta}}{\Gamma(\alpha)\theta^\alpha}$ for $t \in [0, \infty)$. This is the probability function for the gamma distribution.

[2] The complete gamma function $\Gamma(a)$ can be generalised to incomplete gamma function $\Gamma(a, x)$ such that $\Gamma(a) = \Gamma(a, 0)$. This incomplete gamma function is given by $\Gamma(a, x) \equiv \int_x^\infty t^{a-1}e^{-t}dt$.

Our wavelet transform is linear transform ie. $Y \to X\beta$ and our signal energy "levels" are the square of wavelet coefficients. This leads us to consider the properties of the Gamma distribution,

3.2.1. Gamma Densities.

DEFINITION 3.3. A Gamma(α) random variable with $\alpha > 0$,
has the density $f(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)}e^{-x}$ on $x > 0$,
where $\Gamma(\alpha)$ is the constant that makes the density integrate to one.

DEFINITION 3.4. The Gamma function is defined on $\alpha > 0$
by $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$.

then as $Y = \beta X$ is a Gamma(α, β) random variable, we obtain

$$f(y) = \frac{y^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)}e^{-y/\beta}.$$

Now if we let Z be a standard normal random variable and let $Y = Z^2$ then using an intermediate step, let $W = |Z|$.

W has a cumulative distribution function

$$F_w(w) = P(W \le w) = P(-w \le Z \le w) = F_z(w) - F_z(-w)$$

By differentiating

$$f(w) = f_z(w) + f_z(-w) = \frac{2}{\sqrt{2\pi}\sqrt{y}}e^{-w^2/2} \text{ on } W > 0.$$

We may now use our transformation, and its inverse is $W = \sqrt{Y}$ with its derivative being $1/(2\sqrt{Y})$.

With the density being $f(y) = \frac{1}{\sqrt{2\pi}\sqrt{y}}e^{-y/2}$ for $y > 0$.

From this we note that the square of a standard normal variable is Gamma(1/2, 2)[3].

Therefore as $\alpha \to \infty$ we conclude that the gamma density is approximately normal.

This theory enables us to conclude that the signal energy "levels", (squared wavelet coefficients) are the sum of squares of normal random variables and will have a Gamma distribution.

[3]If $X \sim \Gamma(k = v/2, \theta = 2)$ then X is indentical to $\chi^2(v)$ the chi-square distribution with v degrees of freedom. Conversely, if $P \sim \chi^2(v)$ and c is a positive constant, then $c.P \sim \Gamma(k = v/2, \theta = 2c)$.