Foulis, Alison Anne (2010) *A multilayered approach to the automatic analysis of the multifocal electroretinogram.* PhD thesis.

A Multilayered Approach to the Automatic Analysis of the Multifocal Electroretinogram

Alison Anne Foulis

Submitted in fulfilment of the requirements for
the degree of Doctor of Philosophy

Department of Clinical Physics
Faculty of Medicine
University of Glasgow

UNIVERSITY
of
GLASGOW

August 2010

Abstract

The multifocal electroretinogram (mfERG) provides spatial and temporal information on the retina's function in an objective manner, making it a valuable tool for monitoring a wide range of retinal abnormalities. Analysis of this clinical test can however be both difficult and subjective, particularly if recordings are contaminated with noise, for example muscle movement or blinking. This can sometimes result in inconsistencies in the interpretation process. An automated and objective method for analysing the mfERG would be beneficial, for example in multi-centre clinical trials when large volumes of data require quick and consistent interpretation. The aim of this thesis was therefore to develop a system capable of standardising mfERG analysis. A series of methods aimed at achieving this are presented. These include a technique for grading the quality of a recording, both during and after a test, and several approaches for stating if a waveform contains a physiological response or no significant retinal function. Different techniques are also utilised to report if a response is within normal latency and amplitude values.

The integrity of a recording was assessed by viewing the raw, uncorrelated data in the frequency domain; clear differences between acceptable and unacceptable recordings were revealed. A scale ranging from excellent to unreportable was defined for the recording quality, first in terms of noise resulting from blinking and loss of fixation, and secondly, for muscle noise. 50 mfERG tests of varying recording quality were graded using this method with particular emphasis on the distinction between a test which should or should not be reported. Three experts also assessed the mfERG recordings independently; the grading provided by the experts was compared with that of the system.

Three approaches were investigated to classify a mfERG waveform as 'response' or 'no response' (i.e. whether or not it contained a physiological response): artificial neural networks (ANN); analysis of the frequency domain profile; and the signal to noise ratio. These techniques were then combined using an ANN to provide a final classification for 'response' or 'no response'. Two methods were studied to differentiate responses which were delayed from those within normal timing limits: ANN; and spline fitting. Again the output of each was combined to provide a latency classification for the mfERG waveform. Finally spline fitting was utilised to classify responses as 'decreased in amplitude' or 'not decreased'. 1000 mfERG waveforms were subsequently analysed by an expert; these represented a wide variety of retinal function and quality. Classifications stated by the system were compared with those of the expert to assess its performance.

An agreement of 94% was achieved between the experts and the system when making the distinction between tests which should or should not be reported. The final system classified 95% of the 1000 mfERG waveforms correctly as 'response' or 'no response'. Of those said to represent an area of functioning retina it concurred with the expert for 93% of the responses when categorising them as normal or abnormal in terms of their P1 amplitude and latency. The majority of misclassifications were made when analysing waveforms with a P1 amplitude or latency close to the boundary between normal and abnormal.

It was evident that the multilayered system has the potential to provide an objective and automated assessment of the mfERG test; this would not replace the expert but can provide an initial analysis for the expert to review.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Author's declarations

The material presented in this thesis is the author's own work with the following exceptions:

Dr Stuart Parks wrote the software for the custom build multifocal system.

Sinead Walker, Angela McCall and Ann McQuiston helped to test the volunteers when collecting normative data.

Definitions

ANN	Artificial neural network
CRT	Cathode ray tube
DFT	Discrete Fourier transform
ECG	Electrocardiogram
EEG	Electroencephalogram
ERG	Electroretinogram
EOG	Electro-oculogram
FERG	Focal electroretinogram
FFT	Fast Fourier transform
LCD	Liquid crystal display
LED	Light emitting diode
mfERG	Multifocal electroretinogram
mfVECP	Multifocal visual evoked cortical potential
PERG	Pattern electroretinogram
RMS	Root mean square
RPE	Retinal pigment epithelial
SLO	Scanning laser ophthalmoscope
SNR	Signal to noise ratio
VECP	Visual evoked cortical potential

List of presentations

Material contained in this thesis has been disseminated at national and international meetings. Details or oral and poster presentations are listed below:

Foulis A, Parks, Keating D. A multilayered approach to the automatic analysis of multifocal electroretinogram waveforms: the artificial neural network. Poster presentation at The Association for Research in Vision and Ophthalmology meeting, Fort Lauderdale, USA, May 2008.

Foulis A, Parks S, Keating D. A multilayered approach to the automatic analysis of multifocal electroretinogram waveforms: the artificial neural network. Oral presentation at the British Society for Clinical Electrophysiology in Vision conference, Cardiff, September 2008.

Foulis A, Parks, Keating D. A multilayered approach to the automatic analysis of multifocal electroretinogram waveforms: analysis of the Fourier domain profile. Poster presentation at The Association for Research in Vision and Ophthalmology meeting, Fort Lauderdale, USA, May 2009.

Foulis A, Parks, Keating D. The use of the Fourier domain profile as part of a multilayered approach to the automatic analysis of the multifocal electroretinogram. Presented at the International Society for Clinical Electrophysiology in Vision, Padova, Italy, July 2009.

Foulis A, Parks, Keating D. A multilayered approach to the automatic analysis of multifocal electroretinogram waveforms: objective grading of signal quality using the Fourier domain profile. Poster presentation at The Association for Research in Vision and Ophthalmology meeting, Fort Lauderdale, USA, May 2010.

1 Introduction

The aim of this thesis was to develop a system capable of automatically analysing the multifocal electroretinogram (mfERG). This is a clinical test used to assess the function of the retina, the neural tissue lining the back of the eye. Prior to discussing the techniques used to achieve this objective analysis, the main parts of the human visual pathway are described, in addition to a review of a number of methods used to assess the health of this system. Particular attention is paid to tests designed to detect functional rather than structural abnormalities as the mfERG, the subject of this thesis, is one such test.

1.1 The eye

The eye is extremely sophisticated in its function. It receives light from the outside world, focuses it on the retina and then processes the information to provide us with a meaningful view of the world. The retina enables us to differentiate colours from one another, to see very fine detail and to see in conditions ranging from dim light to bright sunlight. Many structures in the eye are involved in ensuring that the light arriving at the retina is well focused, allowing a clear image to be formed (1). A number of these structures are shown below:
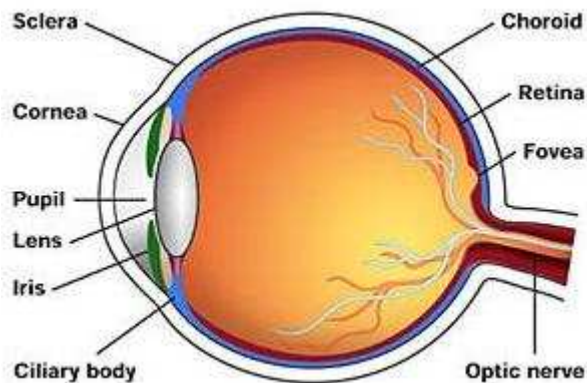
Figure 1.1 Schematic of the eye detailing the main structures (adapted from www.healthyeyes.org.uk).

The majority of the surface of the eyeball is surrounded by a dense, fibrous connective tissue, referred to as the sclera; this acts as a protective layer for the eye. The cornea is a transparent layer at the front of the eye. Light entering the eye must be refracted if focussing is to be achieved; almost two-thirds of this takes place at the air-cornea interface. The lens is responsible for further refraction, and is also required to make continuous adjustments in order to focus on objects at various distances from the eye.

When focussing on distant objects the lens becomes elongated in shape, while it becomes more spherical when focussing on close objects. Both the iris and ciliary body (mainly consisting of the ciliary muscle) are responsible for this adjustment to the shape of the lens. The choroid, which delivers nutrients and oxygen to the outer retina is situated between the sclera and the retina (1).

1.2 The retina

A simplified schematic of the retina is shown in figure 1.2. It consists of a number of layers of cells, each of which has a different role:
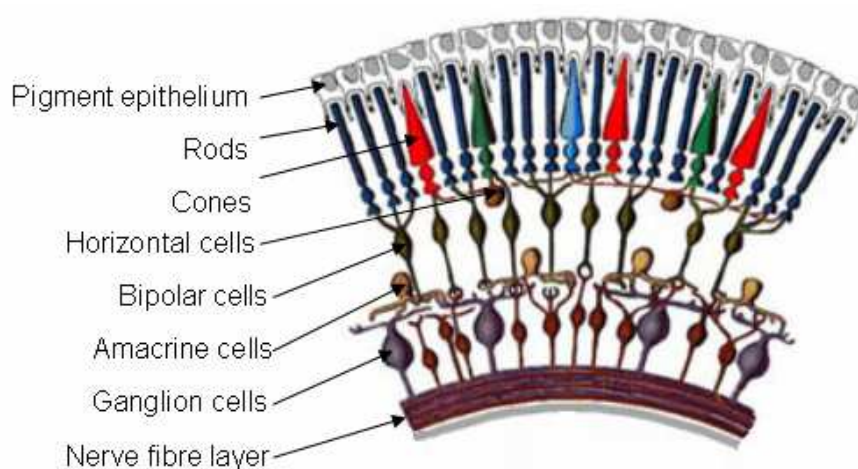
Figure 1.2 Schematic of the retina showing the main structures (reproduced from www.webvision.com).

The pigment epithelium of the retina has various functions, one of which is to absorb light entering the retina, thus preventing it from being reflected back through the layers of cells.

1.2.1 The photoreceptors

The photoreceptors detect light entering the retina. Two groups of photoreceptor exist: rods; and cones. The rods are used to see in dim light while the cones provide us with colour vision and enable us to form clear, sharp images. The rods are most sensitive to blue-green light, with a maximal sensitivity to a light wavelength of approximately 500nm. To enable colour vision three types of cone exist: the S-cone; the M-cone; and the L-cone. Each cone type is particularly sensitive to a different part of the light spectrum, with peak sensitivities of approximately 445nm (blue), 540nm (green) and 565nm (red) for the S-cone, the M-cone and the L-cone respectively (2). This can be seen in figure 1.3:
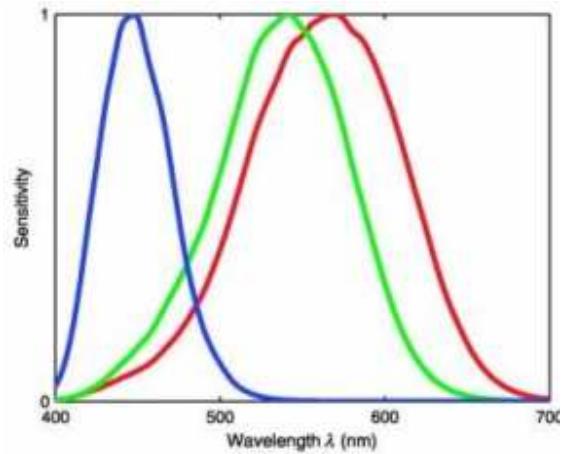
Figure 1.3 A sensitivity plot of the three types of cone to different wavelengths of light (adapted from Lewis *et al.* (2)). The S-cone (blue), the M-cone (green) and the L-cone (red) have peak sensitivities of approximately 445nm, 540nm and 565nm respectively.

The number of rods found in the retina far exceeds that of cones; it was shown by Curcio *et al.* that the average human retina contains 92 million rods and 4.6 million cones (3). The distribution of these varies markedly over the surface of the retina. The following diagram represents the density of each of the two types of photoreceptor across the horizontal meridian of the retina:
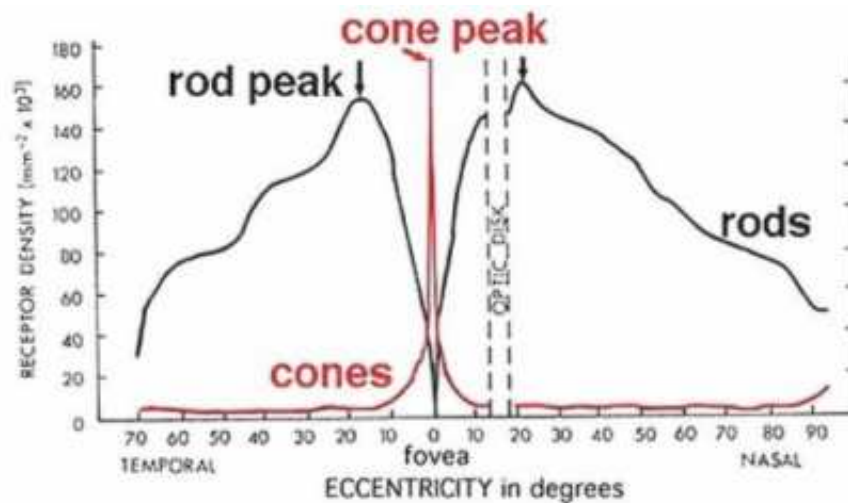
Figure 1.4 Distribution of rods and cones in the retina (adapted from Osterberg *et al.* (4)). The highest density of cones is found centrally, at the fovea while the highest rod density is 18 degrees from the centre.

It can be seen that at the fovea, the part of the retina responsible for sharp, detailed vision there are no rods, only cones. The highest density of cones is found in this region to achieve a high central visual acuity. Cones are present throughout the retina, but with a lower density than is seen at the fovea. This will become relevant in section 1.5.5.1. The highest density of rods is found in a ring surrounding the fovea at approximately 18 degrees from the foveal pit. No photoreceptors are found where the optic nerve is located; this is referred to as the blind spot.

1.2.2 Transmission of information through the retina

Information is transmitted through the retina in one of two paths: a direct; and an indirect path. The former route runs directly from the photoreceptors to the bipolar cells, and finally to the ganglion cells. Horizontal and amacrine cells are involved in the indirect path. Horizontal cells are located at the synapse between the photoreceptors and the bipolar cells, forming lateral connections within the retina; these can inhibit communication between the photoreceptors and the bipolar cells. The amacrine cells form a layer between the bipolar cells and the ganglion cells and like horizontal cells they form lateral connections (1). The retina contains approximately 0.7 to 1.5 million ganglion cells (5); considerable processing therefore takes place in the retina, thus reducing the amount of information to be processed at a later stage (1).

1.3 Transmission of information from the eye to the brain

The axons from the ganglion cells form the nerve fibre layer, converging at the optic disc. These fibres, which form the optic nerve, penetrate the eye and arrive at the optic chiasm. Approximately half of the fibres from the optic nerve proceed towards the lateral geniculate nucleus on the same side of the brain while the other half cross to the lateral geniculate nucleus on the opposite side of the brain. Axons from each lateral geniculate nucleus then terminate in the primary visual cortex of the cerebral hemisphere, where more complex visual processing takes place. This process is shown in figure 1.5:
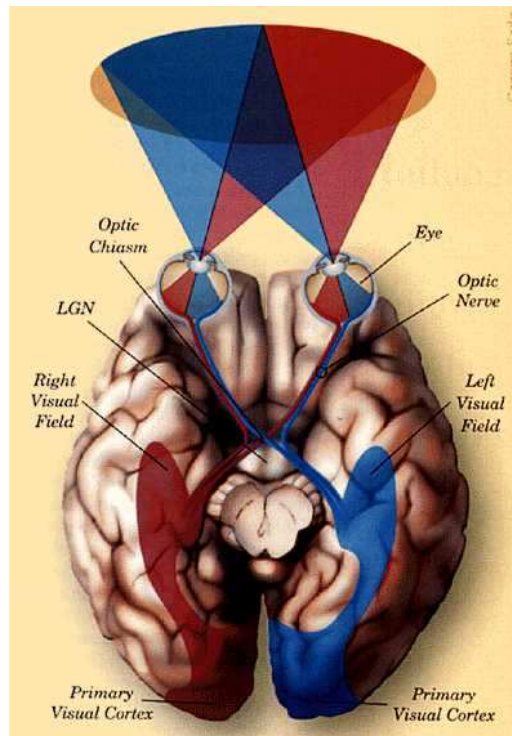
Figure 1.5 Schematic of the human visual pathway showing the transmission of information from the eye to the brain (adapted from www.cs.nps.navy.mil/people/faculty/capps/4473/projects).

1.4 Methods employed to assess vision

A number of techniques are utilised to investigate the integrity of the visual pathway, a selection of which are discussed in section 1.4.

1.4.1 Visual acuity

Visual acuity is a measure of a person's ability to see fine detail. The Snellen test, developed by Snellen in 1868, is commonly used to assess this. The following illustration is an example of a Snellen chart:

A
D F
H Z P
T X U D
Z A D N H
P N T U H X
U A Z N F D T
N P H T A F X U

Figure 1.6 The Snellen chart, employed to measure visual acuity (adapted from www.prosportsvision.co.uk/images/snellen).

The card is typically placed 6m from the person; they will read each line in turn, starting at the top and continuing until the final line, or the point at which they can no longer identify the letters. Visual acuity is recorded as a fraction, with the numerator being the distance between the card and the subject (i.e. 6m) and the denominator being the distance from which a normal eye could see the final letter seen by the person. When someone is unable to visualise the letters on the chart cruder methods are used such as a person's ability to count fingers, to detect hand movements, or their perception to light. One eye is tested at a time. Although the Snellen test is a standard method used to assess a person's visual acuity it remains subjective and is reliant upon patient cooperation (6).

1.4.2 Perimetry

While the visual acuity evaluates a person's central vision it provides no information on their visual field. The visual field is the space in which a person can detect an object when maintaining a steady gaze in one direction. Typically from fixation this extends 60^0 superiorly, 75^0 inferiorly, 60^0 nasally and 105^0 temporally. Damage to any part of the visual pathway can cause field defects within this region, therefore by measuring the function of the visual field, abnormalities can be detected. The aim of perimetry is to plot the sensitivity of the visual field and hence abnormalities can be visualised with ease. Essentially patients are presented with light stimuli and indicate when they can see the stimulus. Perimetry has an important role to play in the diagnosis and

monitoring of various conditions, for example, glaucoma, however it is subjective and relies upon the cooperation of the patient. It can also suffer from poor repeatability and can be relatively time consuming as the patient is tested monocularly and each point in the visual field is tested in a serial fashion (6).

1.5 Visual electrophysiology

Electrophysiology, a branch of physiology which studies the electrical phenomena associated with physiological processes, offers an objective means of assessing various parts of the visual pathway. The mfERG, the subject of this thesis, is one such test. The main tests used in this field are presented in section 1.5. For each of the tests described a response is evoked by stimulating the eye with a pattern or flash stimuli. This electrical potential is then measured and used to determine if a specific part of the visual pathway is intact. The visual evoked cortical potential (VECP) is discussed first.

1.5.1 The visual evoked cortical potential

The VECP is the response evoked by a visual stimulus. It is generated in the occipital cortex, the part of the brain responsible for processing visual stimuli, and can be measured to establish the function of the central visual pathway. To acquire this response the patient looks at a stimulus, commonly a black and white checkerboard screen. The black squares switch to white and vice-versa, causing a response to be evoked. To record this response an electrode is placed on the patient's scalp, over the visual cortex. Signal averaging is performed to recover the response from both electrical noise and noise associated with muscle artefacts. The latency of the recovered response is used for the clinical diagnosis of various conditions including multiple sclerosis, optic atrophy and optic nerve compression (7).

The evoked potential can also be used as an objective means of assessing a person's visual acuity. This is beneficial when communication is a problem or when someone is thought to be exaggerating their visual loss. To achieve this the size of the checks used for the stimulus is decreased until no response is evoked in the visual cortex. The highest spatial resolution at which a response is recovered can then be correlated to their visual acuity (8;9). This objective technique is useful clinically however it only provides information on the integrity of the central visual pathway; abnormalities affecting the peripheral vision are not therefore detected.

1.5.2 The electro-oculogram

An additional test is the electro-oculogram (EOG), a test used to assess the function of the retinal pigment epithelial (RPE) layer and the outer retina. The EOG is an indirect measurement of the potential difference across the eye. When in the dark this potential decreases, reaching a minimum referred to as the dark trough; in light conditions the potential initially decreases and then, in the healthy eye, it increases, reaching a maximum termed the light peak. The Arden index, the ratio of the light peak to the dark trough is calculated. No international normative range currently exists for this ratio however a ratio of less than 1.5 is considered to be abnormal, a ratio of greater than 2 is thought to be normal while values in between are equivocal (10). For many conditions in which the RPE or photoreceptor layer is affected both the electroretinogram (ERG; this is an examination used to assess the function of the outer and mid retina and is discussed in section 1.5.3) and the EOG are abnormal, meaning that the EOG is not essential. It is however of importance when diagnosing Best's disease and its variants, as this condition has a normal ERG but an abnormal EOG (8).

1.5.3 The electroretinogram

The ERG measures the electrical response of the retina to a light stimulus. A corneal electrode is used to measure the potential difference across the cornea while reference and ground electrodes are typically placed on the patient's forehead and ear (or temple), respectively. Several types of ERG tests exist, one of which is the flash ERG.

1.5.3.1 Flash electroretinogram

A flash of light evokes a global response from the outer and mid retinal layers (11). The intensity of the flash, the frequency at which it is presented and the state to which the eye is adapted, (i.e. light or dark adapted) all determine the degree of influence from specific cells in the recovered response. When the eye is light adapted, a response elicited from a flash of light is predominantly derived from the cone pathway, while the presentation of a low intensity flash when the eye is dark adapted provides information on the rod pathway (11). International standards are used when conducting this examination to ensure consistency of the test (12). The five most commonly recovered responses are shown in figure 1.7. The first three responses: the rod; the combined rod-cone; and the oscillatory potential, are all obtained when the

eye has been adapted to darkened conditions. The flash intensity is 0.01 cd.s/m^2 for the rod response; this is increased to 3.0 cd.s/m^2 for both the rod-cone response and the oscillatory potential. The bottom two responses shown in figure 1.7, the cone and the flicker responses, are acquired after a period of light adaptation (10 minutes); a flash intensity of 3.0 cd.s/m^2 is utilised in each case. The former is a single flash while the latter is presented at 30Hz.
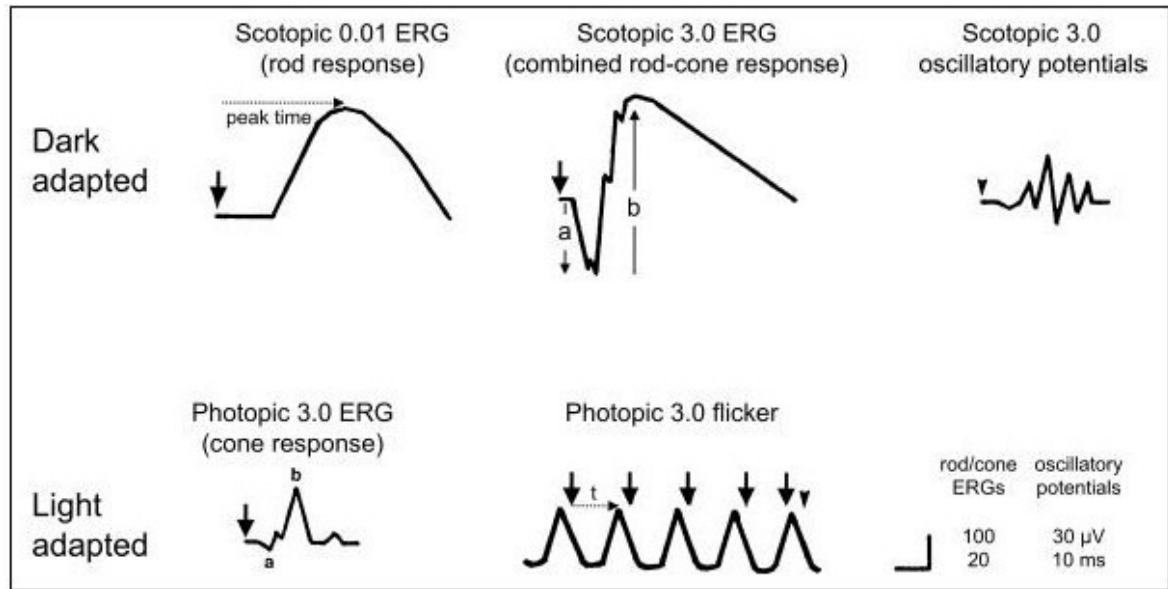
Figure 1.7 The 5 ERG responses typically recovered, as defined by international standards (12). The three responses recovered from the dark adapted eye are shown in the top line: the rod; the combined rod-cone; and the oscillatory potential response. Two types of response acquired from the light adapted eye are shown on the bottom line: the cone; and the flicker response. The magnitude and latency of the b-wave, and when present, the a-wave, are measured to assess if the response is normal.

Two main components make up these responses: the a-wave; and the b-wave. Bush *et al.* suggested that the main part of the retina contributing to the a-wave is the OFF bipolar cells (a type of bipolar cell which depolarises when stimulated) (13). ON bipolar cells (bipolar cells which are hyperpolarised when stimulated) are considered to generate the b-wave (14). It is also likely that the activity of the cones contributes to the a-wave (13) and that the OFF bipolar cells contribute to the b-wave (14). The amplitude and latency values of the b-wave, and where appropriate the a-wave, are measured and compared with normative data to determine if a response is normal.

The ERG is used to monitor many conditions including retinitis pigmentosa, cone-rod dystrophies and vitamin A deficiency (15). It is however a global assessment of function thus localised function cannot be measured. When functional loss is highly specific, such as in a central scotoma, the person may present with normal ERG responses but have very poor visual acuity. This led to the development of the focal electroretinogram (FERG).

1.5.3.2 Focal electroretinogram

The FERG was designed to elicit a physiological response from a smaller area of the retina. Typically a flickering stimulus of approximately 10^0 (or smaller) is used to evoke the ERG response; this stimulus is surrounded by a background light of constant illumination. The scope of this technique is however limited as responses can only be elicited from a small number of areas in one testing session due to time constraints (8).

1.5.3.3 Pattern electroretinogram

A third variation of the ERG is the pattern electroretinogram (PERG). The PERG may reflect the function of the inner retinal layers and can provide information on the patient's central visual function (16;17). A checkerboard stimulus comprising black and white squares is utilised to stimulate the patient; the black squares change to white and vice versa, evoking a response. The responses recovered are however very small, long recording times are required, and a high degree of variability is seen between subjects (8).

1.5.4 Limitations of conventional electrophysiology

Electrophysiology is an objective means of assessing aspects of the visual pathway, making it is less dependent upon patient cooperation than tests such as the Snellen chart or perimetry. Local defects can however be missed. This limitation has led to the development of the multifocal electroretinogram (mfERG) and the multifocal visual evoked cortical potential (mfVECP), objective methods used to map the function of the retina and the visual field of the visual cortex respectively.

1.5.5 Multifocal techniques

When using multifocal techniques multiple areas of the retina, or visual field, are stimulated simultaneously, yet independently from one another, eliciting responses from individual areas. A map of function can therefore be plotted, providing a more detailed clinical picture.

1.5.5.1 The multifocal electroretinogram

The mfERG, as first described by Sutter and Tran (18) provides objective spatial and temporal information on the function of the outer and the mid retina. Since its introduction it has been used widely in both the research and clinical setting. It is a valuable tool for monitoring a wide range of retinal abnormalities including retinitis pigmentosa (19-21), diabetic retinopathy (22-24), retinal vein occlusions (25;26), Stargardt's macular dystrophy (27), and drug toxicity, for example that associated with Vigabatrin (28) or Chloroquine (29).

The patient preparation is identical to that employed for the ERG, as described in section 1.5.3. The main difference lies in the stimulus and the method by which the final responses are obtained. An example of a stimulus is shown in figure 1.8, with a response corresponding to each stimulating element. In this example 61 scaled elements are being used to stimulate the retina therefore evoking 61 individual responses.
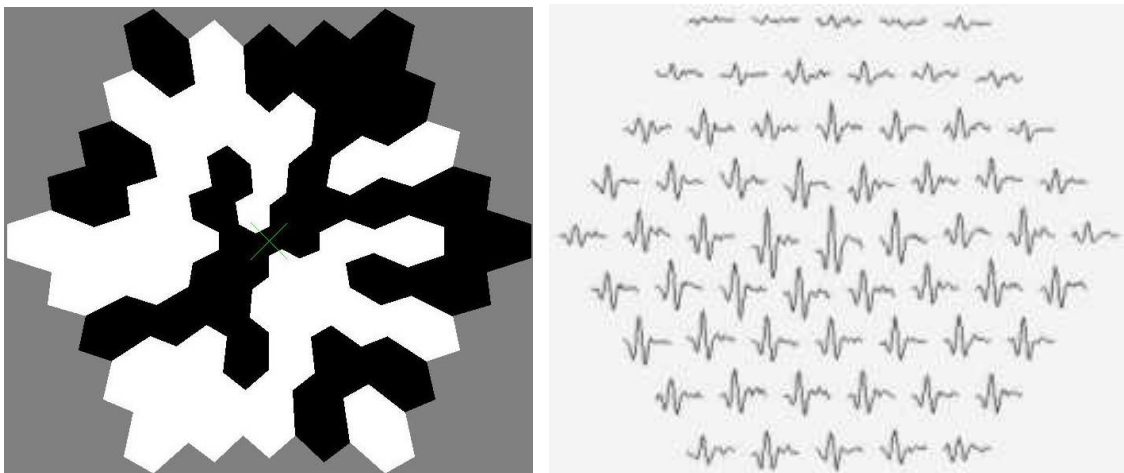
Figure 1.8 The mfERG stimulus (left) and corresponding trace array (right). Each element in the stimulus is used to evoke an individual response in the trace array, thus 61 responses, corresponding to the 61 elements, can be seen.

It can be seen that the size of the hexagons increases with eccentricity; this is to obtain responses of similar amplitude across the field. This is to account for the density of photoreceptor cells across the visual field (figure 1.4) and the adaptation variation across the retina. Unique mathematical sequences determine the luminance of each of the elements throughout the recording period, switching them between black and white. Recordings typically take approximately 4-8 minutes, depending on the length of sequences used to drive the luminance of the elements. Each local response is recovered by cross correlating the sequence used to control it against the recorded

data. This is done for every area to yield the final trace array. An individual waveform taken from a healthy individual is shown:

Figure 1.9 An individual mfERG response with the three main turning points (N1, P1, and N2) labelled.

Three clearly defined turning points can be seen, referred to as N1, P1 and N2. The amplitude and the latency of each of these are measured to determine if the corresponding area of retina is functioning normally. Latency measurements are made from the start of the flash. The N1 and N2 amplitudes are measured from the baseline, while the P1 amplitude is measured from the N1 trough. Abnormalities are reflected as a reduction in amplitude and/or a delay in the response. If no physiological response is acquired from an area of the retina, the mfERG waveform will consist solely of noise. The mfERG will be discussed in greater detail in chapter two and will be the subject of the remainder of this thesis.

1.5.5.2 The multifocal visual evoked cortical potential

The multifocal visual evoked cortical potential (mfVECP) enables VECP responses to be evoked from many different areas of the visual field simultaneously. It can be useful for monitoring patients with conditions such as optic neuritis (30;31) and glaucoma (32;33). It has also been used to test patients suspected of malingering (34;35). A typical stimulus is shown in figure 1.10, along with the responses obtained from a normal individual:
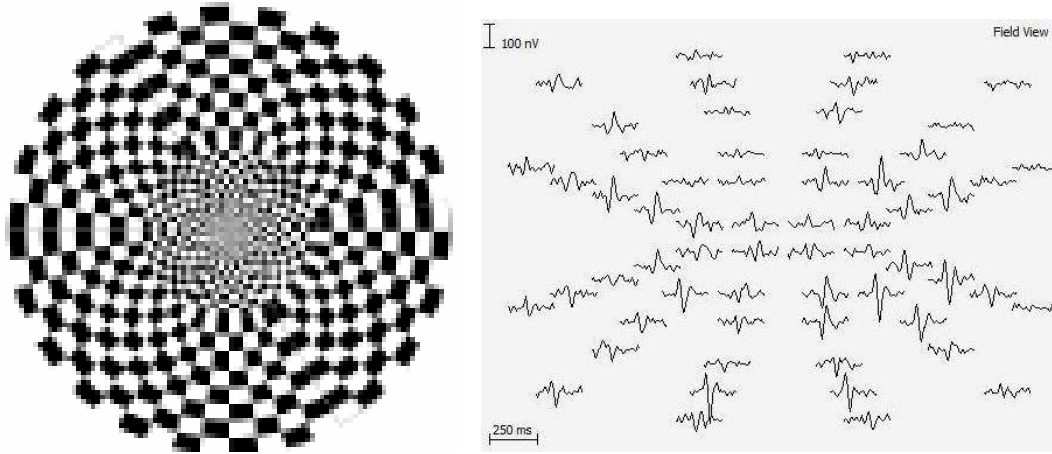
Figure 1.10 The mfVECP stimulus (left) and corresponding trace array (right). Each of the 60 sectors forming the stimulus are used to evoke a response from the visual cortex; these can be seen in the trace array.

As for the standard VECP a checkerboard pattern is used to evoke the responses from the visual cortex. The stimulus for the mfVECP typically comprises 60 sectors, each of which is made up of 16 squares: 8 white; and 8 black. Each of the 60 sectors is controlled by a different pseudo-random binary sequence. As with the mfERG the sequences used to control each of the sectors are unique. These are cross correlated against the recorded data to reveal the 60 local VECP responses, enabling the detection of localised abnormalities.

1.6 Conclusions

This chapter has introduced the visual system and a variety of methods used to assess its integrity. The advantages and disadvantages of these techniques have been considered. Multifocal techniques used to map the function of the outer/mid retina and the visual field have also been discussed. The mfERG is the focus of this thesis therefore the following chapter will study this technique in greater detail.

2 The multifocal electroretinogram

This chapter provides an in depth description of the multifocal electroretinogram (mfERG). The properties of the sequences used to control the stimulus and the techniques required to recover the waveforms from the recorded data are presented, in addition to a number of factors affecting the final responses. These include the type of stimulus, the choice of electrode and the filtering bandwidths selected.

2.1 Hardware

The hardware utilised to record the mfERG can be seen in figure 2.1:
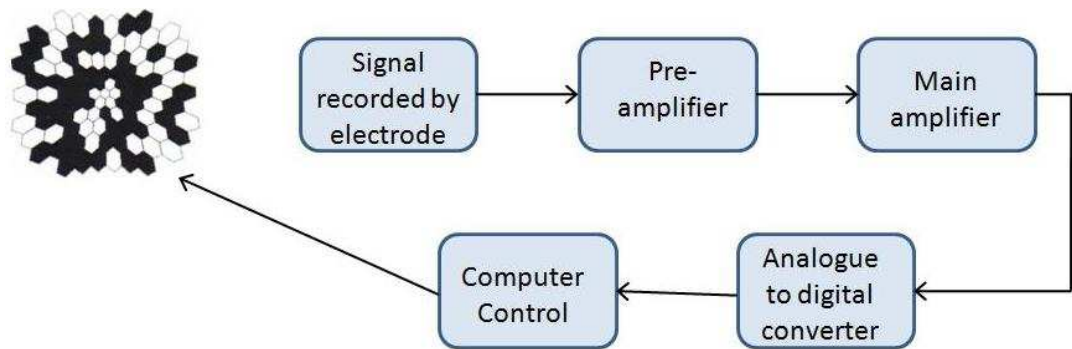
Figure 2.1 An overview of the mfERG hardware (adapted from Keating *et al.* (36)).

The computer generates the sequences which determine the luminance of each stimulus element throughout the test; these sequences are subsequently transferred to the stimulator to enable testing to start. The mfERG signal is recorded from the electrodes, typically placed on the cornea. The magnitude of this signal is very small (an order of tens of nanovolts) therefore amplification of the signal is required to ensure that it is within the operating range of the analogue to digital converter. As a result of the low amplitude of the mfERG signal, it is highly susceptible to noise. The data are finally converted from an analogue to a digital format and delivered to the computer where individual responses are recovered from the signal. To guarantee the accurate recovery of responses the computer must synchronise the rate at which the stimulus is updated with the data acquired at the electrode. A number of commercial mfERG systems are available however a custom built system was used to collect all the data for this thesis.

To ensure that an individual mfERG response corresponds to one particular area of the retina it is essential that the sequences used to control the stimulus are independent from one another (orthogonal).

2.2 The sequences

A group of pseudo random binary sequences, referred to as m-sequences are utilised (18). These are chosen as opposed to random sequences as they have better orthogonality, an essential property for use in the mfERG.

2.2.1 Creating an m-sequence

The generation of a sequence can be demonstrated using a shift register, a circuit which shifts the array stored in it. A primitive polynomial is selected, the terms of which determine the feedback taps of the shift register. Modulo 2 addition (an exclusive OR operation) is carried out on the bits at the tap positions, thus producing a new bit. This bit is shifted into the left hand side of the register. The bit previously at this position is then moved out of the register, forming the first term in the m-sequence. This process is repeated until $2^n - 1$ bits are produced, representing the m-sequence. Figure 2.2 demonstrates this using $x^4 + x + 1$ as the initial polynomial:
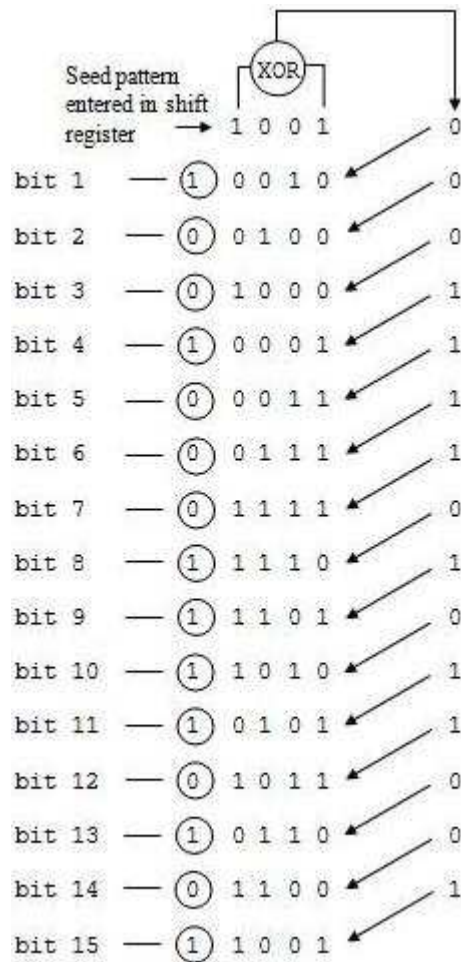
Figure 2.2 The method by which an m-sequence is created (adapted from Keating *et al.* (36)). The primitive polynomial, $x^4 + x + 1$, defines the seed pattern entering the shift register. A modulo 2 addition is performed on this input, producing a new bit, in this case 0. The original bits are shifted left, thus the first bit of the original seed pattern forms the first term of the m-sequence. This is continued until the sequence is 2^n-1 bits long.

2.2.2 Decimation of the m-sequence

To form the set of orthogonal sequences the original m-sequence is decimated. This is demonstrated in figure 2.3. The sequence is filled into the rows of a predetermined number of columns (this must be a power of 2). This process is repeated until the length of each column is 2^n-1 (i.e. 15 in this case). Each resulting column is a sequence which can be used to control an element in the stimulus, hence the number of columns used determines the number of elements available to form the stimulus.

1001
0001
1110
1011
0010
0011
1101
0110
0100
0111
1010
1100
1000
1111
0101

Figure 2.3 Decimation process employed to create multiple sequences from the original m-sequence. The original sequence is used to fill the rows from left to right. This process is continued until each column in 2^n-1 long. The pink highlights the start of the sequence. Each column contains a sequence which can be utilised to control an element of the stimulus.

Each sequence is assigned to a particular element of the stimulus, toggling the luminance between black and white, depending on the state of the sequence (i.e. 0 or 1). It should be noted that each sequence is the same however it starts at a different point.

2.2.3 Properties of an m-sequence

An m-sequence generated using a polynomial of order n has the following properties:

- its length is 2^n-1;

- it contains $2^{(n-1)}$ ones and $2^{(n-1)}$-1 zeros;

- the modulo 2 addition of the sequence with a shifted version of itself produces a third shift of the same m-sequence;

- any bit pattern of length n is unique.

To ensure orthogonality of the sequences the original polynomial must be chosen with care and the shift between the sequences must be greater than the time period of the evoked retinal response. The possibility of using contaminated sequences to control

the mfRG stimulus is increased when the length of the sequence is decreased. Consequently the polynomial utilised to generate the m-sequences for the mfERG is of a much higher order than that shown in figure 2.2. The signal to noise ratio of the recording is also increased by using a higher order polynomial as the resulting sequences are longer thus enhancing recording times. A 15th order polynomial was selected for the department's mfERG system thus a sequence of length 2^{15}-1 was created (32767 steps). The set of m-sequences were subsequently generated by decimating the original sequence over 128 columns thus forming 128 orthogonal sequences. There were therefore unused sequences when using a stimulus of 61 or 103 elements, set ups commonly chosen.

2.3 Cross correlation

The signal measured at the corneal electrode is a combination of the responses from each area of the retina. To obtain the response evoked by a particular element the raw data signal is cross correlated with the m-sequence controlling the element. This is done for each element in the stimulus, enabling an individual response to be derived for every area of the retina stimulated during testing. The two main responses which are reported are the first and the second order response.

2.3.1 First order response

The first order waveform represents the evoked response when presented with a high luminance. It is attained by:

Σ responses to a white element – Σ responses to a black element

Segments of the raw, uncorrelated data, for example 300ms, are added to or subtracted from a memory buffer to obtain the final response. In this case the data are added to the memory buffer if the first bit in the segment is 1. If the first bit is 0 the data segment is subtracted from the memory buffer. This is done for the entire data train to acquire the localised response. The final signal is therefore formed from 2^n-1 overlapping segments, $2^{(n-1)}$ of which are added to the memory buffer and $2^{(n-1)}$-1 of which are subtracted. An example of such a response is shown below (left). This process is carried out for every element thus returning a first order response corresponding to each stimulating element. This can be seen in figure 2.4 (right):
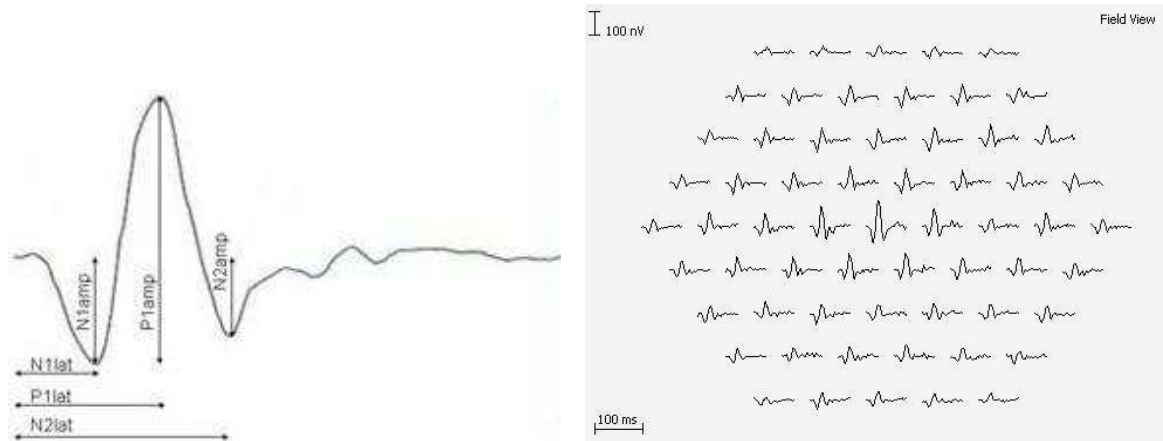
Figure 2.4 An individual mfERG response (left) and a trace array comprising 61 responses (right). The principal measurements (amplitude and latency of N1, P1, and N2) are shown on the individual waveform.

2.3.2 Second order response

The second order response represents the evoked response resulting from a change (Δ) in luminance. To acquire this waveform the data segment is added when there has been a change of state (i.e. 0 to 1 or 1 to 0) and is subtracted when there is no change (i.e. 1 to 1 or 0 to 0). The response is therefore obtained by:

$$\Sigma \text{ responses when } \Delta \text{ luminance} - \Sigma \text{ responses when no } \Delta \text{ luminance}$$

Again to return an array of second order responses the cross correlation process is performed for every element in the stimulus.

2.3.3 Origins of the response

It can be seen that the first order mfERG waveform has an initial trough followed by a peak, similar to that seen in the conventional ERG. However, unlike the conventional ERG the mfERG waveforms are not temporally intact but are instead a composite response comprising components from previous and subsequent stimuli (37;38).

A study by Hood *et al.* (39) compared the conventional full field ERG with the mfERG in an attempt to increase the understanding of the mfERG. They decreased the rate of mfERG stimulation and showed that there was good correlation between the a-wave of the ERG and the N1 of the mfERG response. The ERG b-wave and P1 of the mfERG waveform also corresponded well. When the stimulation frequency was increased to 75Hz, the frequency commonly used to perform the mfERG, the

relationship between the ERG a-wave and the N1 of the mfERG remained intact. The correlation between the ERG b-wave and the P1 component of the mfERG was however less stable at the higher stimulation frequency, indicating that the P1 component is more influenced by non-linear retinal processing mechanisms.

A number of animal studies have been conducted to investigate the cellular origins of the mfERG. These involved using pharmacological blocking agents to either enhance or inhibit activity from particular layers of the retina (40-42). It was suggested that the outer and the mid retina are the main contributors to the mfERG. Reports have also hypothesised that the inner retina contributes to the second order response (43;44). Keating *et al.* (37) however argued that the first and second order responses are recovered from the same data, the only difference being the way in which the responses are added and subtracted. Components seen in one should therefore be seen in the other. It is predominantly the first order responses which are studied when reporting the mfERG test therefore these will be utilised for the remainder of this thesis.

The recovered mfERG responses are influenced by the experimental protocol employed to acquire them. A number of these will be discussed, including the choice of stimulator, the type of electrode utilised and the frequency bandwidth of the amplifier.

2.4 Type of stimulator

The mode of stimulation affects the mfERG response characteristics. A standard cathode ray tube (CRT) computer monitor is commonly used however a back projected liquid crystal display (LCD) was the principal mode of stimulation throughout this study. Both of these will be considered in addition to the scanning laser ophthalmoscope (SLO) and an array of light emitting diodes (LED), each of which can be used as a mode of stimulation.

2.4.1 CRT and LCD

The main difference between a CRT and an LCD stimulating device is their output during a period of high luminance. This can be seen below where a photodiode was used to measure the luminance of an element, illuminated first by a CRT and

subsequently by an LCD device. The element was driven at 75Hz by the sequence 1 1 0; the period between each step of the sequence was 13.3 ms.
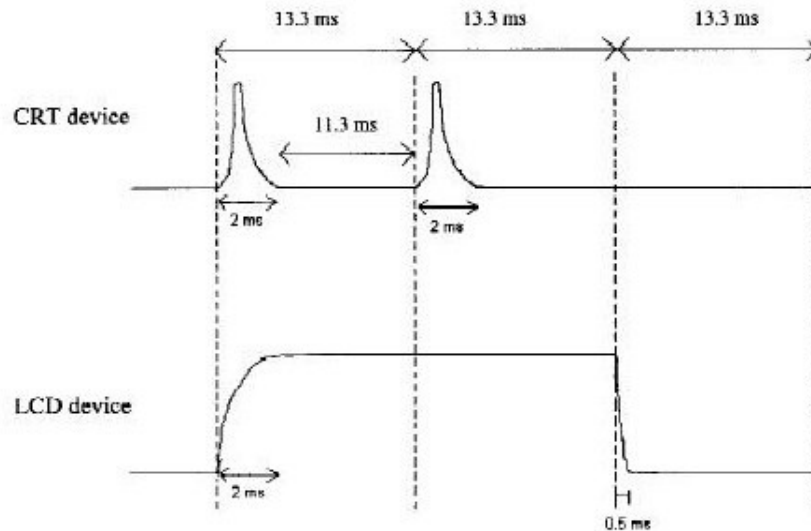
Figure 2.5 The output of a photodiode when stimulated with a CRT device (top) and an LCD device (bottom) by the sequence 1 1 0 (adapted from Parks *et al.* (45)). When there is a 1 in the sequence the CRT produces a short pulse (2ms) and then a dark period (11.3ms). The LCD differs to this; it produces a period of high luminance for the entire on-state of the sequence. When successive 1s are present in the sequence a CRT device produces a pulse of high luminance for each 1, whereas the LCD remains at a high luminance for the duration of the on-state.

When a pulsed raster based method such as the CRT is used a short pulse of light is produced, followed by a longer period of low luminance. In contrast, when using a square-wave based method such as the LCD the luminance remains high throughout the frame. The transition to a high luminance is the main physiological contributor to the mfERG response (45). It can be seen that when two consecutive 1s are present in the sequence two separate responses would be evoked by the CRT device while the LCD would produce only one response at the start of the period of high luminance. The type of stimulator used therefore impacts on the final responses acquired. The following simplified schematic diagram illustrates the differences in the final cross correlated responses when using the two types of stimulating device. The four possible consecutive steps of the m-sequence (1-1, 1-0, 0-1 and 0-0) have been shown in isolation to aid this explanation:
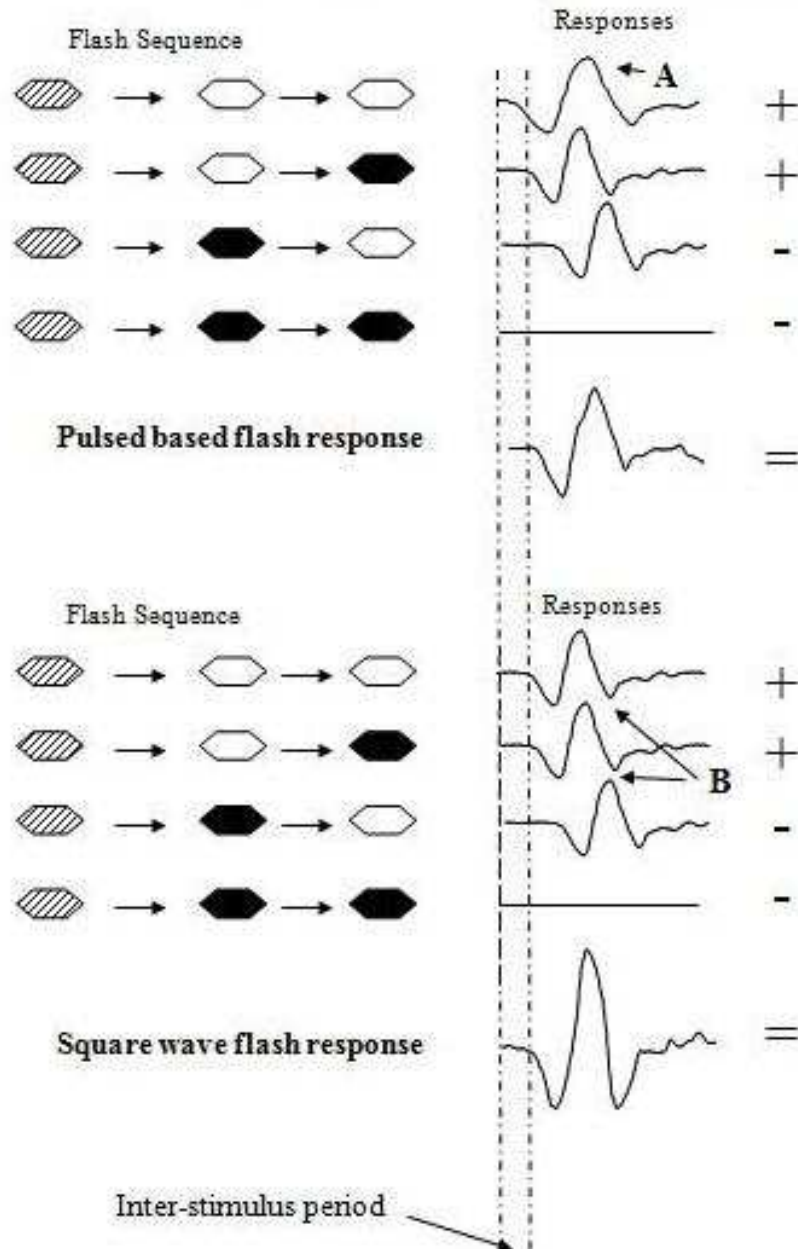
Figure 2.6 A comparison of the cross correlated responses acquired using a CRT stimulator (top) and an LCD stimulator (bottom) (adapted from Parks *et al.* (45)). The principal difference in the final responses acquired using LCD or CRT devices can be attributed to the differences in the illumination profile of each when there are successive 1s in a sequence (figure 2.5). 'A' shows the elongated response resulting from consecutive 1s in the sequence when using a CRT device. 'B' demonstrates why a larger N2 component is evident when using an LCD stimulus; the troughs from the 1-1 and 1-0 part of the sequence align with the peak from 0-1.

Two main differences can be seen: the first is the final CRT response is more elongated than that acquired using the LCD, while the second is the larger N2 amplitude of the LCD generated waveform. These variations can mainly be attributed to the illumination profiles of each device (figure 2.5) and the way in which the retina responds when there are consecutive 1s in the sequence. When there are successive 1s in the sequence the response evoked by the CRT is stretched (A in figure 2.6) relative to that acquired from the LCD. This is the result of separate flashes stimulating the retina within the response time period. The increased N2 amplitude obtained when

stimulating with an LCD device can be attributed to the alignment of the troughs for the 1-1 and 1-0 parts of the sequence with the positive peak of the 0-1 combination (B in figure 2.6). N2 is therefore dominated by the interaction between consecutive stimuli and the P1 component of the preceding stimulus (37). Responses evoked using a pulsed based stimulus have a greater non-linear contribution. They also have a slightly larger amplitude than those acquired from a square-wave based system. This is because the CRT device evokes a response for both the 0-1 and 1-1 steps of the sequence, while the LCD stimulus only evokes a response for the 0-1 parts of the m-sequence. Consequently the contributions to the overall mfERG response are increased when using a CRT device, thus increasing the response amplitude. On the other hand, those evoked using a square-wave based stimulus include information on the retina's ability to recover from a more extensive period of high illumination.

2.4.2 SLO and LED

SLO, like CRT, is a raster based technique. It runs at a slower frame rate than a CRT and presents the stimulus directly onto the retina. It allows the fixation of the patient to be monitored which can be advantageous when testing patients with poor fixation, those who fixate eccentrically or those suspected of malingering (46-48). The field of stimulation is however smaller than that of other stimulating devices meaning that assessment of more peripheral problems is not possible.

With using an LED stimulus the luminance level is constant and the width of a pulse of light can be chosen by the user. This has led to it being used by several researchers (37;49) and has led to a greater understanding of the components of the first and second order responses (37).

2.5 Type of electrode

In addition to the method of stimulation affecting the final responses it has been shown that the type of electrode used impacts on the recovered signals. The electrode utilised for the mfERG is the same as that used for conventional ERG recordings. Researchers have used several different types of electrode to measure electrophysiological responses, each of which offers its own advantages and disadvantages. The electrodes can be divided into three main categories: contact lens; lid hook; and fibre electrodes. Examples of these can be seen in figure 2.7:

Figure 2.7 Different types of electrode used for ERG recordings. From left to right: contact lens; lid hook; and fibre electrode (adapted from www.webvision.com).

When selecting an electrode a number of factors should be considered, including the signal to noise ratio, patient comfort and the reliability and repeatability of the recording. Contact lens electrodes are commonly used when recording the mfERG (19;50;51). When compared with other electrodes they have been shown to have the largest signal amplitude (52) however they can be uncomfortable and have a risk of corneal and conjunctival abrasions (53).

Lid-hook electrodes are inserted into the lower fornix of the eye and are bent over to lie on the cheek (54). They are better tolerated by patients and do not interfere with the optics of the eye which is important when recording the mfERG. This has led many people to use them as an alternative to the contact lens electrode (55;56). They are however very flexible and can slip out of the eye.

The fibre electrode consists of a number of conductive threads which are placed either in the tear film on the surface of the cornea or in the lower fornix (57). The largest recordings are obtained from the former location however the impact of eye movement and blink artefacts is greater. In general fibre electrodes are well tolerated by patients and they do not interfere with the eye's optics. The higher patient tolerance has resulted in this being a popular electrode for recording the mfERG (58;59).

A study carried out by Mohidin *et al.* (60) compared four types of electrode when recording the mfERG. These included the JET contact lens, the gold foil (type of lid hood electrode), the fibre electrode and the c-glide (lid hook) electrode. It was found that responses acquired from the contact lens had the largest amplitude and that those obtained using the c-glide electrode had the greatest variability.

2.6 Amplifiers and filter bandwidth

In addition to the choice of stimulus and electrode, the amplification and filter bandwidth also affect the final mfERG response. The uncorrelated data are both

amplified and filtered prior to being converted from an analogue to a digital format. Standard differential physiological amplifiers are typically used for this purpose. These should have a high gain to enable the production of clear signals however saturation of the signal must be avoided; a gain of 100000 is commonly used (61;62). The amplifiers should also have a high common mode rejection ratio. This is a measure of an operational amplifier's ability to reject signals common to both inputs, such as noise.

To decrease both patient and experimental noise high and low pass filters are used when recording the raw mfERG data signal. Cut off frequencies for the high pass and low pass filters are typically 10Hz and 100Hz (23;63;64) or 10Hz and 300Hz (65-67). It was however shown by Keating *et al.* that the cut off frequency chosen for the high pass filter has a significant affect on the shape of the cross correlated responses; waveforms acquired using 10Hz were distorted relative to those acquired using a 3Hz high pass filter (68). This was reiterated by Seeliger *et al.* who reported that negative mfERG responses were significantly distorted when using a cut off frequency of 10Hz as opposed to 2Hz (69).

2.7 Patient factors

In addition to the choice of hardware affecting the final responses, various patient factors can impact on the cross correlated waveforms, a number of which are described in section 2.7.

2.7.1 Pupil size

The diameter of a patient's pupil affects the mfERG response as the amount of light entering the eye is determined by the size of the pupil. A study by Gonzalez *et al.* showed that a smaller diameter results in a decreased P1 amplitude and an increased P1 latency (70); dilation is therefore recommended (71). Pupil size can be affected by iris colour and age.

2.7.2 Age

An increase in latency and a reduction in response amplitudes have been reported in older people; (65;72). In the first of these two studies the age of the subjects ranged from 9 to 80 years old while for the second study two distinct groups were utilised: one

incorporated people ranging from 19-30 while the other included people ranging from 60-74. A number of studies have demonstrated that the central responses are more affected by age than those positioned peripherally (73;74).

2.7.3 Patient compliance

Both blinking and losses of fixation can result in saturation of the signal. This decreases the portion of raw data available for cross correlation thus impacting on the quality of the final responses. Patients are therefore encouraged to minimise these during testing. It is also essential that a patient maintains fixation throughout a test to ensure spatial accuracy of the test. A number of approaches have been successfully used to monitor fixation during testing including the use of an SLO stimulus (10), an infrared television fundus camera (11) and an eye tracking device described by Chisholm *et al.* (75).

2.8 Aims of thesis

The objectivity of the mfERG and the spatial information it provides make it an attractive test for assessing the function of the outer/mid retina. The responses do however have a low amplitude (in the order of nanovolts). Consequently they are easily contaminated by noise, especially when response amplitudes are decreased as a result of compromised retinal function. Differentiating between normal and compromised retinal function can for example become difficult when waveforms are recovered from a noisy recording. The analysis process can therefore be both difficult and subjective, sometimes resulting in inconsistencies in the interpretation of a test. The expansion of this technique is therefore limited, as experimental and analytical consistency is essential in many circumstances, for example in multicentre clinical trials. An intelligent system capable of analysing a mfERG recording in an objective, accurate and consistent manner would therefore be advantageous. The aim of this project was to develop a technique to achieve this.

An objective method for grading the recording quality both during and after an examination would be beneficial, both to the operator and to those subsequently reporting the test. This would enable the operator to be warned of problems during the testing session and would provide the person reporting the test with a greater knowledge of patient cooperation. Of those recordings deemed to be of a sufficient

standard to be reported, the system should be able to differentiate between waveforms with and without a physiological response. For those representing an area of retinal function the response should ideally be classified as normal or abnormal, both in terms of its amplitude and latency. Ideally the system would also have the ability to compare trace arrays from sequential visits, allowing a patient's condition to be monitored.

A number of methods have been proposed by researchers to decrease the subjectivity of the mfERG analysis however there are limitations associated with a number of them; these will be discussed in the following chapter. It was also decided to investigate techniques applied successfully to the analysis of other similar physiological signals, with a view to applying these to the interpretation of the mfERG to realise the objective of this research. These will be discussed in chapter 3.

2.9 Conclusions

The creation and properties of the sequences used to control the mfERG stimulus have been described, in addition to the recovery of the individual responses from the raw data signal collected at the electrode. Factors affecting the final cross correlated responses have also been considered. Finally, the limitations of the mfERG have been discussed in addition to the principal aims of this thesis.

3 Current techniques used to analyse the mfERG and other physiological signals

It was discussed in chapter 2 that analysis of the mfERG can be both difficult and subjective when a waveform is obtained from a poor recording, or from an area of retinal dysfunction. Differentiating between no significant retinal function and a decreased response can for example be problematic; furthermore, determining the amplitude and the latency of the main components (N1, P1 and N2) can be subjective. Consequently a number of methods have been proposed by researchers to improve the objectivity of mfERG analysis. These will be discussed in addition to their limitations. Approaches used to analyse other physiological signals such as the electroretinogram (ERG), the visual evoked cortical potential (VECP) and the pattern electroretinogram (PERG) will also be described.

3.1 Techniques employed to analyse the mfERG responses

Techniques used to distinguish waveforms with a physiological response from those with no function are described first.

3.1.1 Peak to tough

Hood *et al.* (19) defined a minimum P1 amplitude criterion: waveforms with a peak to trough value of less than a defined value were said to have no significant retinal function. This method is of limited use as the presence of noise can impact greatly on the response amplitude (18).

3.1.2 Signal to noise ratio

An alternative method is to calculate the signal to noise ratio (SNR). Different ways of deriving the SNR have been proposed.

3.1.2.1 Noise window and signal window

Zhang *et al.* calculated the SNR for the mfVECP by comparing the latter part of the cross correlated wave with the earlier part (76). The waveform was effectively split

into two components: a signal window; and a time window. This can be seen in figure 3.1:
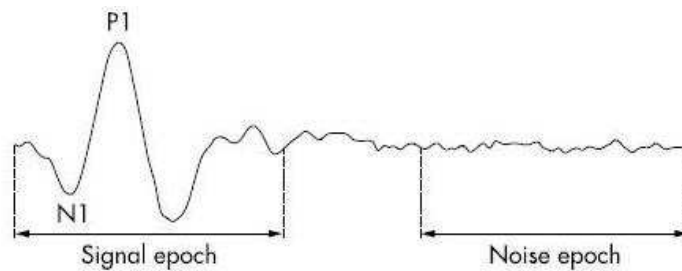
Figure 3.1 Calculating the SNR using a noise window and a signal window (adapted from Han *et al.* (77)). The signal epoch represents the signal window while the noise epoch is the noise window.

The assumption was that the signal window contained both noise and signal while the noise window comprised only noise. The following equation was used to calculate the SNR for each waveform, i, in the trace array:

$$SNR_i = \frac{RMS(signalwindow)_i}{RMS(noisewindow)_i} - 1 \qquad (3.1)$$

where RMS is the root mean square.

An alternative method was also described where the denominator was the average noise window RMS for all waves in the trace array:

$$SNR_i = \frac{RMS(signalwindow)_i}{RMS(noisewindow)_{average}} - 1 \qquad (3.2)$$

The latter was found to yield a lower false positive rate (i.e. stating there was a physiological response when there was not) and has since been applied to the analysis of the mfERG (77-79). This technique however assumes that the data in the noise window contain no contribution from the evoked retinal response.

3.1.2.2 Dead sequences

An alternative method to calculate the SNR utilises unused m-sequences, termed dead sequences. When creating m-sequences to control a 60 element mfVECP or a 61 element mfERG stimulus at least 64 m-sequences are generated using the decimation process; this is dependent on the number of columns over which decimation is performed (refer to section 2.2.2); consequently there are unused (dead) sequences. By

cross correlating one of these dead sequences with the raw, uncorrelated data it can be assumed that the recovered response contains only noise. The following equation can therefore be used to calculate the SNR:

$$SNR_i = 20\log_{10}\frac{RMS(wave)_i}{RMS(deadsequencewave)} \qquad (3.3)$$

MacFarlane *et al.* (80) used this method to calculate the SNR for mfVECP waveforms and showed that there was a good correlation between this approach and the Zhang *et al.* method when healthy people were tested. This technique has also been applied to the analysis of the mfERG by Keating *et al.* (81). Healthy volunteers were tested; the length of the m-sequence and the stimulus frequency were both varied and it was shown to be a viable method for calculating the SNR. It has not as yet been applied to the analysis of clinical mfERG data.

Several approaches to assess the amplitude and latency of a response have been proposed, one of which is described in section 3.1.3.

3.1.3 Minimum and maximum values

The minimum and maximum turning points can be used to locate N1, P1 and N2 when analysing a waveform such as that seen in figure 3.2:
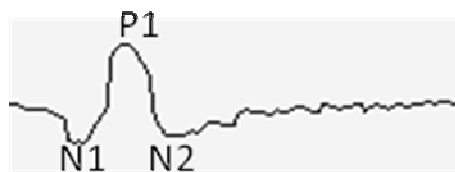
Figure 3.2 A mfERG response which is relatively simple to analyse as it contains three clear turning points corresponding to N1, P1 and N2.

It is however often the case that P1 is not the maximum value, and that the two main troughs are less clearly defined than those seen in figure 3.2. This can arise when a recording is contaminated with patient noise or when retinal function is compromised. The following waveform demonstrates this:
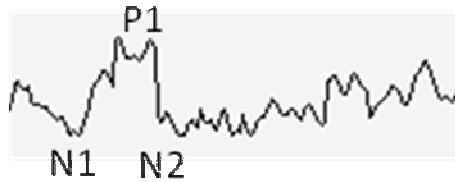
Figure 3.3 A difficult mfERG response to analyse as it contains no clear turning point at P1 and the troughs, N1 and N2, are less clearly defined. Locating P1 is more subjective in this instance as it does not correspond to the maximum value.

3.1.4 Scalar product

An alternative approach to assess the response amplitude and latency is to calculate the scalar product. This provides a measure of the deviation from an ideal response and has been shown to be sensitive to changes in both amplitude and latency (18). It is calculated by:

$$scalar product = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + ... + a_n b_n \qquad (3.4)$$

where a is the mfERG waveform vector, b is the ideal response vector and n is the number of data points in the vector. Increases in latency or reductions in amplitude are reflected as a reduction in the scalar product value. This technique has a relatively good immunity to noise however problems can arise if an inappropriate template is selected for the ideal response. Keating *et al.* showed that the scalar product can detect a change in latency but argued that the scalar product is less sensitive than measuring the actual latency value (82). It has been shown that the P1 timing is particularly important when studying the diseased retina, for example in conditions such as retinitis pigmentosa (19;20) and cone dystrophies (83) therefore a technique should ideally find the actual P1 latency value, rather than detect a change in latency.

3.1.5 Template fitting

Hood *et al.* (84) proposed a method for measuring the amplitude and the latency of a mfERG response. It involved stretching the template of a normal wave in time as well as varying its amplitude until the best fit to the local patient response was found. This was established using the least-squares fitting method. It has since been used to analyse the data in many studies (23;85;86). This method however relies on the patient waveforms being a similar shape to the template response which is not always the case. Furthermore, by stretching the template the later components on the waveform

are more delayed, relatively, than the earlier components which may not always be the case in clinical waveforms.

The possibility of shifting a template in time has also been investigated. In this case all components are delayed by the same value. It has however been shown that the stretching method is more sensitive than the shifting method, both in patients with retinitis pigmentosa (84) and in those with diabetic eye disease (23;87) although Schneck *et al.* argued that the template stretching method does not fully represent the mfERG response in the diabetic eye; the later components were generally not more delayed than the earlier components.

3.1.6 Spatial averaging

Post processing of the cross correlated responses has also been used with a view to easing the interpretation of the mfERG signals. One such technique is spatial averaging; this involves averaging the response from each hexagon with the mean of its surrounding waveforms. It is commonly used to improve the appearance of the mfERG responses (24;88) but it greatly decreases the spatial resolution of the test, one of the greatest advantages of the mfERG technique.

To avoid compromising spatial resolution artefact removal and digital filtering techniques can be utilised with a view to reducing the influence of noise on the waveform.

3.1.7 Artefact removal

One widely used commercial mfERG device, the VERIS system (developed by Electro-Diagnostic Imaging, Inc. USA), has a built-in algorithm which removes artefacts caused by patient blinking and eye movements. The cross correlation process used by the VERIS system differs to that described in section 2.3; it is performed using the complete uncorrelated data set as opposed to small sections of data. Consequently each of the recovered mfERG responses are distributed along the correlated data. The data between each of these correlated responses can be attributed to patient noise; this artefact removal method zeroes these parts of the data. A reverse cross correlation is then performed, producing an alternative uncorrelated data set. This is compared with that of the original raw data signal; where there are significant differences between the two, the original signal is replaced with the new signal. Cross correlation is then

repeated; the appearance of the resulting responses should therefore be improved. This technique has been applied to the data in many studies (24;74;89) however it has been shown that the shape of the cross correlated responses can be distorted when using this feature (90).

3.1.8 Digital filtering

Seeliger *et al.* (63) designed a filter for post processing of the mfERG data based on the frequency spectrum of the cross correlated waveforms. They found that the principal frequencies forming the waveform were within the 10-60Hz range therefore a bandpass filter (one which passes frequencies within a specified range and attenuates frequencies outwith this region) was designed with lower and upper frequencies of 9.4 and 56.4Hz respectively.

Bock *et al.* (91) investigated using a 50Hz notch filter to minimise the effect of electrical noise during a recording. A notch filter is a bandstop filter (one which attenuates frequencies within a specified range and passes all other frequencies) with a narrow stopband, in this case concentrating on 50Hz. When the first order cross correlated responses were compared with those acquired using standard filter settings of 10-300Hz it was shown that the shape of the waveform was comparable however the P1 component was moderately delayed therefore this method should be used with caution.

3.2 Analysis of electrophysiological signals in vision

It is also of interest to consider analysis approaches employed in other areas of electrophysiology, with a view to applying them to mfERG analysis if appropriate as similar problems are encountered when interpreting the signals. Three techniques have been selected: 1) analysis of signals in the frequency domain; 2) artificial neural networks; 3) wavelet analysis. These were chosen as they have each showed promising results when applied to the analysis of other physiological signals, thus may have a potential role in reducing the subjectivity of the mfERG interpretation process.

3.2.1 Studying signals in the frequency domain

Physiological responses are often studied in the frequency (Fourier) domain as it enables information to be extracted which is not available in the time domain. This is the case when reporting a steady-state response (one obtained at a higher rate of stimulation meaning that responses evoked by successive stimuli overlap one another). Fourier analysis has been utilised to ease the analysis of the steady-state PERG (17), the steady-state VECP (92-95) and the oscillatory potentials of the ERG (96). It was shown by Meigan *et al.* that studying the frequency components of steady-state responses can be useful for quantitatively deciding if a waveform contains a true physiological signal (97). The stimulus frequency is known in each case therefore when a recording is viewed in the frequency domain all frequencies which are not a multiple of the stimulus frequency are known to be noise. This is useful for increasing the objectivity of the analysis process.

Klistorner *et al.* used the Fourier domain with a view to reducing the intersubject variability seen in the mfVECP (98) by viewing the frequency components of the raw electroencephalogram (EEG), the electrical activity of the brain data. The VECP data were scaled according to the Fourier spectrum of the EEG signal, resulting in less variability between people, hence widening the scope of the technique for serial measurements and clinical use.

3.2.2 Artificial neural networks

Artificial neural networks (ANNs) are an attempt to emulate tasks unique to the human brain, and like humans they learn by example. Networks can ultimately be trained to perform a specific function such as pattern recognition, predicting future events based on existing data or the categorisation of data. The idea is that with sufficient training a network should be able to provide an accurate output when presented with new data. They have been used for a wide range of functions, for example predicting cancer survival rates (99-101) and classifying physiological signals such as the electrocardiogram (ECG), the electrical activity of the heart (102;103).

Guven *et al.* have recently used ANNs to classify the PERG (104), the VECP (105) and the EOG (106) into one of two categories: normal; or abnormal. Accuracy rates of 98%, 97% and 94% respectively were achieved. ANNs have also been applied to the analysis of the ERG by Lipoth *et al.*; responses were classified with a 95% success rate

(107). In addition to classifying signals Fisher *et al.* demonstrated that ANNs can be used to identify the latency of PERG responses by cursoring the main peak; a high performance was achieved even when signals were contaminated with large amounts of noise (108).

3.2.3 Wavelet analysis

Wavelet analysis can be utilised to decrease the noise present in a signal. A series of high pass and low pass filters are used to decompose the signal: the low pass filters produce approximations of the data while the high pass filters provide the detail. Wavelet coefficients are obtained for both the detail and the approximations. Thresholding is then carried out on the detail data: all wavelet coefficients less than a defined threshold are set to zero. The signal is then reconstructed using the remaining wavelet coefficients. The idea is that the reconstructed signal should be less noisy while maintaining the shape of the original signal. This technique has been applied successfully to the analysis of physiological signals such as the ECG (109), the EEG (110) and the VECP (111).

Wavelet analysis has also been used to classify signals, for example the EEG (112) and the myoelectric signal, the response representing neuromuscular activity (113). Wavelets have since been employed to differentiate between normal and abnormal transient PERG responses (those acquired using a lower stimulus frequency therefore there is no overlap of responses to successive stimuli) (114;115). These studies demonstrated that this technique has the potential to improve the objectivity of the analysis process of the PERG.

A recent study used the wavelet transform for the analysis of the mfERG (116) with a view to identifying glaucoma markers. Differences between the wavelet decomposition of recordings with and without glaucoma were observed thus making the distinction easier than using conventional analysis methods.

3.3 Overview of techniques employed for this research

It is evident that many approaches have been utilised to improve the objectivity of mfERG analysis but that there are limitations associated with these. Methods applied successfully to the interpretation of other physiological signals have been discussed;

this thesis investigates the applicability of these to the analysis of the mfERG with a view to improving the consistency of mfERG analysis.

The possibility of using the frequency domain to assess the quality of a recording and to differentiate an area of functioning retina from a dysfunctional region is discussed in chapter 4. This technique was chosen as the Fourier domain can be used to differentiate signal from noise; this is a useful property to realise these two aims.

Chapter 5 investigates the ability of ANNs to classify mfERG waveforms as delayed, within normal timing limits or as no significant response; neural networks were selected as their ability to classify data should enable the mfERG data to be categorised in a consistent manner.

Two methods utilised to calculate the SNR were described in section 3.1.2; that using a noise window and a signal window, and that using a waveform acquired from an unused m-sequence to represent the noise. The first of these has been applied to the interpretation of the mfERG however the second approach has thus far only been utilised for the analysis of the mfVECP and mfERG data acquired from healthy control subjects. The ability of these two approaches to distinguish an area of functioning retina from a region with no significant retinal function is therefore assessed in chapter 6 with a view to finding the optimal method. Digital filtering and wavelet techniques are also studied in chapter 6; the aim was to improve the appearance of the mfERG data prior to analysing it, thus easing the interpretation process. Finally the responses are analysed using curve fitting techniques as opposed to defining the maximum and minimum values or fitting templates to the data; the aim was to categorise the responses as normal or abnormal, both in terms of their amplitude and latency.

To optimise the performance of the system, each approach is combined in chapter 7 to form a multilayered system.

3.4 Collecting normative data

Prior to studying the potential of each of these techniques it was essential to establish a normal range for the particular experimental set-up used in the department as it was shown in chapter 2 that factors such as the type of stimulus and the choice of electrode

impact on the final mfERG responses. Normative data were collected using the following protocol. 20 healthy subjects ranging in age from 18 to 72 were tested.

3.4.1 Methods: mfERG protocol

3.4.1.1 Acquiring the data

Recordings were conducted using a custom built mfERG system designed to stimulate a 90^0 field of vision and were carried out in accordance with the International Society for Clinical Electrophysiology of Vision (ISCEV) guidelines (71). Tests were performed binocularaly with pupils maximally dilated using Tropicamide (1%). DTL fibre electrodes (Diagnosis LLC) were used to measure the evoked response from each eye, while the ground and reference electrodes were attached to the forehead and the outer canthi respectively. The patient fixated on a cross located at the centre of an array of 61 empirically scaled hexagons, each controlled by an independent m-sequence. Patient fixation was monitored by the operator. The orthogonal sequences were created by decimating a 15 bit m-sequence over 128 columns. The stimulus was presented by a back projected LCD system at a rate of 75Hz, while data were sampled at 1200Hz. Dual bandpass filters of 10-100Hz and 3-300Hz were selected however all analysis was done using the 10-100Hz data as this is more commonly used in the literature. A 12 bit analogue to digital convertor with a gain of 100000 was used. The recording period was split into 16 equal segments, each lasting approximately 30 seconds, to improve patient tolerance.

3.4.1.2 Analysis of the data

The amplitude and the latency of the N1 and P1 components are the principal measurements used when analysing mfERG responses (71) therefore normal ranges were calculated for each of these. When interpreting the responses it is typical to average the responses in each concentric ring (71); this can be seen for one of the participants:
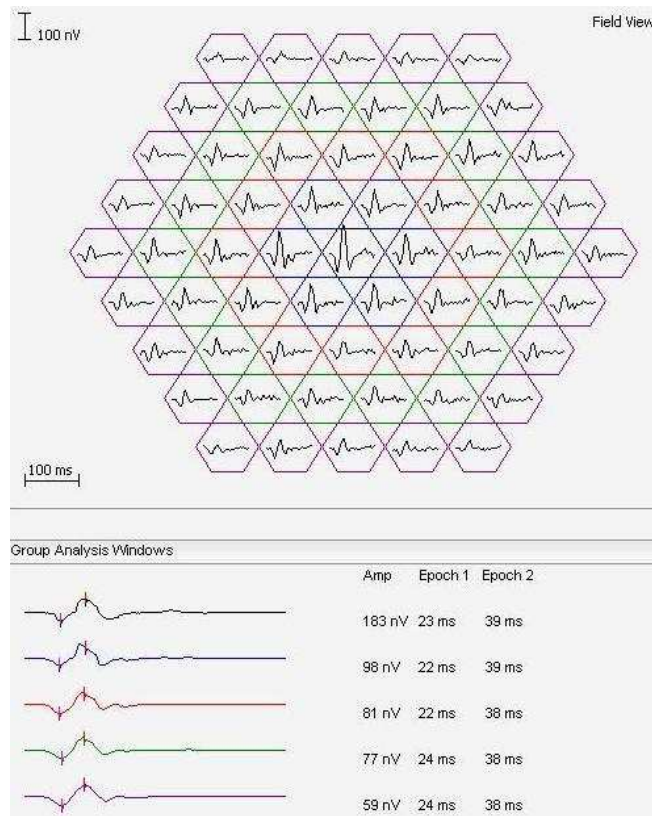
Figure 3.4 Analysis of the mfERG using concentric rings. The average P1 amplitude (amp), N1 latency (epoch 1) and P1 latency (epoch2) are shown for each concentric ring.

This mode of analysis was utilised for both eyes for each of the 20 volunteers thus 40 N1 and P1 amplitude and latency values were obtained for each ring. The 5th and 95th percentile were subsequently calculated for the four parameters, yielding a normal range for each. These were calculated as opposed to standard deviations as the mfERG data are typically non-parametric (71).

3.4.2 Results

The following table shows the normative range for the N1 and P1 amplitude and latency of the central mfERG response and each of the concentric rings:

	N1 amplitude (nV) 5th -95th percentile	N1 latency (ms) 5th -95th percentile	P1 amplitude (nV) 5th -95th percentile	P1 latency (ms) 5th -95th percentile
Central response	33-71	23-26	86-180	38-42
Ring 1	26-54	22-25	66-145	37-42
Ring 2	23-56	22-26	56-133	37-42
Ring 3	20-45	24-25	51-111	37-42
Ring 4	13-28	24-26	30-69	38-42

Table 3.1 Normative data for the mfERG. The 5th-95th percentile has been shown for both the N1 and P1 amplitude and latency.

The overall trend for the N1 component was a reduction in amplitude from the central response to ring four. The 95th percentile for the latency varied slightly, with values of 25ms or 26ms for each of the rings. As was seen with N1, the P1 amplitude of the central waveforms was larger than that of the peripheral responses despite the scaling of the stimulus. The latency was however consistent across the field. These results imply that the latency of the P1 component is slightly more consistent than that of N1; this has caused some people to report only the P1 data (23;117). It was therefore decided to concentrate on the latency and amplitude of the P1 component throughout this thesis: all responses with a P1 latency greater than 42ms, irrespective of their location in the trace array will be classified as delayed. The position of a response in the trace array must however be considered when determining if it is normal or decreased in amplitude. For central waveforms a P1 amplitude of less than 86nV can be considered to be decreased while those in rings one, two, three and four with an amplitude smaller than 66nV, 56nV, 51nV and 30nV respectively are compromised. It is difficult to compare these values directly with studies in the literature as the experimental set up used is slightly different in each case, for example the type of stimulus or electrodes used. It has however previously been reported that the P1 latency does not vary greatly across the retina (56) and that the response density decreases with eccentricity (18;56), trends which were observed in the current data.

3.5 Conclusions

Approaches currently employed for the analysis of the mfERG have been discussed, in addition to their limitations. Techniques successfully applied to the interpretation of physiological signals other than the mfERG have also been considered; these included artificial neural networks, analysis of signals in the frequency domain and wavelet analysis. The possibility of applying these techniques to the analysis of the mfERG is investigated in the following chapters. Normative data for the mfERG were also presented.

Chapter 4 studies the potential role the Fourier domain has to play in the assessment of recording quality and in the classification of mfERG data as 'response' or 'no response'.

4 Analysis of the Fourier domain profile

It was stated in chapter 2 that one of the main aims of this thesis was to develop a method for grading the recording quality of the mfERG, both during and after a test. It was also important for the system to differentiate a physiological response from a waveform with no significant retinal function. This chapter investigates the possibility of using the frequency domain to achieve each of these objectives; this approach was selected as the frequency domain is commonly used to differentiate signal from noise (97;118-120), a valuable property to enable these two aims to be realised. The frequency components of the raw, uncorrelated data signal were examined initially with a view to establishing a method for assessing the recording quality; an automated grading system was subsequently developed based on the findings. To distinguish a retinal response from an area of retina with no function, cross correlated responses were studied in the frequency domain.

When studying the uncorrelated data in the Fourier domain the frequency profile inherent to the mfERG stimulus was revealed. The experimental factors determining this were investigated. Finally the impact of compromised retinal function on the Fourier profile was studied.

4.1 The Fourier transform

Fourier analysis is used to view the frequencies embedded within a signal. It is based on the principle that any signal acquired in the time domain can be decomposed into sinusoidal components. This process is achieved by the use of Fourier transforms and has been used extensively in many scientific and engineering fields including image processing, signal processing, acoustics and communications engineering (121;122). To transform a function from the frequency to the time domain, inverse Fourier transforms are performed.

4.1.1 Continuous Fourier transform

The continuous time Fourier transform transforms a function $x(t)$ which is continuous in the time domain to one which is continuous in the frequency domain $X(u)$, using

$$X(u) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ut} dt \qquad\qquad (4.1)$$

where u is the frequency measured in Hz and t is the time measured in seconds.

$X(u)$ is a complex function. The magnitude of each frequency component (i.e. the relative contribution of each) is found, in addition to the phase shift of each frequency.

The inverse transform is

$$x(t) = \int_{-\infty}^{\infty} X(u)e^{j2\pi ut} du \qquad\qquad (4.2)$$

4.1.2 Discrete Fourier transform

It is not always the case that data are continuous. When working with data stored in a computer, for example, the data are discrete therefore discrete Fourier transforms (DFTs) are more appropriate. DFTs are utilised to transform a signal which is discrete in the time domain, $x[n]$, to one which is discrete in the frequency domain, $X[k]$.

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi kn/N} \qquad\qquad (4.3)$$

where $k = 0,1,2...N-1$

The resolution achieved in the frequency domain is determined by $\frac{1}{N\Delta t}$ where Δt is the interval between each sample in the time domain. It is assumed that the data $x[n]$ are periodic, with a fundamental period of N. In other words the interval over which the transform is performed, the analysis interval, is one cycle of a repeating series in time. The following example, containing both 10Hz and 20Hz is used to demonstrate the DFT:
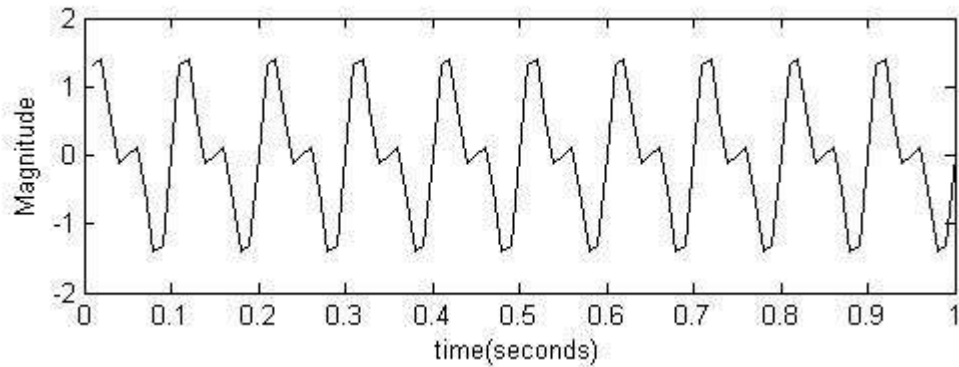
Figure 4.1 10Hz and 20Hz signal in the time domain.

The data were sampled at 100Hz while the DFT was performed using an analysis interval of one second (i.e. containing an integer number of cycles of the two principal frequencies). The resulting frequency plot is shown in figure 4.2:
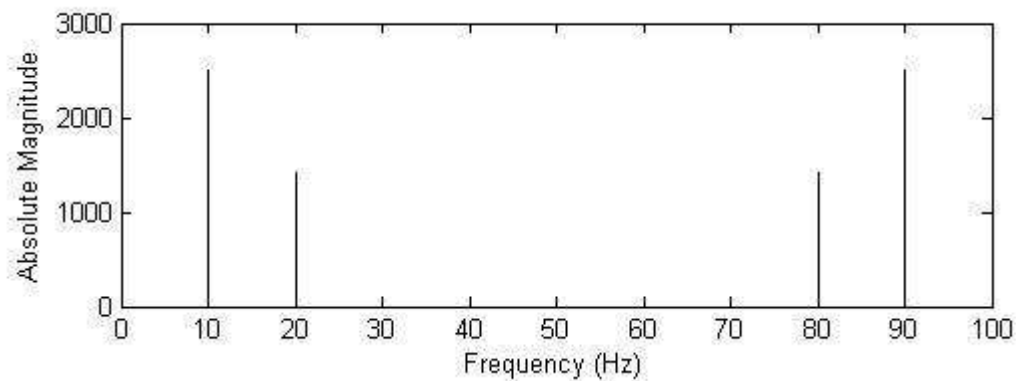
Figure 4.2 10Hz and 20Hz signal in the frequency domain. This plot is symmetrical around 50Hz, the Nyquist frequency.

Both the 10Hz and 20Hz components can clearly be seen in figure 4.2. The frequency-magnitude plot is symmetrical around the Nyquist frequency (half the sampling frequency). This is because the DFT is periodic, with one period extending from 0Hz to the sampling frequency; real signals are therefore always symmetric around the Nyquist frequency when transformed from the time to the frequency domain. Consequently it is common to plot the frequencies between 0 and the Nyquist frequency.

Problems can however be encountered when using DFTs, the most common of which are aliasing and leakage. Aliasing, when high frequencies are mistaken for lower frequencies, occurs when the sampling frequency is insufficient to represent the highest frequency components in the signal. This is avoided by ensuring that the rate at which the data are sampled is at least twice that of the highest frequency in the time domain data. Leakage arises when the analysis interval does not contain an integer number of

cycles of each frequency in the signal. This is reflected as a smearing of the signal's frequencies in the frequency domain. This can be seen in figure 4.3, where the signal in figure 4.1 was transformed to the frequency domain using an analysis interval of 0.93 seconds (i.e. the analysis did not include an integer number of periods of the 10Hz and 20Hz frequencies). A spread of frequencies can clearly be seen. Again the plot is symmetrical around 50Hz:
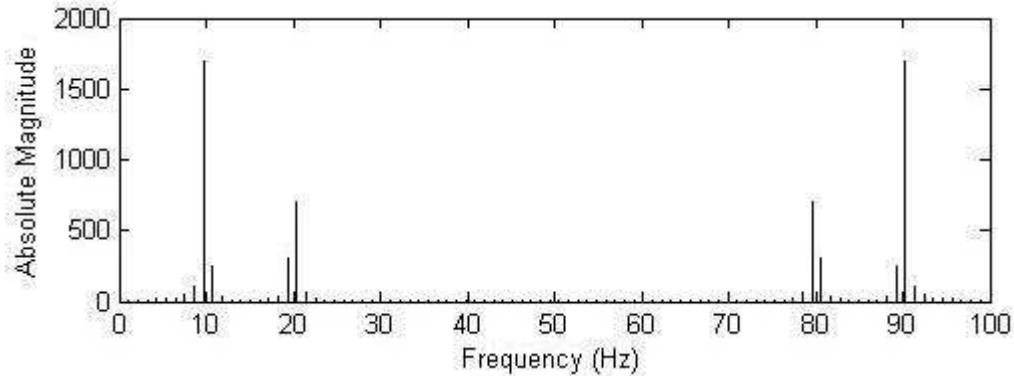
Figure 4.3 Demonstrating leakage in the frequency domain for a signal comprising 10Hz and 20Hz components. The peaks are not limited to 10Hz, 20Hz, 80Hz and 90Hz as in figure 4.2, but have spread; this is evidence of leakage.

To reduce the effect of leakage windowing techniques can be used. These involve multiplying the signal by a function which starts and finishes with a value of zero and reaches a maximum value of one.

To transform a signal from the frequency to the time domain the inverse transform is used:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{j 2\pi nk / N} \qquad (4.4)$$

The computational time required to perform a DFT is considerable, particularly when working with large data sets. When evaluating an N point DFT, for each value of k, N complex multiplications and N-1 complex additions are performed. The DFT therefore requires in the order of N^2 calculations.

4.1.3 Fast Fourier transform

Fast Fourier transforms (FFTs) have been developed to decrease this computational time. A number of different FFT algorithms exist, the most common of which is the Cooley-Tukey algorithm. By exploiting the periodic and symmetrical nature of the

DFT when the length of the signal is 2^n the FFT can decrease the number of computations to an order of $N\log_2 N$. If the signal is not 2^n long, sequences can be filled with zeroes or they can be resampled. FFTs were used for all Fourier transforms in this chapter due to the increased computational speed they offer. The use of the Fourier domain to assess recording quality is discussed initially.

4.2 Assessing recording quality

To monitor the integrity of the mfERG recording the frequency components of the raw, uncorrelated data signal were studied. When using the Fourier domain to distinguish signal from noise for steady-state responses such as the PERG or the VECP the stimulus frequency is known. All frequencies which are not a multiple of the stimulus frequency can therefore be identified as noise. Specific frequencies cannot however be studied when analysing the mfERG as the stimulus comprises multiple stimulating frequencies. Frequency profiles were therefore examined as opposed to individual frequencies.

The aim was to establish patterns inherent to acceptable and unacceptable recording conditions, thus enabling an objective grading system to be developed. Complete recordings were studied initially to emulate assessing the quality of an entire test. The Fourier profile of individual data segments was subsequently analysed to simulate grading a test's integrity during the examination.

4.2.1 Transforming uncorrelated mfERG data from the time to the Fourier domain

When collecting the uncorrelated data the recording period is split into 16 equal segments, each lasting approximately 30 seconds. There is an overlap of these segments; the initial 16 points of the m-sequence are repeated at the start while the final 16 steps are repeated at the end. This process is illustrated for the first three segments in figure 4.4:

| 1:16 | Segment 1 (1:2048) | 2032: 2048 | 2049: 2064 | Segment 2 (2049:4096) | 4080: 4096 | 4097: 4112 | Segment 3 (4097:6144) | 6128: 6144 |

Figure 4.4 Demonstrating the overlap of segments in the uncorrelated data. The green represents the steps of the m-sequence which are repeated; data acquired at these points are removed prior to transforming the data to the frequency domain. Data recorded during the orange periods are spliced together.

A program was therefore written to remove these sections of uncorrelated data (refer to appendix 1). A built in Matlab FFT function, based on the Cooley and Tukey algorithm, was then utilised to perform the transformation from the time to the Fourier domain. Frequencies greater than the Nyquist frequency were discarded and phase information was ignored (appendix 1). It was important to establish if leakage and aliasing would be a problem and hence if windowing techniques would require consideration. The frequency resolution achieved in the Fourier domain was also calculated.

4.2.2 Assessing for aliasing and leakage

Upon removal of the overlapping regions of the m-sequence the uncorrelated data comprised 524272 data points as the data were sampled 16 times for every step of the m-sequence. The analysis interval contained 32767 stimulus periods meaning that neither aliasing nor leakage were a problem and hence it was not necessary to employ windowing techniques. The majority of mfERG tests used a stimulus frequency of 75Hz. The resolution achieved in the Fourier domain was calculated by:

$$\Delta f = \frac{1}{N\Delta t} \text{ (refer to section 4.1.2)} = \frac{1}{524272 \times \frac{1}{1200}} = 0.0023 Hz$$

Similarly when a lower stimulation rate of 60.8Hz was selected, leakage and aliasing problems were not encountered and a frequency resolution of 0.0019Hz was achieved in the Fourier domain. Finally, when one segment of the raw data was transformed to the frequency domain aliasing and leakage problems were not encountered; the resolution in the frequency domain was 0.0366Hz.

4.2.3 Testing the transformation from the time to the frequency domain

Before transforming patient data to the frequency domain the program written in Matlab (detailed in appendix 1) was tested to ensure that it produced the results expected of it. The mfERG protocol (section 3.4.1) was used to test the program with two exceptions: the stimulus was a light source with a known frequency of 100Hz instead of a multi-element mfERG stimulus; and a photodiode was stimulated instead of a human eye.   The frequency plot shown in figure 4.5 was obtained:
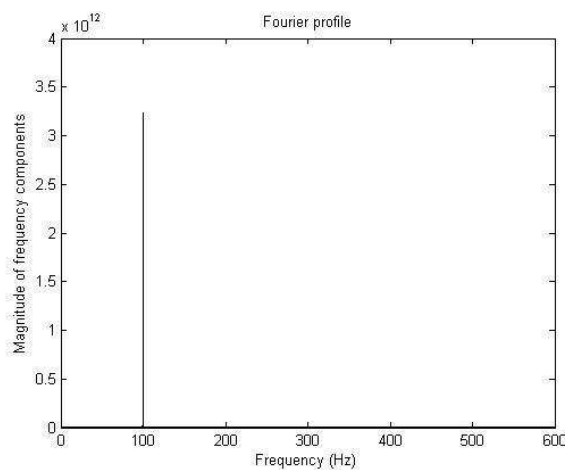
Figure 4.5 Fourier profile when a photodiode was stimulated with a 100Hz light source. A dominant peak was noted at 100Hz therefore the program functioned as required.

A strong 100Hz component can be seen, confirming that the transformation from the frequency to the time domain functioned as expected. It was therefore possible to apply this technique to the analysis of patient data with confidence. Initial experiments were conducted to study both the noise and the signal in isolation, to acquire knowledge of the profile specific to each. These were subsequently studied in combination.

4.2.4 Methods

The mfERG data were acquired in the time domain using the mfERG protocol (section 3.4.1). When isolating features such as the noise or the signal, alternations were made, each of which will be described.

4.2.4.1 Isolating the noise from the signal

A healthy control was set up using the mfERG protocol. To remove the evoked retinal responses they were asked to fixate on a green cross instead of the multi-element

stimulus described in the protocol. Four tests were conducted. These included good compliance (maintaining fixation on the cross throughout the test, minimising blinking and remaining still), loss of fixation, excessive blinking and muscle movement/jaw clenching, problems commonly encountered when testing patients. The exam is typically split into 16 equal segments therefore to investigate the possibility of grading the patient noise after each segment, $1/16^{th}$ of the uncorrelated data were transformed into the frequency domain for each of the four tests. Finally the Fourier profiles of the entire recordings were viewed to study the viability of grading signal quality at the end of a test. The frequency components specific to patient noise were identified and the features consistent with reduced patient cooperation were noted.

4.2.4.2 Isolating the signal from the noise

To acquire data uncontaminated by patient noise a photodiode was stimulated in lieu of a human eye using the 61 element stimulus. The testing protocol described in section 3.4.1 was utilised, with the exception that an amplifier gain of 100 was chosen. This was to minimise saturation of the signal; if a higher gain is chosen the signal saturates when using a photodiode. The uncorrelated data were transformed into the Fourier domain, allowing the frequency profile specific to the mfERG stimulus to be acquired.

4.2.4.3 mfERG recordings with varying compliance and normal retinal function

Having established the Fourier profiles particular to the mfERG stimulus and patient noise, a healthy control was tested using the mfERG protocol (section 3.4.1). As recording quality was of interest the person varied their cooperation, repeating the four recording conditions described in section 4.2.4.1. As before, individual segments of the data, in addition to entire recordings were viewed in the frequency domain to determine if this approach can be used to monitor the recording 'live' and upon completion of the test.

4.2.5 Results

4.2.5.1 Isolating the noise from the signal

Figure 4.6 (left) shows the Fourier profile obtained when the complete uncorrelated data set was transformed to the frequency domain. Frequency components from 0-600Hz, the Nyquist frequency, have been included. The range of the y-axis was determined by the magnitude of the maximum frequency component, found at 0Hz; this is shown in red. The 0Hz contribution is significant relative to all other frequency components making it difficult to extract any meaningful information when studying the profile. Figure 4.6 (right) shows a frequency plot with a maximum value at 8×10^7, enabling contributions at frequencies other than 0Hz to be seen:
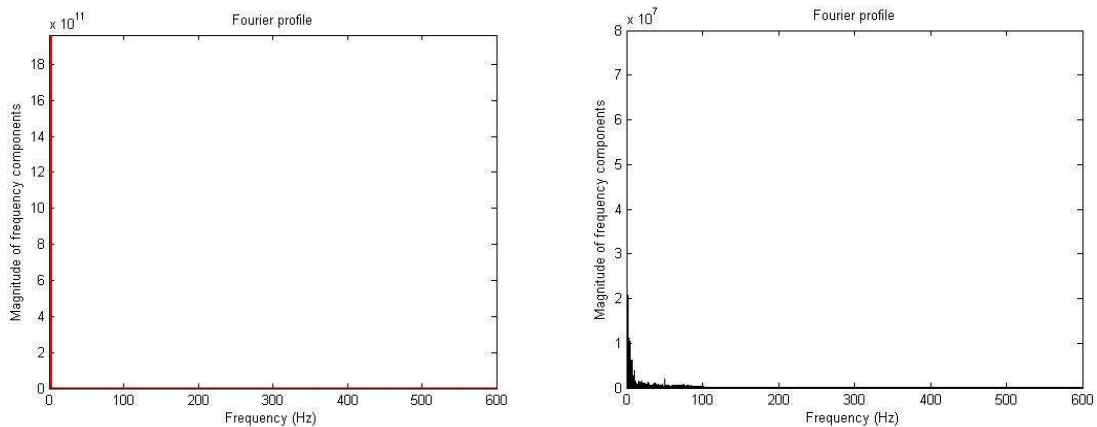
Figure 4.6 Fourier profiles from a full mfERG recording in a compliant subject when the noise was isolated from the signal; the range of each axis for the first profile (left) was defined by the maximum values, while the y-axis was limited for the second profile (right). The magnitude of the component at 0Hz is very large (shown in red) in comparison with that of all other frequencies; consequently when the range of the frequency profile (y-axis) is determined by the magnitude of the maximum frequency component no other frequencies can be seen (left graph). When the y-axis range is limited to 8×10^7 contributions at other frequencies can be seen. These predominantly lie within 0-100Hz.

It is evident that the range of 'interesting frequencies' lies within 0-100Hz. All subsequent Fourier profiles acquired by transforming the complete data set, with the exception of those measured using a photodiode, are shown using a maximum magnitude of 8×10^7 and frequency range of 0-100Hz. A similar problem was encountered when one segment of data was viewed in the Fourier domain: the 0Hz magnitude (shown in red) dominated all other frequency components. This can be seen in figure 4.7 (left). When using a decreased value of 2×10^7 for the y-axis, frequency components in the range 0-100Hz were seen. These magnitude and frequency ranges were therefore used to study one segment of data.
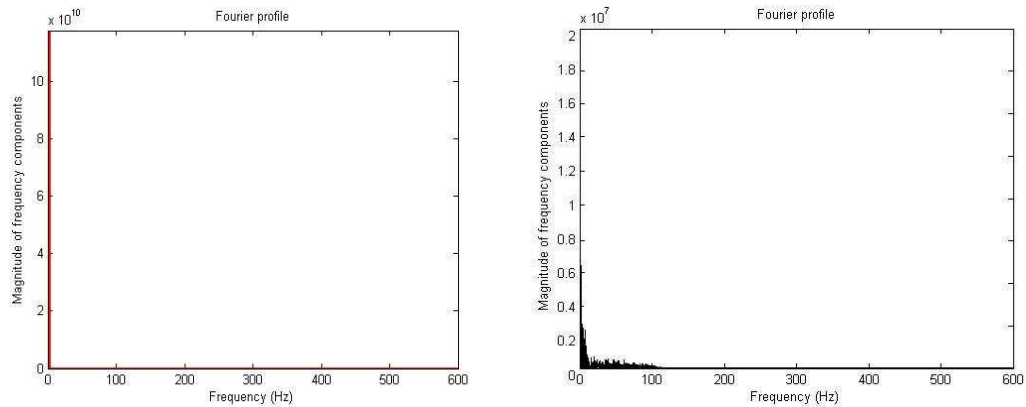
Figure 4.7 Fourier profiles from one segment of a mfERG recording in a compliant subject when the noise was isolated from the signal; the range of each axis for the first profile (left) was defined by the maximum x- and y-values, while the y-axis was limited for the second profile (right). The magnitude of the component at 0Hz is very large (shown in red) relative to that of all other frequencies; if the range of the frequency profile (y-axis) is determined by the maximum frequency component no other frequencies can be seen (left graph). When the y-axis range is limited to $2x10^7$ contributions at other frequencies are evident. These predominantly lie within 0-100Hz.

The patient noise Fourier profiles obtained from the subject when compliant, looking around the stimulus, blinking and increasing their muscle movement (e.g. jaw clenching or fidgeting) respectively can be seen in figure 4.8. The frequency and magnitude ranges detailed above are used. The first column represents the profiles acquired when the complete recording was transformed into the frequency domain while the second column shows those attained when one segment of the uncorrelated data was utilised:

Full recording   **One segment of recording**
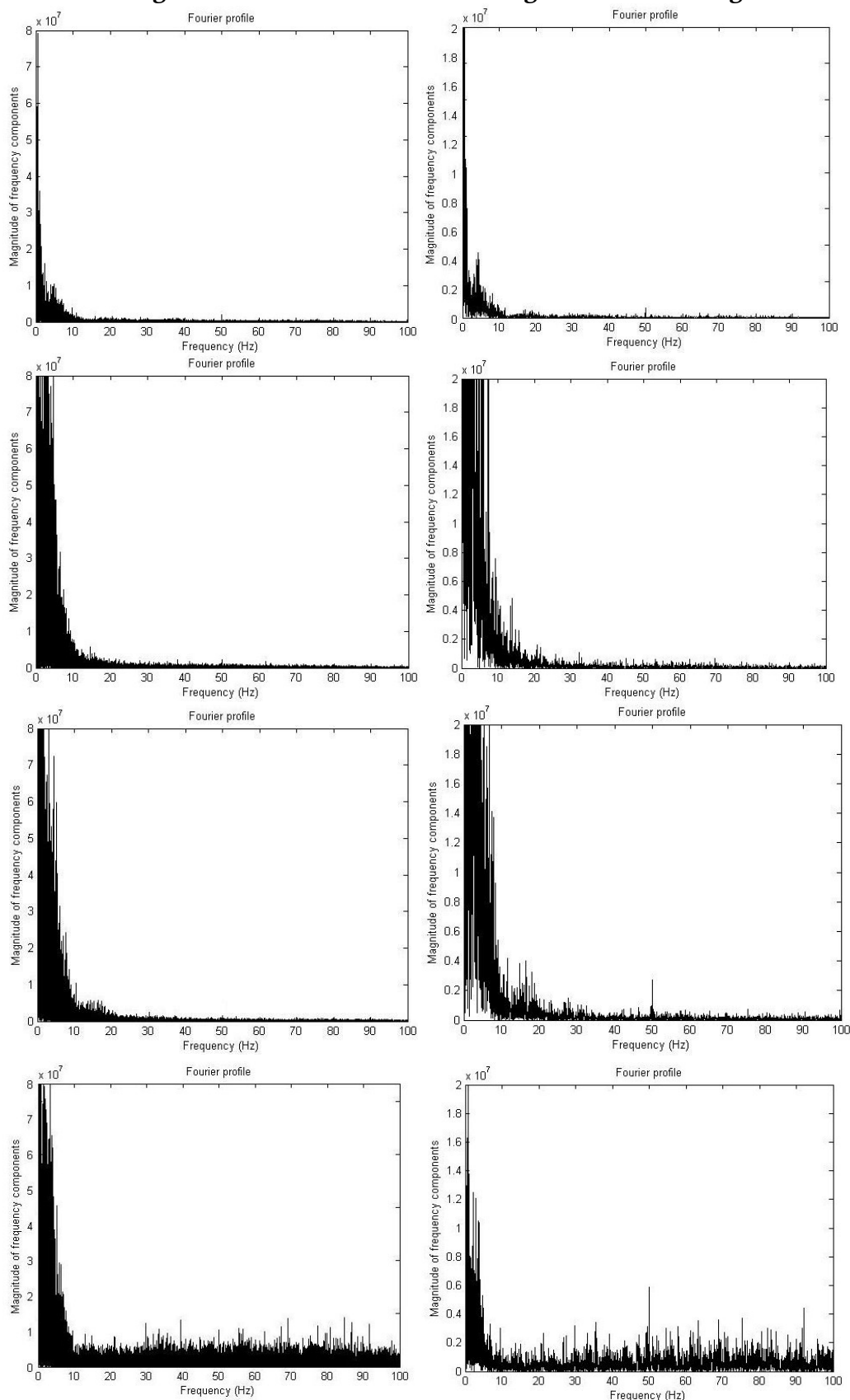
Figure 4.8 Fourier profiles acquired when the noise was isolated from the signal and patient compliance was varied. From top left to bottom right: 1) compliant, full recording; 2) compliant, 1 segment; 3) poor fixation, full recording; 4) poor fixation, 1 segment; 5) blinking, full recording; 6) blinking, 1 segment; 7) muscle movement, full recording; 8) muscle movement, 1 segment. The magnitude of low frequency components increased when the person's compliance decreased; muscle movement was also reflected as an increase in the frequency components across the full frequency range. This was seen both for one segment of data and for the full recording.

It can be seen that the lowest frequencies were the most prominent in each case. For those obtained when the person was less cooperative the prevalence of these low frequencies increased, reflecting greater amplifier saturation. Those acquired when the person increased their muscle movement also showed an increase in the magnitude of components across the frequency spectrum. Although the magnitude of the frequency contributions differed when one segment of data, as opposed to the whole recording was transformed to the Fourier domain, similar patterns were seen.

4.2.5.2 Isolating the signal from the noise

The Fourier profile of the uncorrelated data recorded when the photodiode was stimulated is shown in figure 4.9:
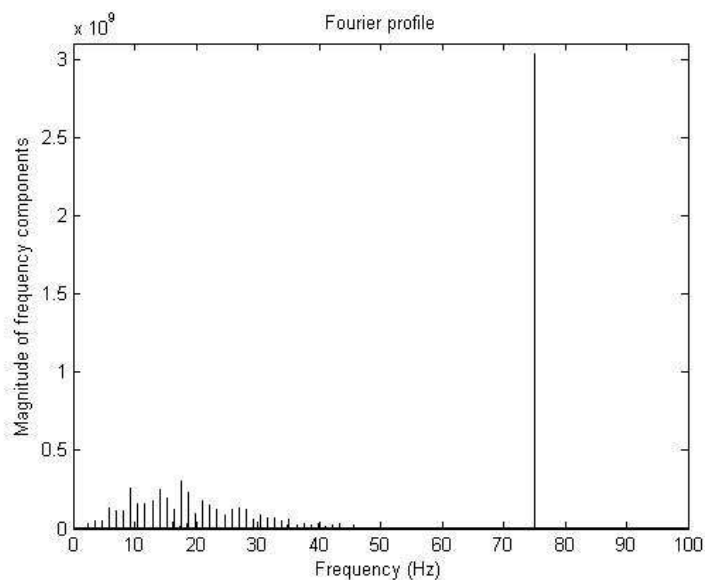
Figure 4.9 Fourier profile when the signal was isolated from noise. Note the response at 75Hz, the stimulus frequency, and the discrete frequency peaks. The separation between the peaks is the stimulus frequency/64.

A strong frequency component was observed at the stimulus frequency, 75Hz, in addition to a normative type frequency distribution of peaks, each separated by the stimulus frequency/64. These represent the main mfERG stimulating frequencies for this particular experimental set up.

4.2.5.3 mfERG recordings with varying compliance

When a healthy, compliant control subject was stimulated with the mfERG stimulus, the following Fourier profile was obtained when the entire uncorrelated signal was transformed to the frequency domain:
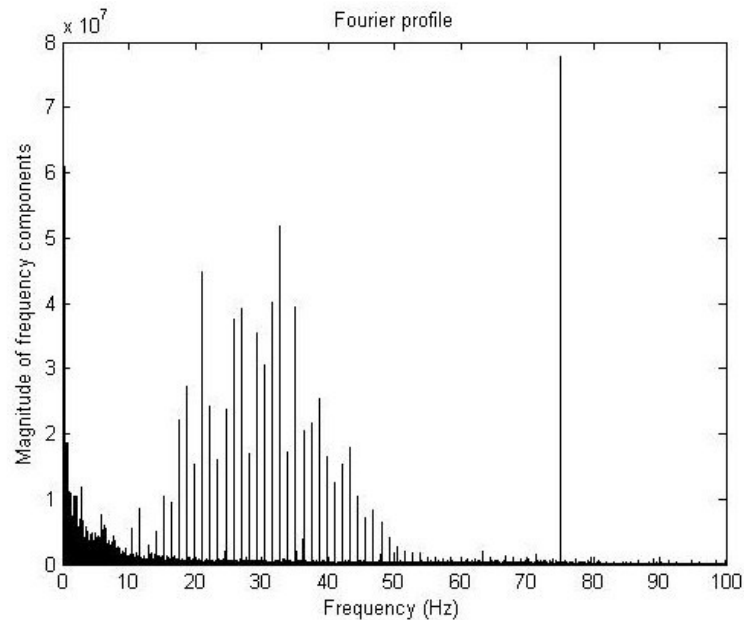
Figure 4.10 Fourier profile acquired from a full mfERG recording when a compliant subject was stimulated with the mfERG stimulus. The stimulus frequency and discrete frequency peaks are evident (as seen with photodiode), in addition to the noise profile seen in a compliant subject when unstimulated.

The features observed when both noise and signal were studied in isolation were present: the low frequency noise; the strong contribution at the stimulus frequency; and the discrete peaks, each separated by the stimulus frequency/64. It should be noted that the exact range of these peaks differed slightly to that of the photodiode. When the uncorrelated data from one segment of this recording were viewed in the frequency domain the following profile was seen:
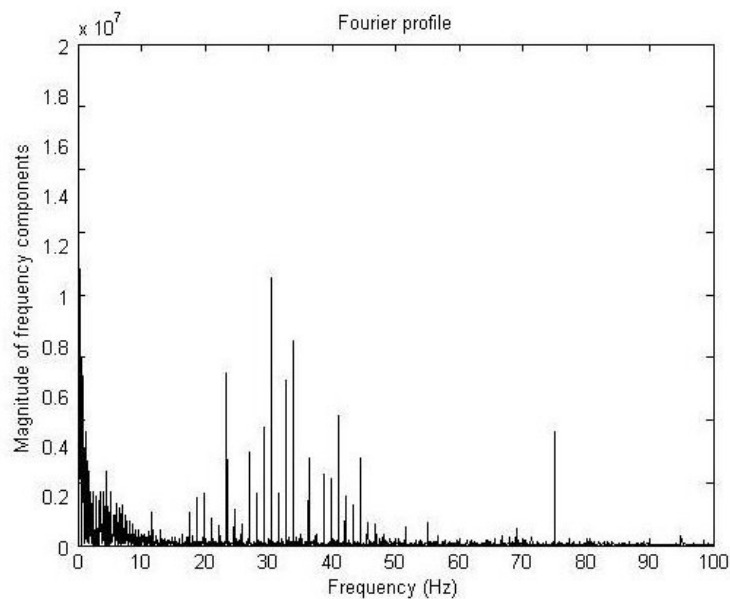
Figure 4.11 Fourier profile from one segment of a mfERG recording when a compliant subject was stimulated with the mfERG stimulus. The discrete frequency peaks and the stimulus frequency can be seen, in addition to the noise profile observed in a compliant subject when unstimulated.

The noise profile acquired from the compliant, unstimulated control subject (shown in figure 4.8) can be seen in this Fourier profile, in addition to the peak at 75Hz and the discrete frequency peaks. The Fourier profiles attained when the control looked around the stimulus, blinked and increased their muscle movement respectively can be seen, both when the entire raw data signal was transformed to the frequency domain and when one segment of the recording was examined in the Fourier domain:
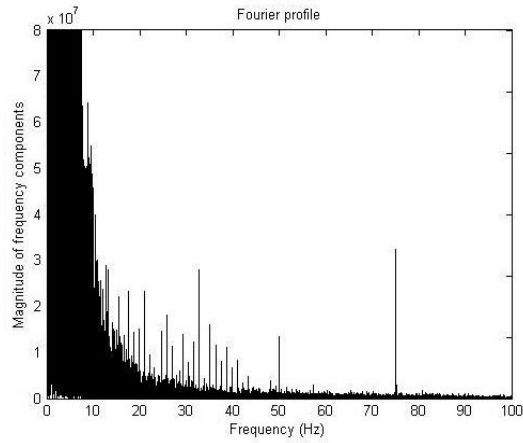
Full recording   **One segment of recording**
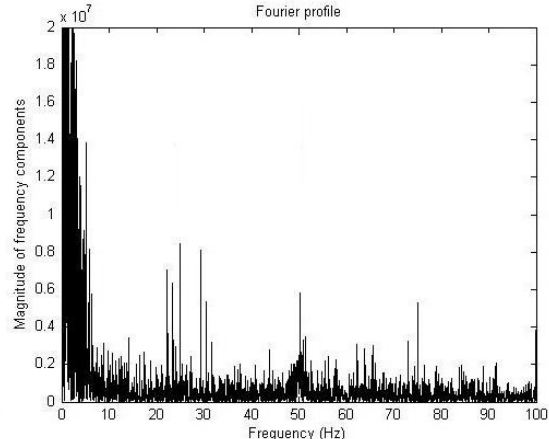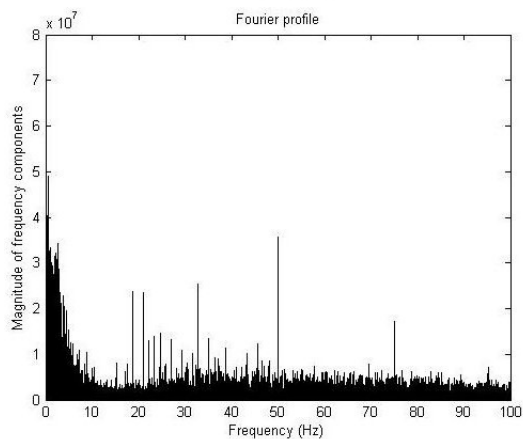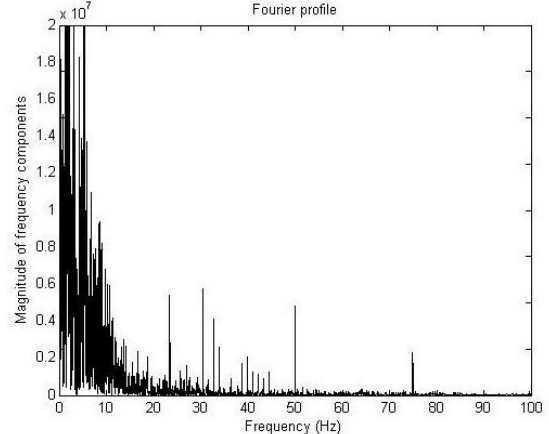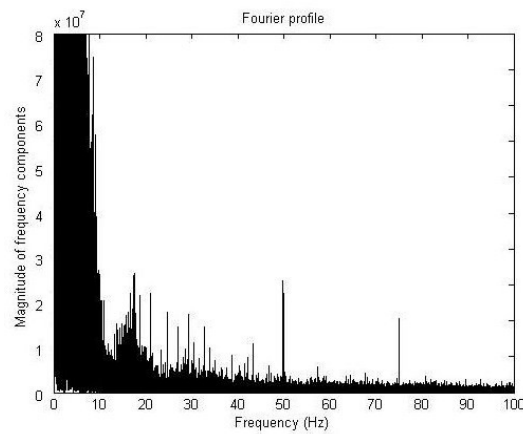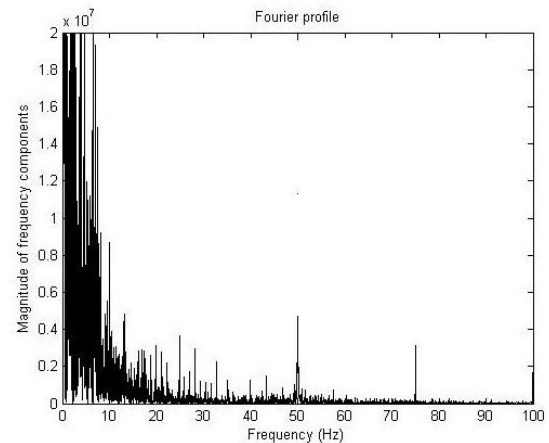
Figure 4.12 Fourier profiles acquired when the subject was stimulated with the mfERG stimulus but varied their compliance. From top left to bottom right: 1) poor fixation, full recording; 2) poor fixation, 1 segment; 3) blinking, full recording; 4) blinking, 1 segment; 5) muscle movement, full recording; 6) muscle movement, 1 segment. As for the unstimulated data shown in figure 4.8, the prevalence of low frequency components increased as compliance decreased, and muscle movement was reflected as an increase in the frequency magnitude in the 0-100Hz range. Frequency peaks, as seen in the photodiode test were evident; these decreased in magnitude as compliance was decreased.

The noise profiles were comparable with those seen when the subject was tested without the mfERG stimulus; there was an increase in the prevalence of low frequency noise when the subject was incompliant and an additional evenly distributed noise profile when they clenched their jaw/increased their muscle movement. Again the discrete peaks and the peak at the stimulus frequency were seen in each profile, although they were smaller in magnitude than those observed when the subject was

compliant. As was seen when the noise was studied in isolation the frequency components were smaller when $1/16^{th}$ of the data set was visualised in the Fourier domain however the features observed were comparable for each recording condition.

4.2.6 Discussion

Deciding if a mfERG recording is of a suitable standard to be reported can be difficult and subjective. Furthermore, distinguishing between normal and compromised retinal function can be problematic when the recording quality is suboptimal. It was therefore the initial aim of this chapter to establish if the frequency domain can be used to differentiate between a good and a bad recording, and if this could be used as a basis for grading recording quality in an objective manner both during and after a test.

It has been shown that by viewing the raw, uncorrelated data in the Fourier domain patient noise can be easily identified from signal and that a good recording is highly distinct from a poor one; the prevalence of low frequency noise increased significantly when the person blinked excessively or lost fixation. Furthermore, muscle noise could be distinguished from eye movement/blinking; an increase in frequency components across the frequency spectrum was noted in the case of muscle noise. This was the case both when studying the complete data set and when examining a smaller section of it, thus enabling it to be used both during and after a test to comment on the integrity of a recording.

As the data were displayed in a highly visual way this method lends itself to providing a quick, objective and effective method for measuring the integrity of a recording. By defining limits for a good recording and an unreportable test an automated grading of quality can be incorporated into a system, therefore providing an operator with a continuous assessment of the incoming signal and producing an evaluation of the overall recording quality upon completion of the mfERG test. This will be developed, tested and discussed in section 4.3, both for the full data set and for one segment of data.

Although the main objective of this section was to assess the recording quality, a number of interesting findings relating to the mfERG stimulus were also noted. When studying the signal in isolation the stimulation frequencies inherent to the mfERG system were revealed: a dominant peak at the stimulus frequency; and discrete peaks separated by the stimulus frequency/64. This frequency distribution was also present

when the control subject was tested, although the exact profile obtained from the eye and the photodiode differed. The photodiode has a linear output whereas the eye is non-linear; the non-linearity of the eye results in higher frequencies when there is an overlap of the evoked responses from multiple areas of the retina. As the frequency profile observed when testing the photodiode was also evident when the control subject was stimulated it can be said that these are the retinal stimulation frequencies and retinal response frequencies.

It was noted when testing the stimulated control that the magnitude of the stimulus peaks decreased when the subject was less cooperative. This can be credited to large sections of the recorded signal containing only patient noise instead of the evoked retinal signal, as well as a greater loss of signal due to increased amplifier saturation.

It was also noted that the magnitude of the frequency components was significantly larger when the entire recording was examined than when only one segment of data was studied. As a larger sample of data were transformed to the frequency domain the contribution at each frequency was greater.

When studying the uncorrelated data in the Fourier domain it was found that the ability of the retina to respond to the main mfERG stimulating frequencies could be viewed with ease. This is of particular interest as it may enable a quick and simple method of extracting temporal information on the function of the retina. The effect of compromised retinal function on the Fourier profile is therefore investigated in section 4.5 while a more in depth investigation into the factors affecting the stimulus profile will be discussed in section 4.4.

4.2.7 Conclusions

It has been shown that analysis of the uncorrelated data in the frequency domain is a viable method for assessing recording quality both during and after a mfERG examination; clear differences between good and bad recordings were observed. Furthermore, distinct patterns were seen corresponding to different types of patient artefacts. The highly visual nature of these findings suggest that this method will lend itself to the development of an objective quality grading for use both during and after a test.

4.3 Automating findings: grading recording quality

A method of quantifying the calibre of a recording is proposed in section 4.3. The aim was to grade the recording quality into one of four categories: 1) excellent; 2) moderate; 3) noisy; and 4) unreportable, for the full recording. This was first done for the low frequency noise and then for muscle noise; finally the magnitude of the 50Hz component was assessed in recordings contaminated with electrical noise. Each of these three noise profiles was studied in isolation and were subsequently combined to provide an overall quality grading. This was compared with that of three experts when presented with 50 previously unseen mfERG tests.

4.3.1 Methods: eye movement/blinking for a complete recording

Grading the recording quality based on low frequency noise present in a complete recording is discussed initially.

4.3.1.1 Defining the limits for excellent and unreportable recordings

60 clinical mfERG recordings, all obtained using the mfERG protocol described in section 3.4.1 were selected. These comprised 30 recordings with very little patient noise and 30 tests which had not been reported due to excessive blinking or eye movement during testing. The uncorrelated data for each recording were transformed from the time domain to the frequency domain. The following trace array and Fourier profile show an example of one of the 30 high quality recordings. A small section of the uncorrelated data can also be seen in figure 4.13.
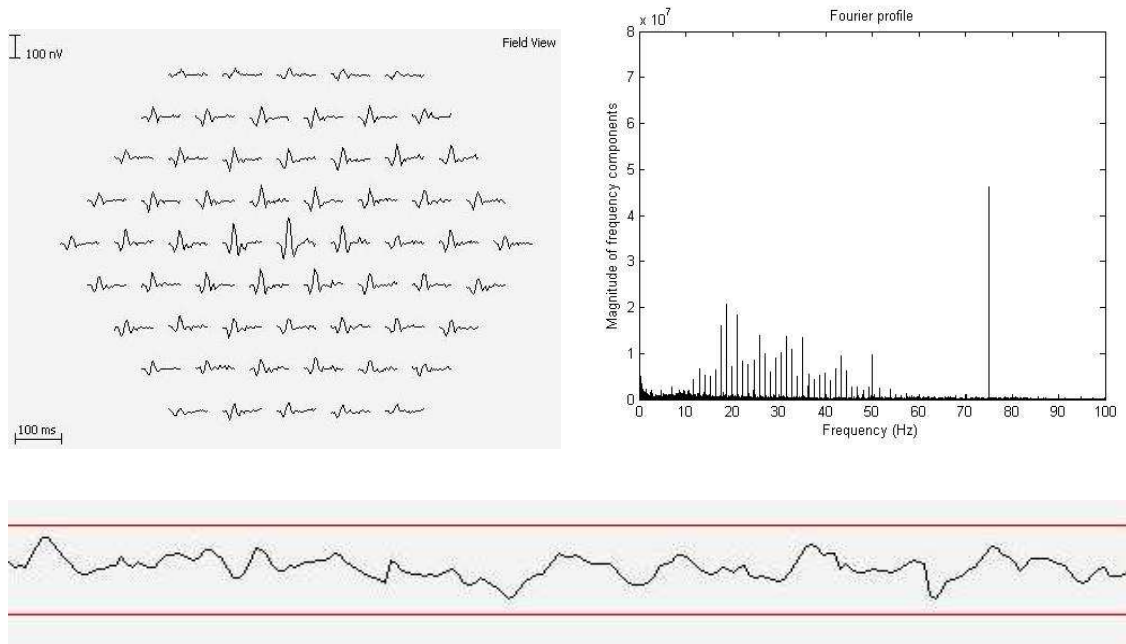
Figure 4.13 An example of a trace array (top left), a Fourier profile (top right) and a portion of the uncorrelated data (bottom) acquired from an excellent recording. The noise profile of the Fourier profile corresponds to that obtained from the compliant person in section 4.2.5.1. Red lines, defining the limits within which data should remain to obtain a high quality recording, are shown; these are arbitrary limits. In this instance the uncorrelated data are within these red lines.

An example of one of the 30 recordings contaminated by low frequency noise is shown below. Again the trace array, the Fourier profile and part of the uncorrelated data have been included:
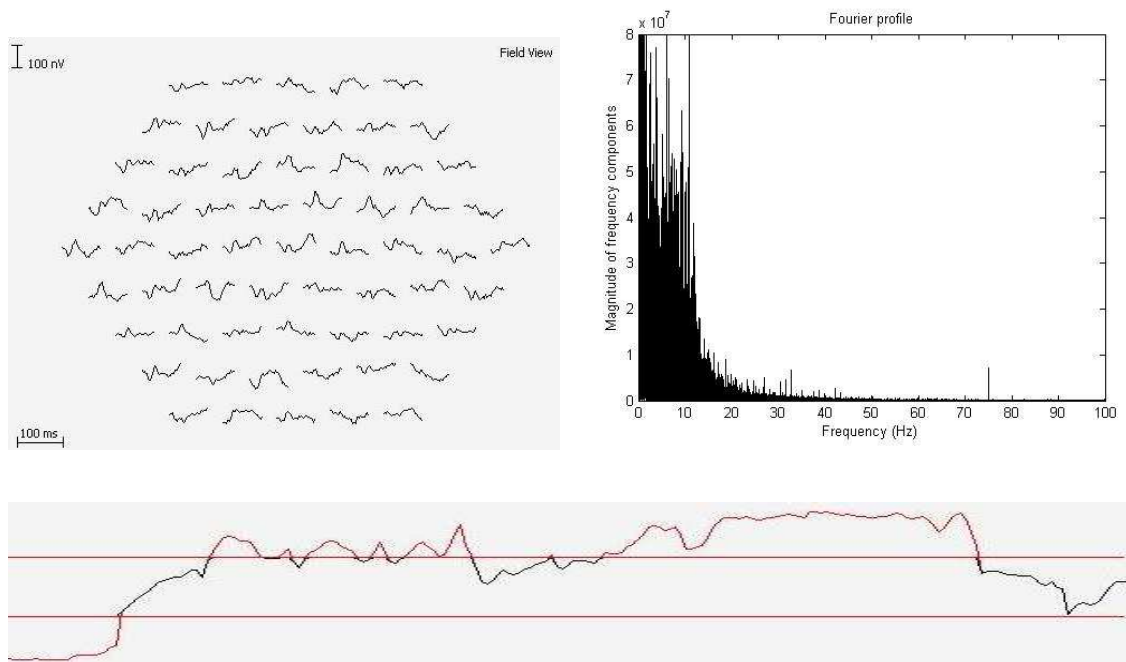
Figure 4.14 An example of a trace array (top left), a Fourier profile (top right) and a section of the uncorrelated data (bottom) obtained from an unreportable recording. The noise profile of the Fourier profile is comparable with that of the person tested in section 4.2.5.1 when looking around the stimulus or blinking excessively: the magnitude of the low frequency components is considerable. It can be seen that much of the section of uncorrelated data is outside the red lines; these are arbitrary limits within which the recorded data should remain to obtain a high quality recording, thus indicating a degradation of recording quality in this instance.

It was decided to examine the 0-10Hz part of the Fourier profile as clear differences between high and low quality recordings existed in this frequency range: the magnitude of the low frequency components increased in the case of blinking and eye movement. The data points corresponding to 0-10Hz on the Fourier plot were therefore selected; this included the first 4370 data points in the frequency domain. It was shown in section 4.2.5.1 that it was only possible to differentiate between a good and a bad recording when the Fourier profiles were plotted with a decreased y-axis. All magnitude values were therefore limited to 8×10^7.

The 30 good recordings were examined first. The magnitude of the frequency components at data points 1 to 4370, with values limited to 8×10^7 were exported from Matlab to an Excel file. From this the 5th, 50th and 95th percentiles were calculated at each data point (using the data from the 30 recordings), enabling 5th, 50th and 95th percentile frequency profile curves to be plotted. In theory it could then be said with a confidence of 95% that a test whose frequency-magnitude profile falls below the 95th percentile curve is of a high standard. This process was also done for the 30 poor recordings. In this case recordings with a frequency profile greater than the 5th percentile curve would be considered unreportable.

The magnitude of the frequency components fluctuates considerably. The following example, showing the 0-10Hz range of a recording contaminated with low frequency noise illustrates this:
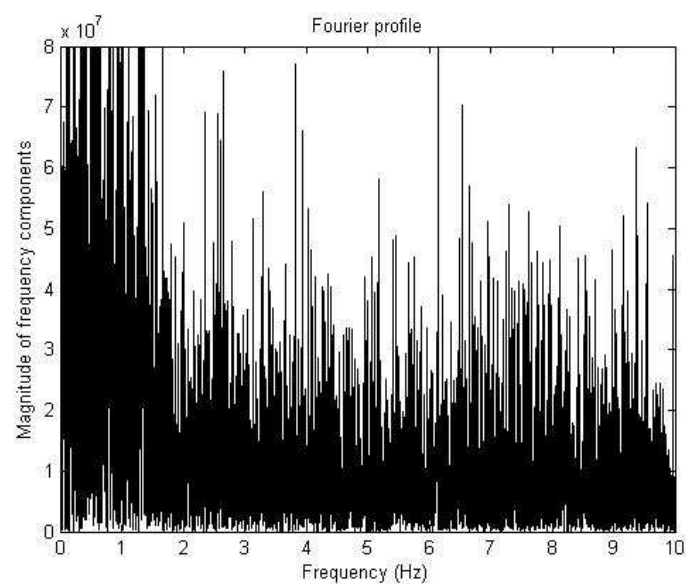
Figure 4.15 Demonstrating the fluctuation in the magnitude of the low frequency components for a recording contaminated with low frequency noise.

In addition to calculating the percentile curves using every data point it was decided to derive them using averaged magnitude values to minimise the impact of this fluctuation. An average value was therefore calculated for every 32 data points for each of the 60 recordings. Averaging over 32 data points was chosen as preliminary investigations showed that when using fewer data points fluctuation remained a problem, while averaging over a greater number of points reduced the information provided in the frequency domain. 5th, 50th and 95th percentile curves were once again acquired for the excellent and unreportable recording groups by calculating the 5th, 50th and 95th percentile at each of the averaged data points for the two groups. Ideally there would be a clear separation between the 95th percentile for the excellent recordings and the 5th percentile for the unreportable recordings. For those recordings with a Fourier profile between the 95th percentile of excellent recordings and the 5th percentile of unreportable recordings further subclassification would be useful as this region encompasses a wide variety of recording qualities.

4.3.1.2 Defining the limits for moderate and noisy recordings

A 3rd curve, referred to as the midline, was plotted with a view to dividing the region between excellent and unreportable. This was done by calculating the average value at each data point of the excellent 95th percentile curve and the unreportable 5th percentile limit. Moderate quality was therefore defined as the region between the excellent 95th percentile curve and the midline curve, while noisy but reportable recordings had a profile between the midline curve and the unreportable 5th percentile curve.

The aim was to develop a system capable of classifying the quality of a previously unseen mfERG recording by comparing its Fourier profile with the limits defined using these 60 recordings. It is however not always the case that the profile of a test will lie exactly between two limits for the full 0-10Hz range. It was therefore decided to calculate the area under each of the three limits (excellent 95th percentile, midline and unreportable 5th percentile). The area was then calculated under the 0-10Hz part of the Fourier profile for a new recording and compared with that of each of the limits. The recording integrity was subsequently classified as:

Excellent, when $A_{test} < A_{excellent\ limit}$

Moderate, when $A_{excellent\ limit} < A_{test} < A_{midline}$

Noisy, when $A_{midline} < A_{test} < A_{unreportable\ limit}$

Unreportable, when $A_{test} > A_{unreportable\ limit}$,

where $A_{excellent\ limit}$ was the area under the excellent recordings' 95th percentile curve, $A_{unreportable\ limit}$ was the area under the unreportable recordings' 5th percentile curve, $A_{midline}$ was the area under the midline curve and A_{test} was the area under the Fourier profile of the new test in the 0-10Hz region. The area was approximated by calculating the summation of the magnitude of each frequency component in the 0-10Hz range.

4.3.2 Methods: eye movement/blinking for one segment of a recording

A smaller section of the uncorrelated data was subsequently studied to enable recording quality to be graded during the testing session. Ideally the recording quality for each segment would also be categorised into one of the four classes defined for the complete recording as similar patterns were seen when studying one segment as all sixteen segments. It would however be very difficult to assess the system's performance as the experts would have to grade the integrity of each data segment simply by studying the uncorrelated data; this would be a highly subjective process. Consequently the number of gradings was decreased to two: 'acceptable'; and 'unreportable'.

To enable these limits to be defined one segment was selected from each of the 30 recordings contaminated by low frequency noise utilised when studying the full recording. These were transformed from the time to the frequency domain (refer to appendix 1), and exported from Matlab to Excel, where the 5th percentile curve was calculated and plotted for this group of noisy recordings. The region above this limit was defined as 'unreportable' while the area under the curve was said to be 'acceptable'. The methodology used was similar to that described in section 4.3.1; the principal differences were the number of data points in the 0-10Hz range, the value to which all frequency magnitudes were limited and the number of data points over which averaging was performed to derive the percentile plots: these were 273 points as opposed to 4370; 2×10^7 instead of 8×10^7; and 2 as opposed to 32, respectively. In this case every 2 points were averaged as there were 16 times fewer data points than when using the complete recording; this was therefore equivalent to 32 when using the full recording. Again the area under the limit was calculated.

To enable the recording quality of subsequent tests to be graded the relevant segment of the new test was transformed from the time to the frequency domain using the program detailed in appendix 1. The initial 273 data points were selected and a plot was derived by calculating the average of every 2 data points. The area under this profile was calculated and compared with that of the 5th percentile of the unreportable recordings. The data segment was categorised as:

Acceptable, when $A_{test} < A_{unreportable\ limit}$

Unreportable when $A_{test} > A_{unreportable\ limit}$

where A_{test} was the area under the profile of the mfERG recording being graded, and $A_{unreportable\ limit}$ was the area under the 5th percentile of the unreportable recordings.

4.3.3 Methods: muscle movement for a complete recording

It was shown in section 4.2.5 that muscle noise and jaw clenching are reflected as an increase in the magnitude of the frequency components across the spectrum. The aim was to grade the recording quality of a complete recording, in terms of muscle noise, into one of four categories: 1) excellent; 2) moderate; 3) noisy but reportable; and 4) unreportable.

4.3.3.1 Defining the limits for excellent and unreportable recordings

30 recordings were selected which were recorded under optimal conditions. An additional 30 recordings which an expert was unable to analyse due to patient movement were chosen. The following images demonstrate the trace array, the Fourier profile and a sample of the uncorrelated data respectively for one of the 30 poor quality recordings:
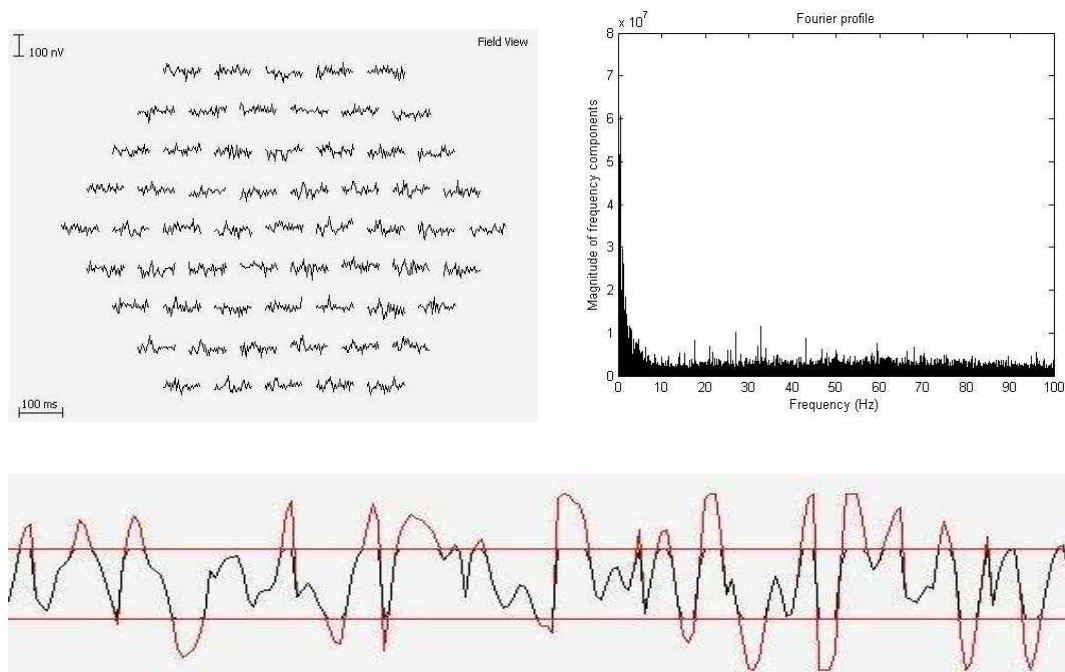
Figure 4.16 An example of a trace array (top left), a Fourier profile (top right) and a portion of the uncorrelated data (bottom) acquired from a recording which was unreportable due to excessive muscle movement. The noise profile of the Fourier profile is comparable with that of the person tested in section 4.2.5.1 when they increased their muscle movement: the magnitude of the frequency components increased across the frequency spectrum. Red lines, indicating the limits within which the data should remain to attain a high quality recording are shown. It can be seen that much of this section of uncorrelated data is outside these red lines, indicating poor recording quality.

The 25-100Hz frequency range was studied for these 60 recordings. This was to ensure the separation of muscle noise and blinking noise as a small number of recordings contaminated by 'blinking' noise contain frequency components of a significant magnitude at frequencies greater than 10Hz. As 50Hz electrical noise was being examined separately it was decided to ignore 50Hz and its harmonic 100Hz when assessing muscle noise.

An additional consideration was that the frequency range being examined contains both noise and signal in the case of retinal function. The aim of this section was to assess the integrity of the recording, irrespective of retinal function, therefore it was necessary to ignore the data at each frequency known to be related to the mfERG stimulus. This included the principal stimulus frequency, 75Hz, and the discrete frequency peaks located every 75/64Hz (refer to figure 4.10), always at the same frequencies. The program detailed in appendix 1 was modified to discard all data with a frequency of less than 25Hz, after being transformed from the time to the frequency domain. Frequency contributions at 50Hz, 100Hz, 75Hz and at the discrete frequency peaks were also removed; this was done with a knowledge of the data points at which these frequency contributions were present. The data were subsequently exported from Matlab to Excel for further analysis. Figure 4.17 demonstrates the removal of the

frequency peaks associated with the mfERG stimulus and 50Hz noise, first for a recording with very little muscle noise and then for a mfERG test containing considerable noise:
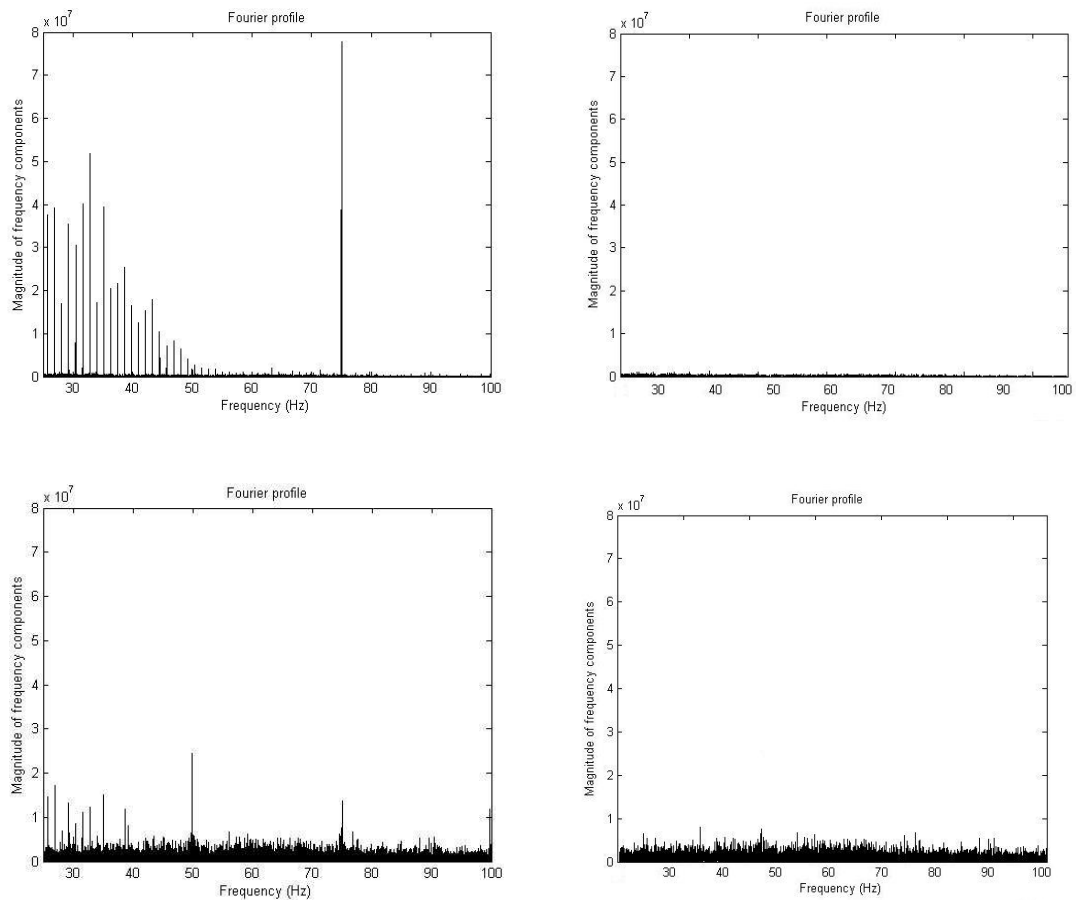
Figure 4.17 Demonstrating the removal of the stimulus associated frequency components from the 25-100Hz range of the Fourier profile. The original Fourier profile obtained from an excellent recording is shown (top left), in addition to its profile after the stimulus peaks have been ignored (top right). The bottom left image is the Fourier profile obtained from a poor recording; this profile after the removal of the stimulus associated peaks can be seen at the bottom right.

As before the 5th, 50th, and 95th percentile of the uncorrelated data were calculated in Excel for the excellent and unreportable recordings. Again magnitude values were taken from every data point, as well as an average of every 32 frequency increments. Limits for excellent and unreportable recordings were therefore defined. As with eye movement/blinking a range of recording quality exists between these classifications.

4.3.3.2 Defining the limits for moderate and noisy recordings

A third curve was derived by averaging the value of the excellent 95th percentile and the unreportable 5th percentile at each data point. This divided the region between excellent and unreportable, thus allowing recordings to be further classified as moderate and noisy but reportable. The area under each limit line was subsequently

calculated. As when grading the recording quality in terms of low frequency noise this value was utilised to classify the recording integrity of all future mfERG tests in terms of muscle movement. This was achieved by transforming the entire uncorrelated data set for the new test from the time to the frequency domain, isolating the 25-100Hz section of the Fourier profile, removing all data associated with the mfERG stimulus and calculating the area under the plot. As for the low frequency noise the recording quality was categorised as:

Excellent, when $A_{test} < A_{excellent\ limit}$

Moderate, when $A_{excellent\ limit} < A_{test} < A_{midline}$

Noisy, when $A_{midline} < A_{test} < A_{unreportable\ limit}$

Unreportable, when $A_{test} > A_{unreportable\ limit,}$

where $A_{excellent\ limit}$ was the area under the excellent recordings' 95th percentile curve, $A_{unreportable\ limit}$ was the area under the unreportable recordings' 5th percentile curve, $A_{midline}$ was the area under the midline curve and A_{test} was the area under the Fourier profile of the new test.

4.3.4 Methods: muscle movement for one segment of a recording

One segment of the uncorrelated data was chosen from each of the 30 mfERG recordings contaminated with patient noise and transformed to the Fourier domain. As before the frequency components in the 25-100Hz range were examined while peaks associated with the mfERG stimulus, as well as 50Hz and its harmonic at 100Hz, were removed in Matlab. The 5th percentile was plotted in Excel for these 30 recordings, enabling the quality to be defined as one of two classifications: those above the limit were 'unreportable' while those below were 'acceptable'. As before magnitude values were taken from every data point and then from averaged data points to plot the 5th percentile curve; an average of every 2nd frequency increment was utilised as this was equivalent to every 32 data points for the entire data set. The area under this limit was calculated to enable the recording integrity of future tests to be categorised by comparing their area with that of the 5th percentile for unreportable segments of a recording: those recordings with an area less than that of the 5th percentile were classified as 'acceptable' while all others were said to be 'unreportable'.

4.3.5 Methods: 50Hz noise for a complete recording and one segment of a recording

Finally the magnitude of the 50Hz component was assessed. An example of a recording containing a large contribution at 50Hz can be seen below. The trace array, the Fourier profile and a sample of the uncorrelated data are displayed:
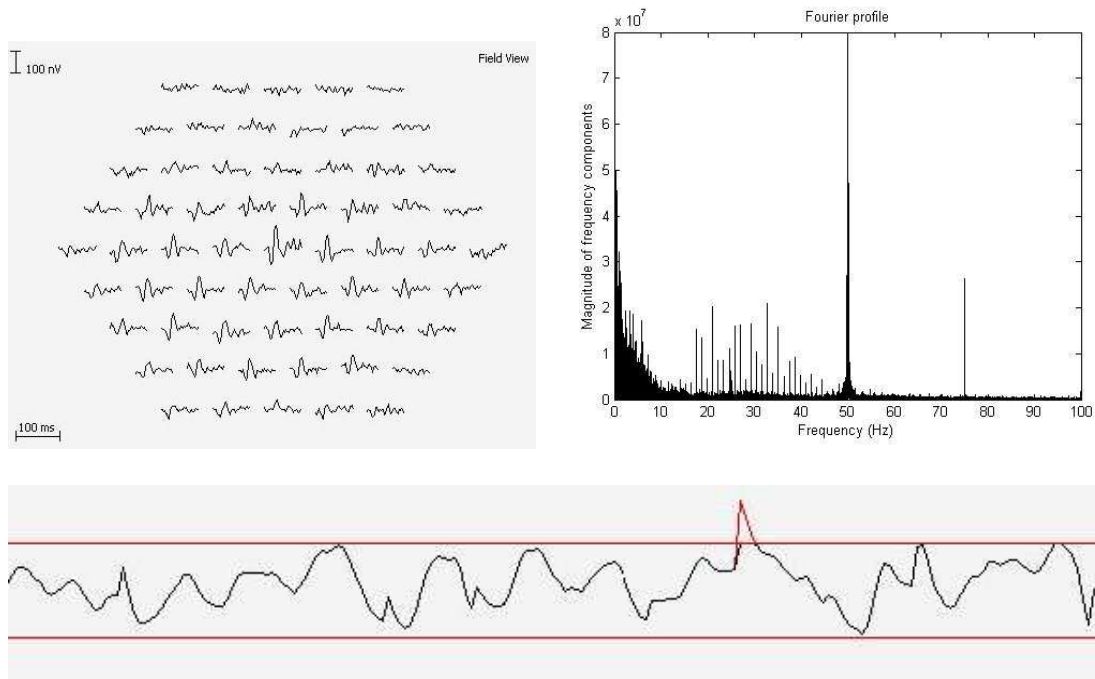
Figure 4.18 An example of a trace array (top left), a Fourier profile (top right) and a portion of the uncorrelated data (bottom) acquired from recording contaminated with 50Hz noise. A strong 50Hz component is evident in the Fourier profile. Red lines are shown along with the uncorrelated data. These are arbitrary limits within which the recorded data should ideally remain to achieve a high quality recording. In this instance the majority of data are within these limits despite the 50Hz noise.

Ten recordings contaminated with 50Hz noise were selected. The uncorrelated data were transformed from the time domain to the frequency domain and the magnitude of the frequency component at 50Hz was established. The magnitude of the 50Hz component was also examined for these ten recordings when one segment of the data was studied in the frequency domain.

4.3.6 Combining and testing the system

The findings for the three types of noise were subsequently utilised to grade the recording quality for 50 previously unseen recordings. The grading provided by the system was compared with that of the experts in each instance. The ability of the system to grade the overall recording quality was assessed first. 50 previously unseen mfERG recordings, all of which were obtained using the mfERG protocol (section

3.4.1), were selected retrospectively. These represented a wide variety of recording quality and retinal function. The uncorrelated data for each recording (the complete data set) were transformed to the frequency domain and subsequently categorised in terms of low frequency noise and muscle noise using the methods and limits defined in sections 4.3.1 and 4.3.3. The magnitude of the 50Hz component was also established allowing comment on electrical noise if appropriate. To provide an overall computational classification the worst grading was chosen.

The tests were examined independently by three experts, each of whom graded the overall recording quality as excellent, moderate, noisy or unreportable. This was done by viewing the final cross correlated trace array as well as the raw, uncorrelated data acquired for all 16 segments of the test (an option made possible using custom built software). When there was a discrepancy in the grading the most common answer was chosen. This was then compared with that of the computer for each of the 50 recordings.

The efficacy of the system to grade the quality of individual data segments was subsequently examined; 50 single data segments were selected, transformed into the frequency domain and presented to the system; these were graded as either 'unreportable' or 'acceptable' both in terms of the low frequency noise and muscle artefacts using the methods and limits defined in sections 4.3.2 and 4.3.4. The magnitude of the 50Hz component was also assessed. The worst of the three gradings was utilised to define the system's classification. The three experts also assessed the quality of these 50 individual segments, categorising them as 'unreportable' or 'acceptable'. This was achieved by viewing the uncorrelated data. The majority expert opinion was then compared with the classification stated by the system for each of the 50 data segments.

It was of interest to assess the interobserver variation; to obtain this the Kappa value was calculated. This method is often used to quantify the agreement between observers (123-125) and takes into account the fact that observers can concur with one another simply by chance (126). A Kappa value of 1 corresponds to complete agreement between people; 0 states that any agreement is a result of chance while -1 indicates a systematic level of disagreement between observers. Excel was utilised to perform these calculations.

4.3.7 Results: eye movement/blinking for a complete recording

It was shown in section 4.2.4.1 that both eye movement and a loss of fixation are reflected as an increase in the magnitude of the low frequency components. The aim was therefore to categorise the integrity of a recording as excellent, moderate, noisy or unreportable, based upon the magnitude of the low frequency components.

4.3.7.1 Defining the limits for excellent and unreportable recordings

The limits for excellent and unreportable are presented initially. When each data point was used to establish the 5th, 50th and 95th percentile noise functions for the high quality group the following plot was obtained:
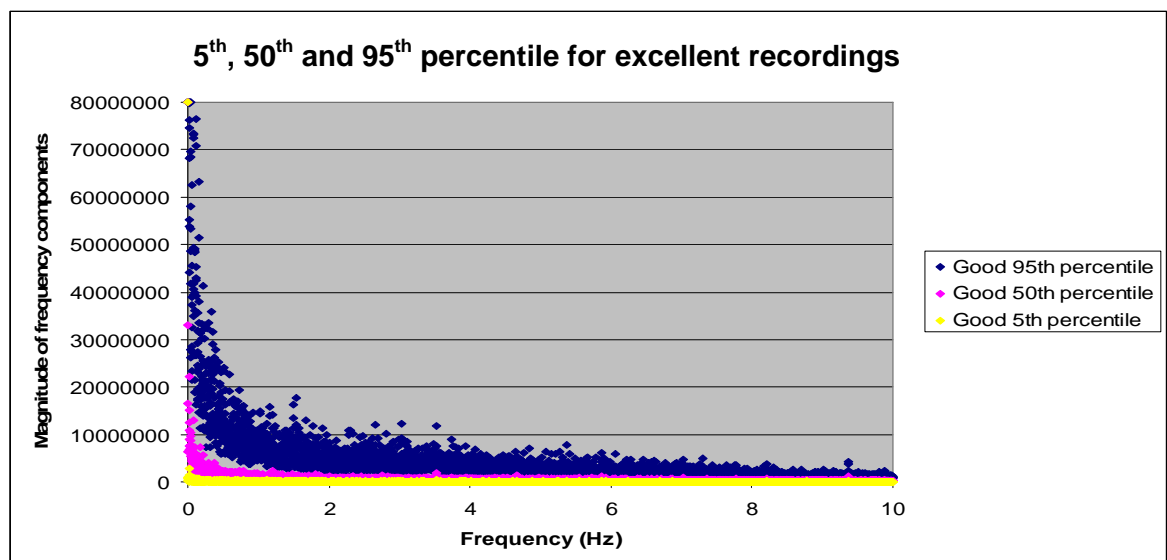
Figure 4.19 The 5th (yellow), 50th (pink) and 95th (blue) percentile curves for 30 excellent recordings when every data point was utilised to derive the plots.

It could therefore be said that mfERG tests with a noise profile within the 95th percentile (derived from the 30 high quality recordings), fulfil the requirement for excellent. The equivalent plot for the 30 recordings contaminated by low frequency noise is shown in figure 4.20:

5th, 50th and 95th percentile for unreportable recordings
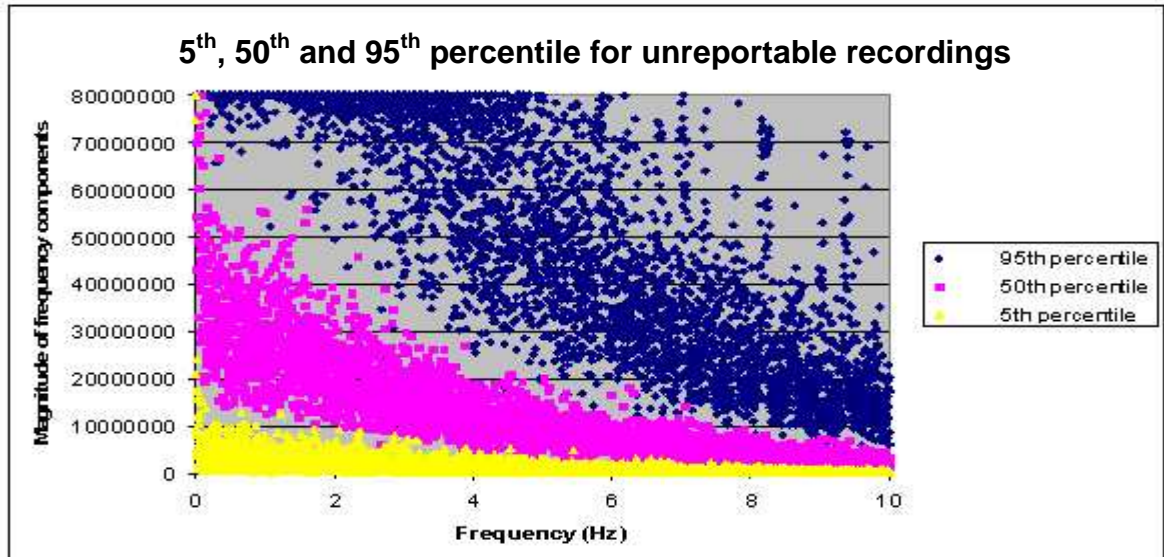
Figure 4.20 The 5th (yellow), 50th (pink) and 95th (blue) percentile curves for 30 recordings contaminated with low frequency noise when every data point was utilised to plot the curves.

From this it would be said that recordings with a frequency profile of a smaller magnitude than the 5th percentile curve would be reportable. The magnitudes of the 5th percentile curve were however surprisingly small. When the excellent 95th percentile and unreportable 5th percentile curves are shown on the same plot the following can be seen:

95th percentile for excellent and 5th percentile for unreportable recordings
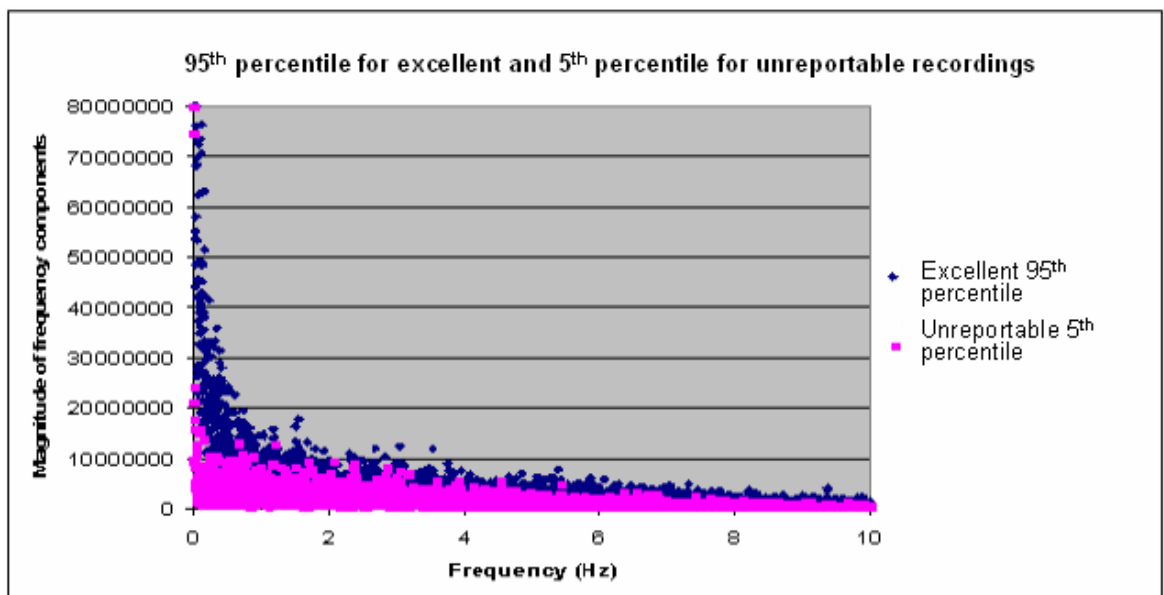
Figure 4.21 The 95th percentile for the 30 excellent recordings (blue), and the 5th percentile for the 30 unreportable recordings (pink) when every data point was utilised to derive the plots. An overlap of these limits is evident, implying that the magnitude of the frequency components was greater for the excellent recordings than the unreportable recordings.

It can be seen that there is an overlap of the limits for excellent and poor recordings when calculating the percentiles using every data point. Furthermore, this graph implies that magnitudes are greater for excellent recordings than for unreportable

recordings, which was not found to be the case when noise profiles were studied. It should be noted that the 5th percentile of the unreportable group was not representative of any of the 30 recordings. This approach could not therefore be used to grade recording quality. When the magnitude was averaged for every 32 data points the following graph was obtained for the 95th percentile of the excellent group and the 5th percentile of the unreportable group:
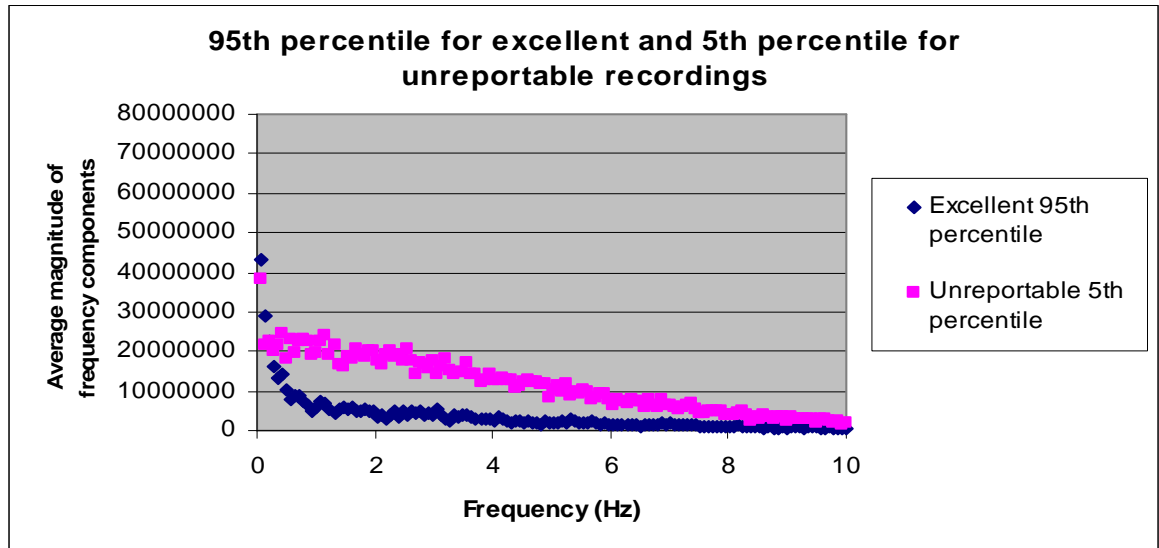
Figure 4.22 The 95th percentile for the 30 excellent recordings (blue), and the 5th percentile for the 30 unreportable recordings (pink) when every 32 data points were averaged prior to calculating the percentiles. The limits defining excellent and unreportable are distinct from one another in this instance.

By averaging the magnitude of every 32 components it can be seen that there is a separation between excellent and unreportable recordings. All subsequent grading for the complete recording therefore averaged every 32 data points (i.e. when grading the quality of previously unseen tests).

4.3.7.2 Defining the limits for moderate and noisy recordings

To sub-classify those mfERG tests with a recording quality in the region between excellent and unreportable, an additional limit was derived by averaging the excellent 95th percentile and the unreportable 5th percentile at each point. This enabled the four classifications to be defined:
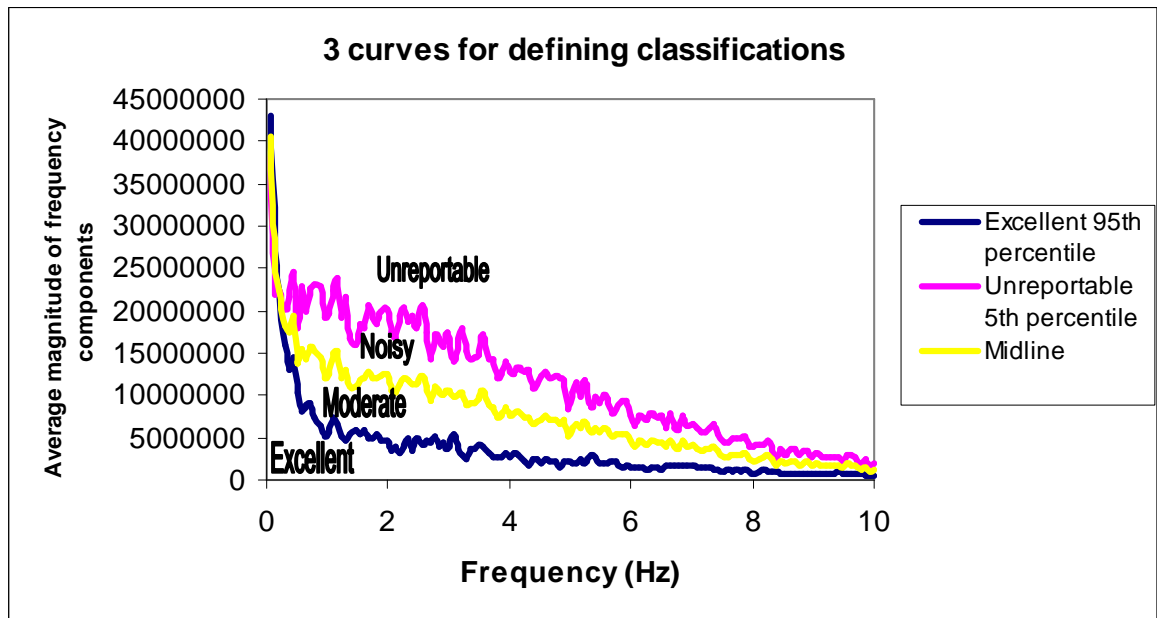
Figure 4.23 The four classifications (excellent, moderate, noisy and unreportable) for recording quality based upon the magnitude of the low frequency components. The excellent 95th percentile (blue), the midline (yellow) and the unreportable 5th percentile (pink) are shown. The quality classifications are defined as: those recordings with a frequency profile below the blue limit are 'excellent'; those between the blue and yellow curves are 'moderate'; those between yellow and pink are 'noisy'; those with a frequency profile greater than the pink limit are 'unreportable'.

The area under each curve was subsequently calculated; this was to enable the quality of additional mfERG tests to be graded by comparing their area with those of the three limits. The area under each of the three curves was:

$A_{excellent\ limit} = 4.93 \times 10^8$

$A_{midline} = 10.30 \times 10^8$

$A_{unreportable\ limit} = 15.66 \times 10^8$

These values were utilised to classify the integrity of future recordings in terms of eye movement and blinking using the method described in section 4.3.1.

4.3.8 Results: eye movement/blinking for one segment of a recording

The following graph displays the 5th percentile for the group of noisy recordings when only one data segment was analysed; this curve was acquired using an average of every 2nd data point (the equivalent to every 32 data points when studying the complete recording) to minimise the effect of the fluctuation of the frequency magnitudes (refer to figure 4.15). The two classifications for signal quality can be seen:
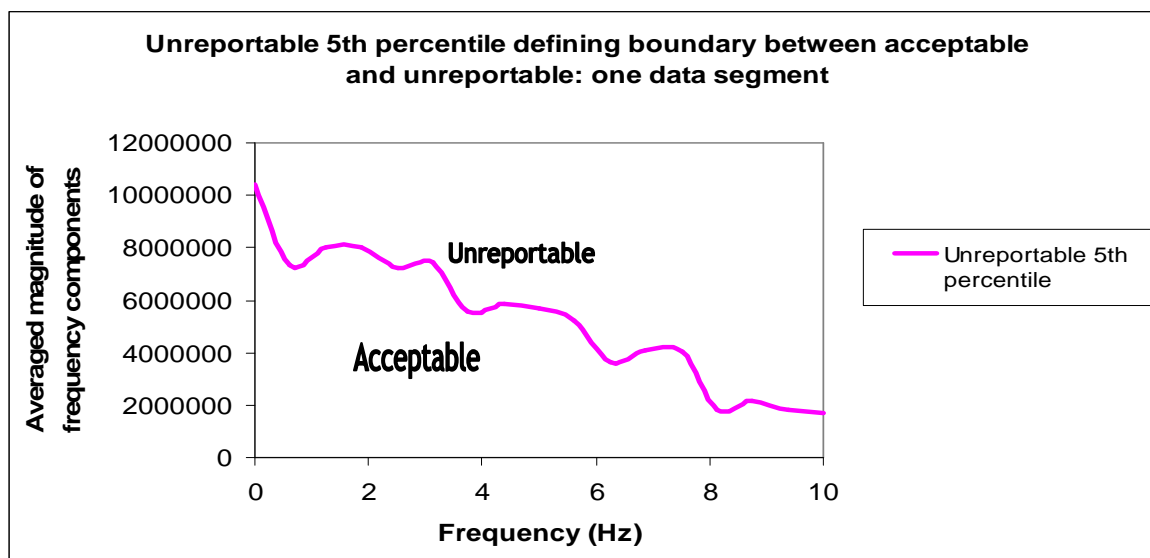
Figure 4.24 The two classifications (acceptable and unreportable) for one segment of a recording in terms of patient noise caused by eye movement and a loss of fixation. Those recordings with a Fourier profile above the 5th percentile of the group of unacceptable recordings are categorised as 'unreportable' while those below the limit are said to be 'acceptable'.

The area under the curve was 9.02×10^7 for the unreportable 5th percentile; this value was utilised to grade the integrity of future recordings.

4.3.9 Results: muscle movement for a complete recording

It was shown in section 4.2.5 that muscle noise and jaw clenching are reflected as an increase in the magnitude of the frequency components across the spectrum. As for the low frequency noise the aim was to grade the recording quality in terms of muscle noise into one of four categories: 1) excellent; 2) moderate; 3) noisy but reportable, and 4) unreportable.

4.3.9.1 Defining limits for excellent and unreportable recordings

The limits for 'excellent' and 'unreportable' are discussed initially. As was found when 0-10Hz frequency components were studied there was an overlap of the excellent 95th percentile and unreportable 5th percentile curves when every data point was utilised to calculate these limits. This can be seen in figure 4.25:
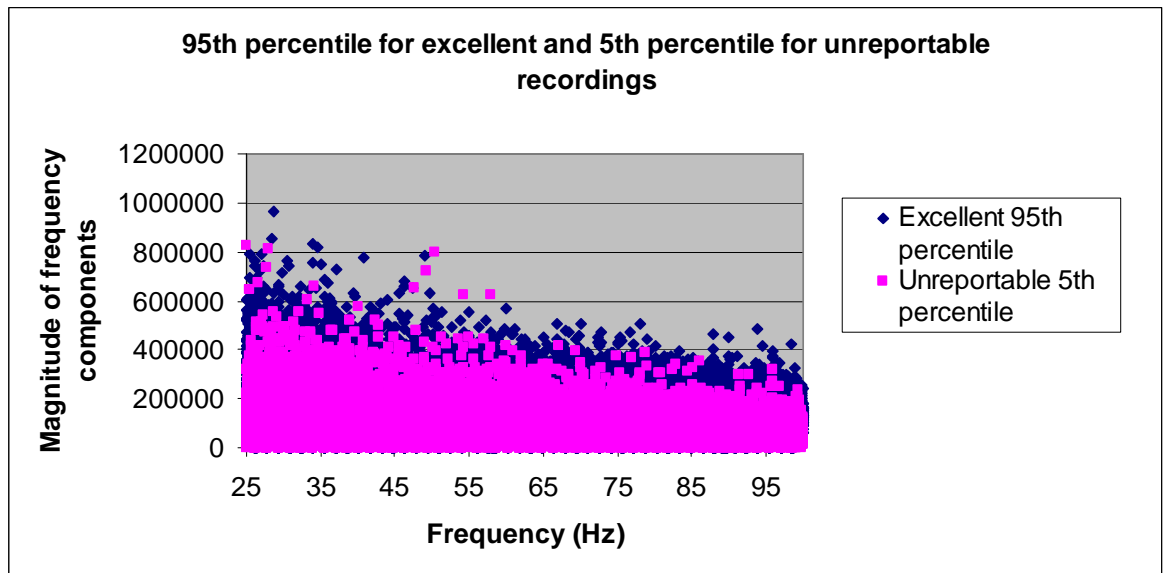
Figure 4.25 The 95th percentile for the 30 excellent recordings (blue), and the 5th percentile for the 30 unreportable recordings (pink) when every data point was utilised to derive the plots. There is an overlap of these plots thus excellent and unacceptable recordings are not distinct from one another when using every data point to calculate the 5th and 95th percentiles.

It was not therefore possible to use this method to differentiate excellent from unreportable recordings. When the 95th percentile for the excellent recordings and the 5th percentile for the unreportable recordings were calculated using averaged values (every 32 data points) the following plot was obtained:
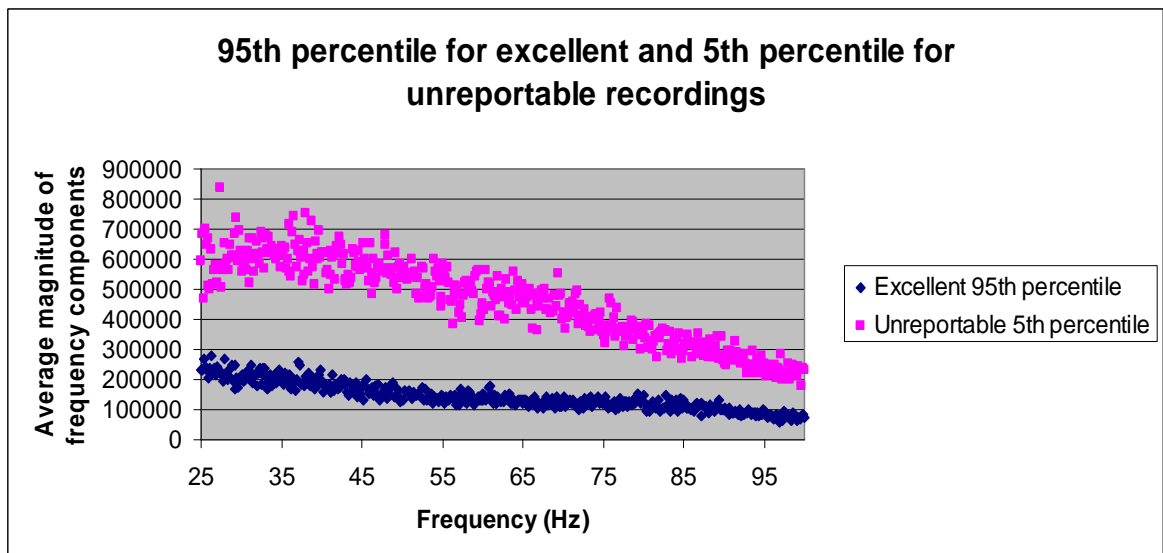
Figure 4.26 The 95th percentile for the excellent recordings (blue) and 5th percentile for the unreportable tests (pink) when every 32 data points were averaged prior to calculating the percentiles. It should be noted that the limits defining excellent and unreportable are now distinct from one another.

These curves are distinct from one another. All subsequent plots therefore used an average of 32 data points prior to plotting the percentile curves (i.e. when categorising the recording quality of all future tests).

4.3.9.2 Defining limits for moderate and noisy recordings

The midline, calculated by averaging the value of the excellent 95th percentile and the unreportable 5th percentile at each point is shown in figure 4.27:
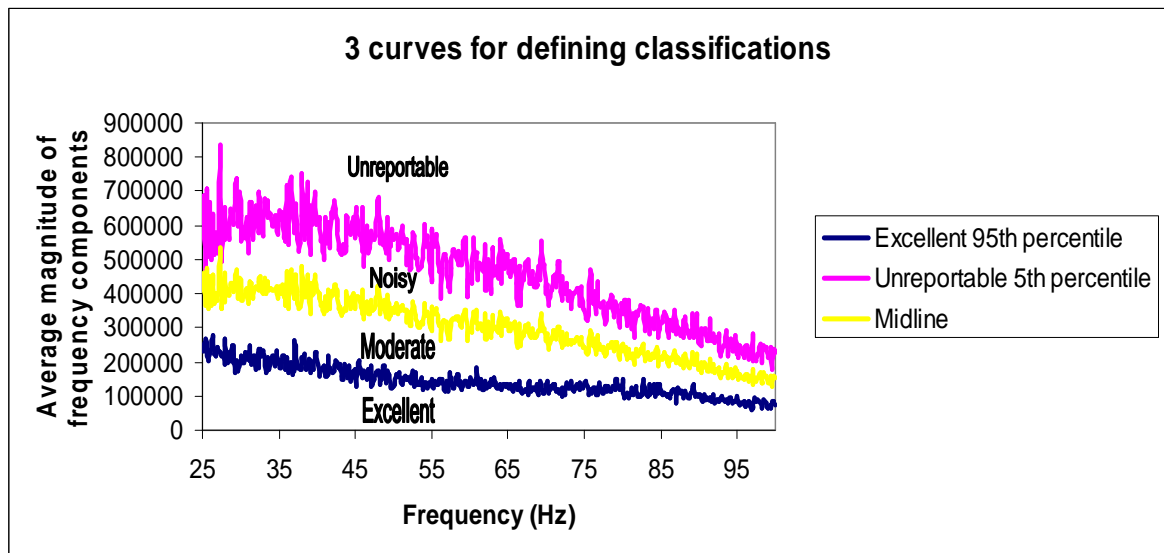
Figure 4.27 The four classifications (excellent, moderate, noisy and unreportable) for recording quality based upon the magnitude of the frequency components in the range 25-100Hz. The excellent 95th percentile (blue), the midline (yellow) and the unreportable 5th percentile (pink) curves are shown. The quality classifications are defined as: those recordings with a frequency profile below the blue limit are 'excellent'; those between the blue and yellow curves are 'moderate'; those between yellow and pink are 'noisy'; those with a frequency profile greater than the pink limit are 'unreportable'.

The area under the curve was 5.22×10^7, 13.00×10^7 and 20.77×10^7 for the excellent 95th percentile, the midline and the unreportable 5th percentile plots respectively. These values were utilised to classify the integrity of future recordings in terms of patient muscle movement during the testing session.

4.3.10 Results: muscle movement for one segment of a recording

When one segment was analysed to define limits for 'unreportable' and 'acceptable' the following plot was obtained:
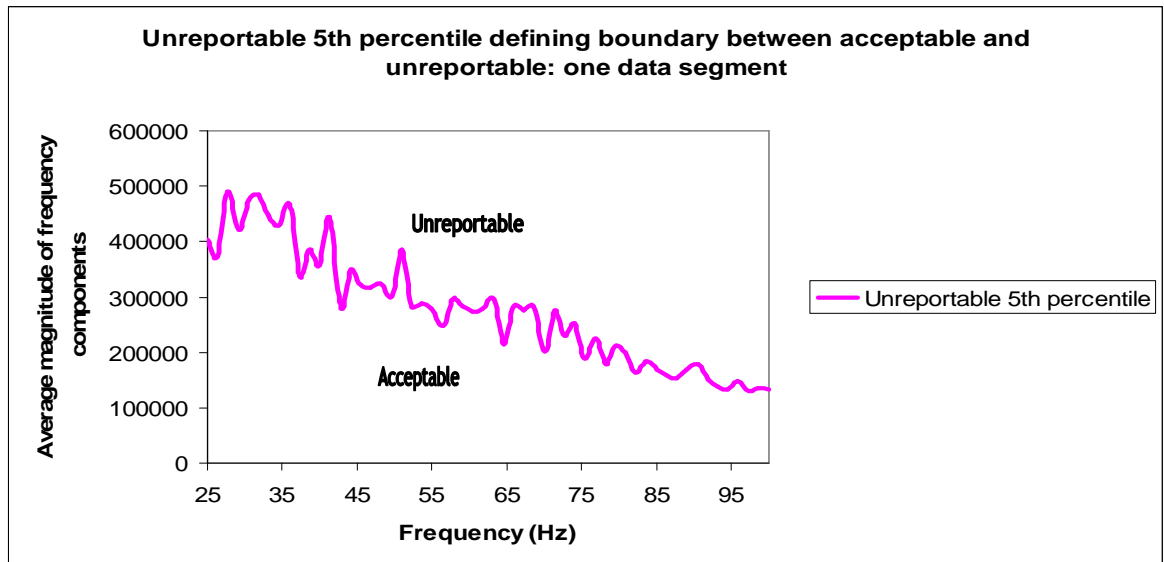
Figure 4.28 The two classifications (acceptable and unreportable) for one segment of a recording in terms of patient noise caused by muscle movement. Those recordings with a Fourier profile above the 5th percentile of the group of unacceptable recordings are categorised as 'unreportable' while those below the limit are said to be 'acceptable'.

In this instance every two data points were averaged to acquire the 5th percentile curve for the unreportable recordings. The area under the curve was 15.49×10^6; this was utilised to define the quality of other mfERG tests.

4.3.11 Results: 50Hz noise for a complete recording and one segment of a recording

For all ten recordings it was observed that the magnitude of the 50Hz component exceeded 8×10^7 when studying the full recording and 2×10^7 when only one data segment was viewed in the Fourier domain. The operator can therefore be warned if any patient recording contains a 50Hz component greater than these values.

4.3.12 Combining and testing the system

Table 4.1 presents the ability of these limits to grade the recording quality of 50 complete mfERG tests when compared with the opinion of three experts:

		Computer's classification			
		Excellent	Moderate	Noisy	Unreportable
Overall expert grading	Excellent	**12**	5	0	0
	Moderate	4	**8**	4	0
	Noisy	0	3	**4**	1
	Unreportable	0	0	2	**7**

Table 4.1 The agreement between the computer and the experts when grading the recording quality for a complete test as excellent, moderate, noisy or unreportable. The computer and the expert agreed for 31 of the 50 recordings (62%).

It can be seen that the expert and the computer agreed on the grading for 31 of the 50 recordings (62%). For each of the 19 examinations where there was disagreement in the classification of recording quality the computer was always within one grading of the overall expert opinion. For 13 of these 19 tests the computer agreed with one of the human experts.

It should be noted that of the 50 recordings assessed by the three experts, a unanimous classification was only achieved for 21 of the tests. For each example at least two people gave the same classification with the third person grading it either one class above or below the others. The overall Kappa value of agreement between the observers was 0.47 corresponding to a moderate level of agreement (127). The Kappa value for each of the four classifications of recording quality (excellent, moderate, noisy and unreportable) was 0.47, 0.23, 0.46 and 0.83 equating to moderate, fair, moderate and almost perfect agreement respectively (127).

It was of interest to assess if the variability between the system and the experts was comparable with that between the three experts; overall Kappa values of 0.48, 0.46 and 0.43 were calculated when each of the two experts and the computer were compared; this is similar to that of the three experts (0.47).

When interpreting the results the most important classification is arguably whether or not a test is of a sufficient standard to report. An agreement of 94% between the experts and the system was achieved for this distinction; 78% of the mfERG tests said

to be unreportable by the experts were given the same classification by the computer while 22% were categorised as being noisy but reportable.

The three experts and the system were then presented with one data segment from each of the 50 mfERG recordings, classifying the integrity of these segments as acceptable or unreportable. The following table compares the majority expert grading with that of the system:

		Computer's classification	
		Acceptable	Unreportable
Overall expert grading	Acceptable	32	5
	Unreportable	3	10

Table 4.2 The agreement between the computer and the experts when classifying the recording quality of one segment of the mfERG as acceptable or unreportable. They concurred for 84% of the tests.

An agreement of 84% was observed between the system and the experts. Of those eight segments for which there was disagreement, the computer agreed with one of the experts for four of them.

An overall Kappa value of 0.64 was calculated for the agreement between the three experts when categorising the data segments into one of two groups; this corresponds to a substantial agreement (127). As when studying the complete recording, it was of interest to assess the agreement between the system and two of the experts. Overall Kappa values of 0.63, 0.61 and 0.60 were calculated; these are similar to that when studying the three human experts.

4.3.13 Discussion

When defining limits to classify the integrity of a recording there was no clear distinction between excellent and unreportable recordings when percentile curves were plotted using every data point. However, when values were averaged, excellent and unreportable recordings were distinct from one another in the frequency domain. This is a reflection of the fluctuation in the magnitude of frequency values; an overall trend was required therefore averaging enabled this to be established.

Prior to presenting the 50 new mfERG recordings to the system, three experts graded the recording quality of each test as excellent, moderate, noisy or unreportable. It became evident that this is both a difficult and a subjective task given that there was only moderate agreement between the three experts. The most consistent grading between the three people was in the case of recordings thought to be too poor to report, when the agreement increased to almost perfect. These findings emphasise the requirement for a more objective method of assessing the integrity of a mfERG recording.

The 50 mfERG tests were subsequently presented to the system and categorised into one of the four groups detailed above. This enabled the classification of the experts and the system to be compared for each recording. It was shown that the computer and experts only agreed on the classification of recording quality for 31 (62%) of the tests. However given the degree of inconsistency between the three people it is difficult to argue that the classification provided by the experts is definitive. It should be noted that the computer agreed with at least one of the experts for 44 of the 50 (88%) recordings. Furthermore, a similar level of agreement was seen between the system and two experts as that reported between the three experts (by assessing their Kappa values). It is therefore difficult to truly assess the efficacy of the system for grading the integrity of a recording, given that its agreement is comparable with that of another expert.

It was of interest that of the 19 cases where there was disagreement between the system and the overall expert opinion, ten were said to be one grading worse by the computer while nine were classed as one grading better. It would be of greater concern if the computer consistently reported the quality as better or worse than the humans as this would indicate that the limits defined in section 4.3 were not suitable. It can be argued that the method by which the computer defined the recording quality was less subjective than that used by the humans as it is simply calculating the magnitude of the frequency components in the Fourier domain. The discrepancies could therefore be a reflection of the subjectivity of the experts assessing the integrity of a recording by viewing the uncorrelated data and the trace array of cross correlated responses in the time domain.

One data segment (1/16th of the recording) was then selected from each of the 50 tests; these were presented to the three experts and the system. In this instance each categorised these 50 data segments as either acceptable or unreportable in terms of

their recording quality. A higher agreement was seen between the three experts (substantial, as opposed to moderate), however they were only required to classify the data into one of two groups as opposed to four, thus minimising the scope for disagreement. The Kappa value of agreement (0.64) was however lower than that calculated when the experts made the equivalent distinction between reportable and unreportable tests when examining the complete recording (0.83). This is likely to be attributed to the fact that when studying the entire recording, both the uncorrelated data and the trace array (comprising the 61 cross correlated responses) were viewed, whereas only the raw, uncorrelated data were assessed when examining one data segment. The raw data were much less familiar than the trace array to the experts thus increasing the difficulty of the task for the experts.

When comparing the system with the overall expert opinion, an agreement of 84% was observed. Of those data segments misclassified by the system, five were given a worse grading by the system, while three were better. As was seen when examining the complete recording, the system's grading was not therefore consistently better or worse than that of the experts.

Finally, the Kappa value of agreement between the three experts was comparable with those values calculated when two experts were compared with the system. It is therefore difficult to argue that the classifications provided by the system were inherently wrong in those instances when there was a disagreement between it and the overall expert opinion.

These results imply that the system can be used with relative confidence when assessing the quality of a recording, both when studying the entire recording and one segment of the data. The system performed particularly well when deciding if a completed test should be reported. Further investigation is however required to improve confidence in the results. The number of examples used to define the limits for excellent and unreportable when studying the complete recording could for example be increased to encompass a wider variety of recording qualities. Or alternatively the number of classifications into which the recording quality is categorised could be decreased; this may improve the agreement between the experts and the system, thus leading to greater reliance on the computer's assessment of recording integrity.

4.3.14   Conclusions

A method for grading mfERG recording quality based upon the Fourier profile of the uncorrelated data has been proposed. Good agreement was seen between the experts and the computer when defining if a completed recording should be reported. Differences were however observed when subclassifying the recording quality and should be the subject of further investigation. A relatively high level of agreement was also observed between the experts and the system when assessing the integrity of smaller sections of the recorded data.

4.4 Investigation into the mfERG Fourier profile

One of the main aims of this chapter was to investigate if the Fourier profile of the raw mfERG data can be used to develop a method for grading recording quality, both during and after a test. In doing so the frequency profile associated with the mfERG stimulus was also found, revealing the frequencies at which the retina was stimulated by, and responded to. To fully understand why this particular frequency distribution was observed further experiments were conducted.

4.4.1 Methods

A number of test parameters fundamental to the stimulus were varied, allowing their impact on the Fourier profile to be studied. This enabled the factors determining the frequency profile to be established. A photodiode was stimulated, with each investigation based on the mfERG protocol described in section 3.4.1. As when isolating the signal in section 4.2.4.2 an amplifier gain of 100 was chosen. The shift between the m-sequences used to drive the stimulus was changed initially. The number of elements forming the stimulus was then varied, followed by the stimulus frequency, and finally the type of stimulating device. As before the raw, uncorrelated data were transformed into the frequency domain using a FFT (appendix 1), allowing the Fourier profiles to be studied.

4.4.1.1 Alternative shift between m-sequences

The set of orthogonal m-sequences used in the mfERG protocol were created by decimating the original m-sequence over 128 columns. By changing this to 256

columns the shift between the m-sequences was altered, therefore changing the stimulus pattern throughout the test. These were used to control the stimulus.

4.4.1.2 Varying the number of stimulus elements

Thus far all stimuli have comprised 61 stimulating elements. This was changed to 1, 7 19, 37 and 103 respectively to establish if the number of elements affects the principal stimulating frequencies. Figures 4.29 illustrate these alternative set ups:
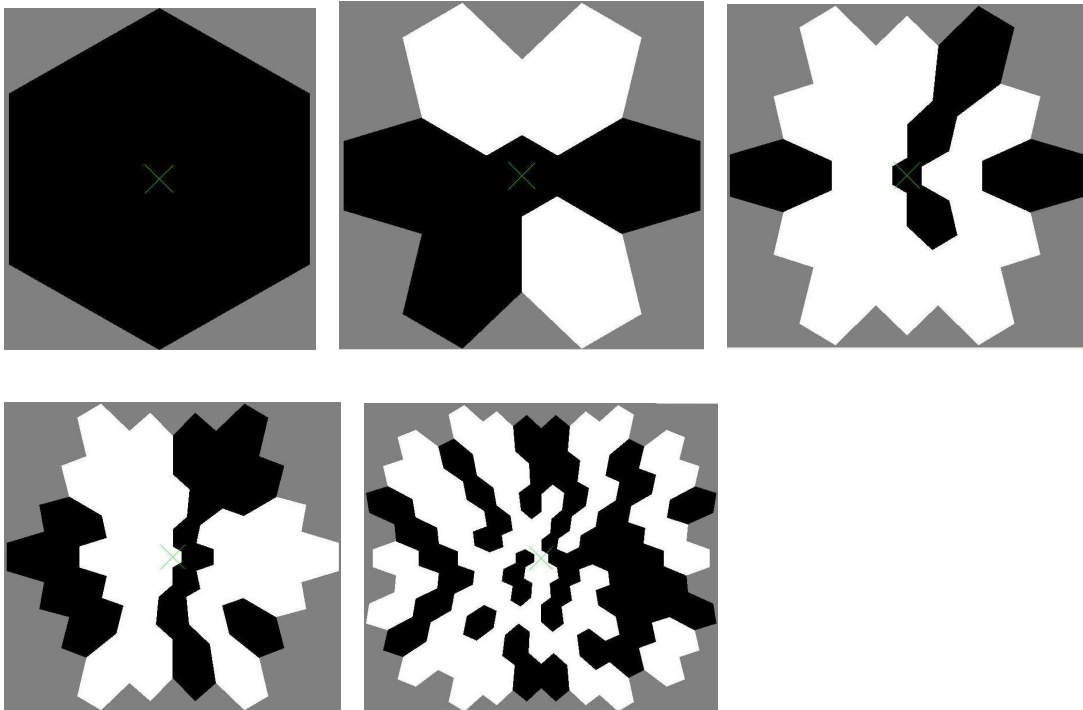
Figure 4.29 Illustrating five mfERG stimuli, each comprising a different number of elements. From top left to bottom right: 1 element; 7 elements; 19 elements; 37 elements; 61 elements; and 103 elements.

4.4.1.3 Different stimulating frequency

A stimulus frequency of 60.8Hz was selected as opposed to 75Hz. A sampling frequency of 972.8Hz, 16 times the stimulus frequency, was chosen.

4.4.1.4 Change of stimulating device

All tests thus far have utilised an LCD stimulus; a CRT device was therefore selected to evaluate the impact of the stimulator on the frequency spectrum. Tests were conducted using 1, 7, 19, 37, 61 and 103 stimulating elements.

4.4.2 Results

4.4.2.1 Alternative shift between m-sequences

When these sequences were used to control the stimulus the following frequency distribution was obtained:
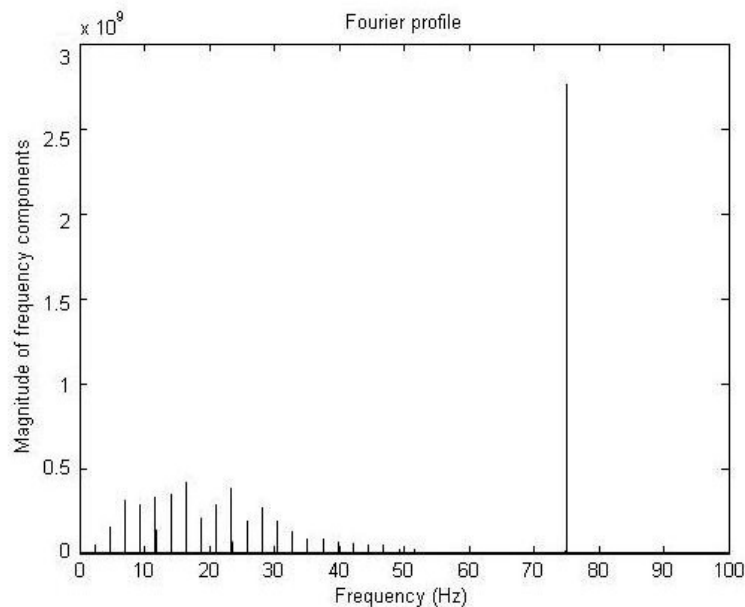
Figure 4.30 The Fourier profile obtained from a photodiode when the shift between the m-sequences controlling the stimulus was changed. A dominant frequency component is evident at 75Hz, the stimulation frequency, in addition to frequency peaks separated by a discrete frequency. This separation is the stimulus frequency/32.

Again the strongest contribution was at 75Hz, the stimulus frequency. As before discrete frequency components were present, however they were now separated by the stimulus frequency/32.

4.4.2.2 Varying the number of stimulus elements

The following figure demonstrates the Fourier profile when using one stimulating element:
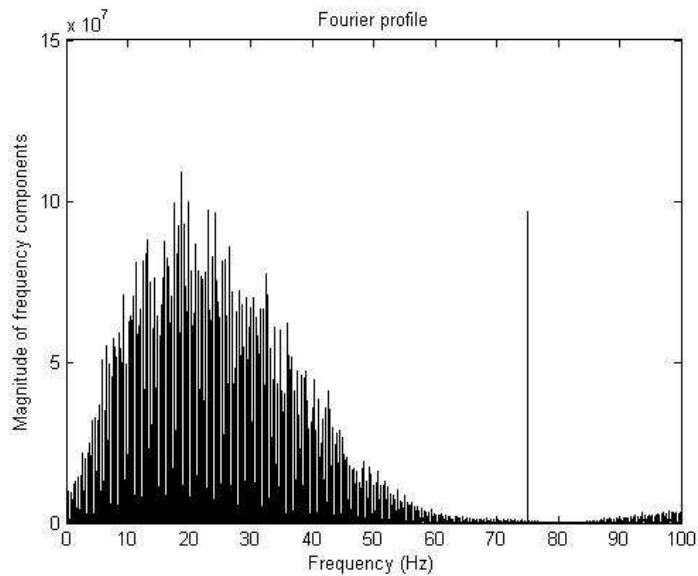
Figure 4.31 The Fourier profile when one element was utilised to stimulate the photodiode; an LCD stimulator was employed. A dominant peak is present at the stimulus frequency (75Hz), in addition to discrete frequency peaks, each separated by the stimulus frequency/192.

As before a distinct contribution was observed at 75Hz, the stimulus frequency ($F_{stimulus}$). Discrete frequency peaks were also noted, however the separation between each of the peaks was greatly decreased; each peak was distributed $F_{stimulus}$/192 apart. It was noted that the magnitude of the stimulus frequency was less than that of the largest discrete peak. The following Fourier profile demonstrates the frequency spectrum found when the photodiode was stimulated by seven elements. A different scale has been utilised; this is to account for the increased magnitude of the component at the stimulus frequency:
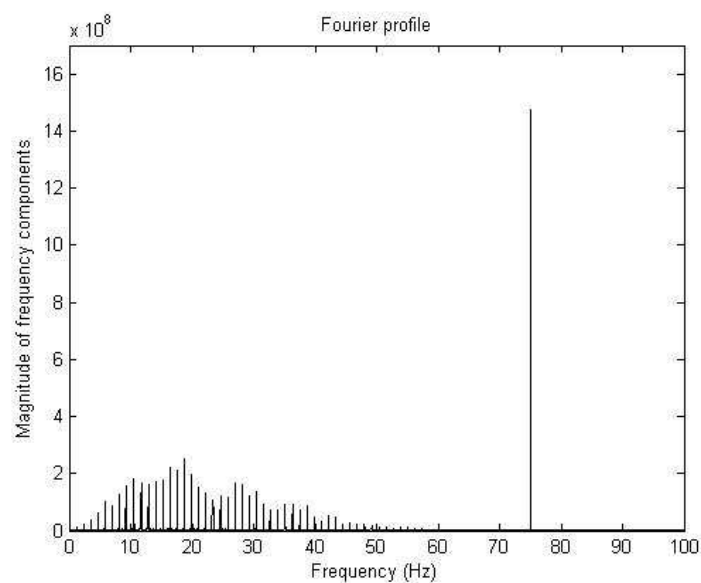
Figure 4.32 Fourier profile when an LCD stimulator presenting a seven element stimulus was utilised to stimulate a photodiode. A dominant peak is evident at 75Hz, the stimulus frequency. Discrete frequency peaks were also present, each separated by the stimulus frequency/64.

It was noted that the largest peak was at the stimulus frequency, in addition to a normative-like distribution of discrete frequency peaks, each separated by the stimulus frequency/64. A similar pattern was observed when the stimulus comprised 19, 37 and 103 elements.

4.4.2.3 Different stimulating frequency

Again a prominent contribution was seen at the stimulus frequency, 60.8Hz. Frequency peaks, separated by $F_{stimulus}/64$ were noted. The pattern seen was similar to that when the photodiode was stimulated using 75Hz, the only difference being the specific frequencies of the peaks.

4.4.2.4 Change of stimulating device

A CRT device was selected to display the mfERG stimulus, while the number of elements was varied. Figure 4.33 illustrates the Fourier profile when one element was chosen:
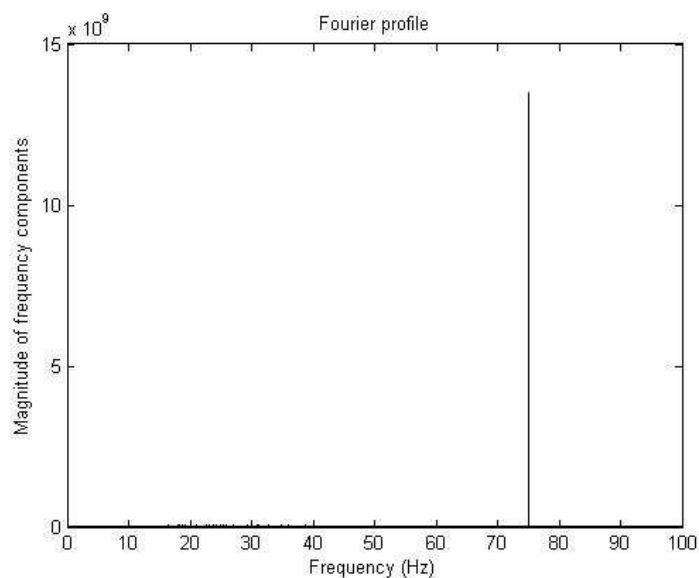
Figure 4.33 The Fourier profile when a photodiode was stimulated by a CRT device displaying a one element stimulus. Only the peak at the stimulus frequency (75Hz) can be seen as its magnitude is significant relative to all other frequency components.

A dominant component was seen at $F_{stimulus}$. Discrete frequency peaks were also present however their magnitude was significantly smaller than that of the 75Hz peak. This was the case for each stimulus pattern therefore all subsequent Fourier profiles obtained when using a CRT stimulator are displayed using a decreased y-axis range to

allow these peaks to be viewed. The discrete peaks present when one stimulating element was used can be seen in figure 4.34:
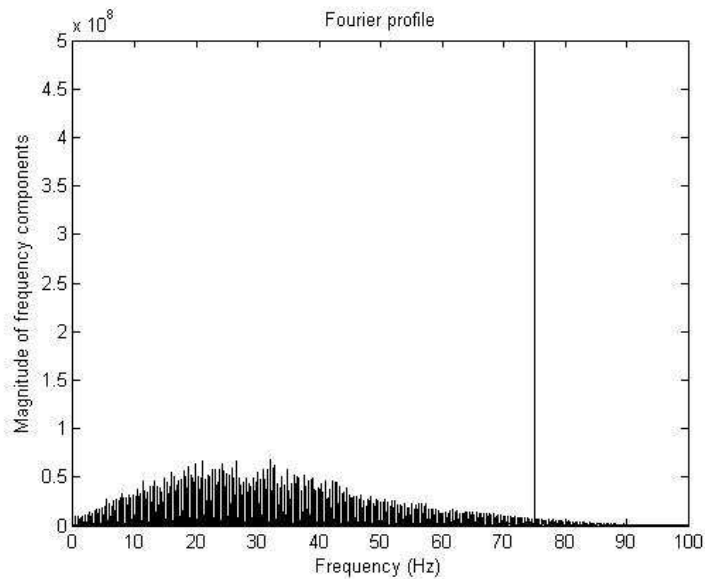
Figure 4.34 The Fourier profile with a decreased y-axis when a photodiode was stimulated by a CRT device displaying a one element stimulus. A peak is present at 75Hz; peaks separated by the stimulus frequency/192 are also evident.

As was observed when the LCD device was used, each discrete peak was separated by $F_{stimulus}$/192. The range over which the peaks were seen was greater than that noted when stimulated using the LCD device. When the photodiode was tested using a seven element stimulus the following profile was obtained:
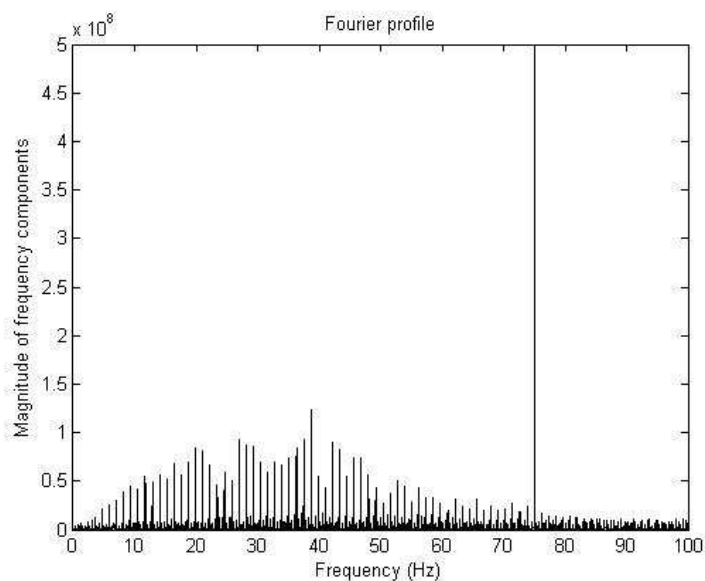
Figure 4.35 The Fourier profile when a photodiode was stimulated with a seven element stimulus by a CRT device. A peak is present at 75Hz; peaks separated by the stimulus frequency/64 are also evident.

Again the most prominent peak was at 75Hz. Discrete frequency peaks, each $F_{stimulus}$/64 apart were noted. The frequency range of the peaks was greater than that

observed when studying the equivalent LCD Fourier profiles. A similar pattern was seen when using a stimulus comprising 19, 37, 61 and 103 elements.

In summary it was found that:

- a peak was present at the stimulus frequency for all experiments;

- discrete frequency peaks were evident for all testing parameters;

- changing the shift between the underlying m-sequences changed the separation between the discrete peaks;

- using one element as opposed to a multi-element stimulus altered the distribution of the frequency peaks;

- altering the stimulus frequency changed the position of the peaks but the separation remained $F_{stimulus}/64$ when using a 61 element stimulus;

- the separation between peaks was the same when using the CRT and LCD stimulating devices, however the range of peaks was greater when the CRT was chosen;

- the stimulus frequency was the dominant frequency with the exception of the one element LCD experiment;

- the magnitude of the peak at the stimulus frequency relative to that of the peaks was greater when using a CRT device than an LCD stimulator.

4.4.3 Discussion

For each experiment there was a peak at the stimulus frequency. This was to be expected as it was the rate at which the stimulus patterns were updated during the test. Of most interest was the finding that the discrete nature of the frequency peaks was altered by changing the shift between the m-sequences or by using one element instead of multiple elements.

When a single element is chosen one m-sequence is used to control the luminance of the screen throughout the test. The transition to high luminance evokes the greatest change in potential in the uncorrelated data. This is illustrated below for a very short part of a sequence using data acquired from a patient:

Figure 4.36 Demonstrating a small portion of an m-sequence (left) and the corresponding response (right). On the left, 5 steps of an m-sequence are illustrated, with white representing the on state (1) and black being the off state (0). The trace on the right demonstrates the response evoked from this short part of the m-sequence.

The fundamental frequencies observed when the entire raw data trace is transformed to the frequency domain are therefore determined by the separation between each transition to high luminance. When multiple elements are used each element is driven by a shifted version of the original m-sequence. An example of the evoked retinal responses for four different sequences can be seen in figure 4.37:
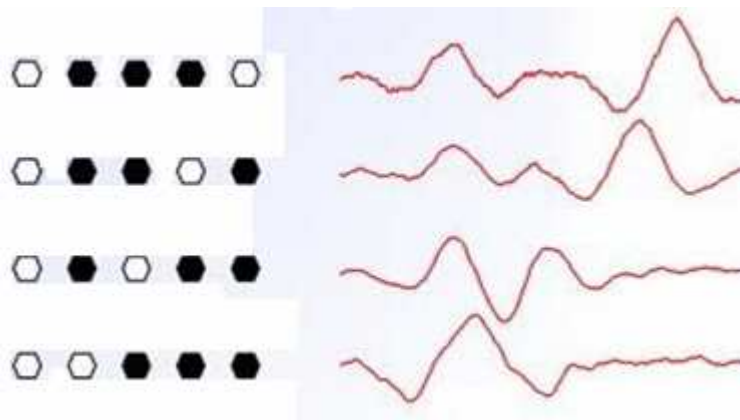
Figure 4.37 Illustrating the responses derived from stimulation at multiple frequencies. On the left, small portions of the m-sequence are shown, with black representing the off state (0) and white corresponding to the on state (1). The resulting responses are shown on the right.

Again the main evoked response is seen when there is a change to a period of high luminance. Each element has a set of fundamental frequencies associated with it, determined by its m-sequence. As the uncorrelated data are a global response to the stimulus it is a superposition of the response to each element. The frequency components of the uncorrelated data were therefore a superposition of the principal frequencies for each element. Consequently these were different to that of a single element, hence the Fourier profile changed when several elements were chosen to stimulate the photodiode instead of one element.

When the sequences used to control each element were created using 256 instead of 128 columns for the decimation process, the relative shift between each of the orthogonal sequences was altered; consequently the stimulus patterns throughout the test differed to those of the original set up thus affecting the fundamental frequencies present. This explains why a different Fourier profile was revealed when using these alternative sequences.

Neither varying the stimulating device nor the stimulating frequency changes the underlying m-sequences. The fundamental frequencies associated with each m-sequence were therefore unaffected and hence the superposition of these fundamental frequencies was comparable with that seen using the mfERG protocol. This explains why the separation between each discrete peak (in relation to the stimulus frequency) was unaffected by the change in stimulation frequency and the method of stimulation.

The differences observed in the Fourier profiles for the CRT and LCD experiments can be explained by the difference in their illumination profile when there are consecutive 1s in the m-sequence (figure 2.5): individual flashes of light are produced by the CRT whereas the luminance remains constant when using an LCD stimulating device.

When there is a 0-1 step in the m-sequence both the LCD and CRT change to a period of high luminance and contribute to the peak at the stimulus frequency seen in the photodiode Fourier profiles. However when there are successive 1s in an m-sequence a CRT device produces a series of pulses at the stimulus frequency. In contrast the LCD device produces a period of high luminance for the duration of the 1s in the sequence. The contribution to the stimulus frequency will therefore differ for each device: a peak is seen at the stimulus frequency for the CRT but not for the LCD when there is a 1-1 step in the m-sequence. This explains why the relative magnitude of the peak at $F_{stimulus}$ was greater when using the CRT stimulating device: it was generated for 0-1 and 1-1 steps in m-sequence as opposed to 0-1 when using an LCD stimulator. This also explains why the peak at 75Hz was not the dominant frequency when stimulating the photodiode with 1 element driven by an LCD device: for much of the test no contribution was made to the 75Hz peak as only a 0-1 step in the m-sequence added to the magnitude of the stimulus frequency peak. When using multiple elements it was always the case that at least one element was changing from a period of low to high luminance therefore each step in the test contributed to the magnitude of the stimulus frequency peak thus it was dominant for all multi-element stimulus set ups. The

frequency range of the discrete peaks was greater when using the CRT device than the LCD device for all numbers of stimulating elements. This is again a result of the difference between the illumination profile of each when driven by a 1-1 part of the m-sequence. The two separate flashes produced by a CRT device as opposed a continuous period of high luminance increases the non-linearity of the output thus producing higher frequencies.

4.4.4 Conclusions

It was demonstrated that using a single as opposed to a multi-element stimulus, or changing the shift between the orthogonal sequences altered the frequency spectrum of the mfERG stimulus. It was concluded that the Fourier profile is determined by the superposition of the fundamental frequencies associated with the orthogonal sequences used to control each stimulating element. Differences between CRT and LCD devices were apparent and were attributed to differences in the illumination profile when set to a period of high luminance.

4.5 Effect of compromised retinal function on the Fourier profile

Having established the frequency profile particular to the mfERG stimulus and why this pattern is seen it was of interest to assess the affect of compromised retinal function on the frequency spectrum of the uncorrelated data. The hope was that this method would enable the simple extraction of temporal information from the mfERG data.

4.5.1 Methods

50 recordings, acquired using the mfERG protocol (section 3.4.1) were selected retrospectively. These recordings encompassed a wide range of retinal function. The processed, cross correlated waveforms were analysed by an expert for each test, with comments made on the amplitude and latency of the P1 component. The uncorrelated data were then transformed into the Fourier domain for each test. Six examples have been used to demonstrate the main findings. These include: diffuse amplitude reductions; reductions and moderate delays; reductions and significant delays; an absent N2 component; no significant retinal function and a recording with a localised

area of dysfunction, respectively. The Fourier profile was compared with that acquired from a healthy retina in each case. As a reminder the Fourier profile taken from a control subject is shown in figure 4.38:
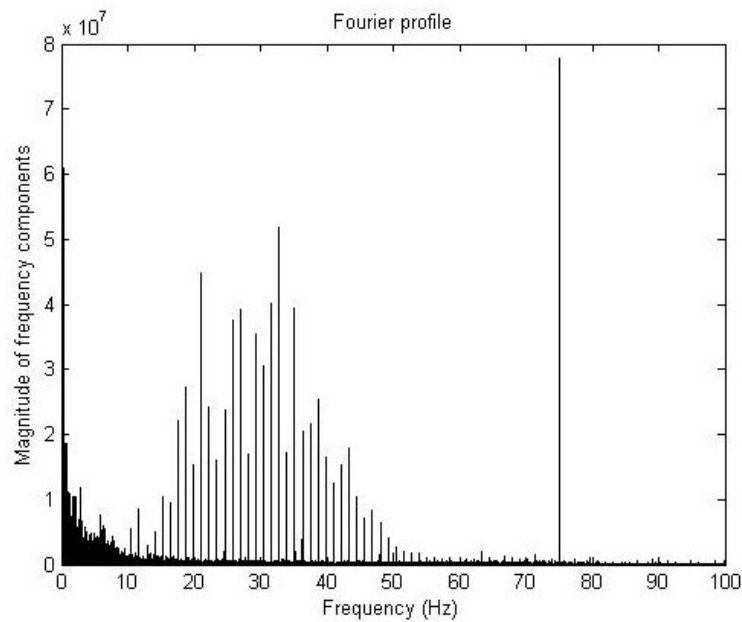
Figure 4.38 The Fourier profile obtained from a healthy and compliant subject. The contribution at the low frequencies is relatively small; this is comparable with the noise profile obtained from the compliant unstimulated subject (figure 4.8). Again, a dominant peak is present at the stimulus frequency (75Hz), in addition to discrete frequency peaks, each of which is separated by the stimulus frequency/64.

4.5.2 Results

It was found that the various mfERG abnormalities affected the frequency spectrum differently. For each example the correlated responses are displayed in the time domain along with the uncorrelated data in the frequency domain.

4.5.2.1 Case 1: diffuse amplitude reductions
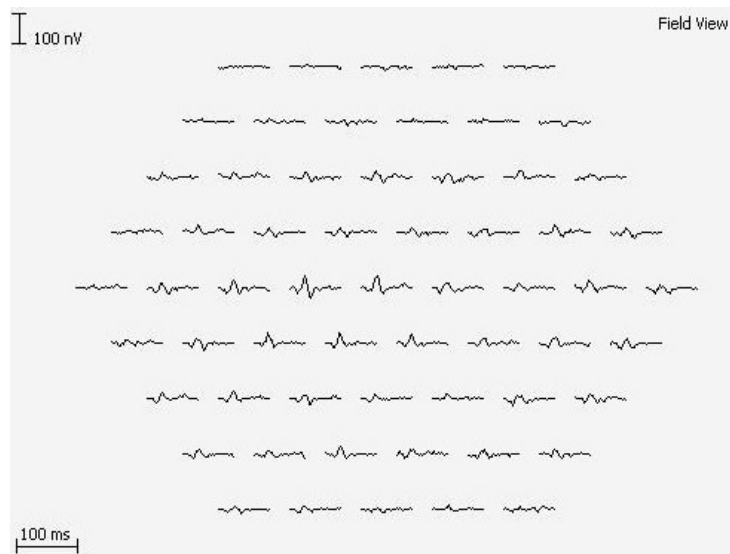
Figure 4.39 A trace array with a diffuse reduction in the amplitude of responses. The P1 amplitude of the cross correlated responses was decreased relative to the normal range defined in section 3.4.2; this was the case for all waveforms.

The trace array showed diffuse P1 reductions, while P1 latencies were within normal limits. When the uncorrelated data were transformed into the frequency domain the following profile was obtained:
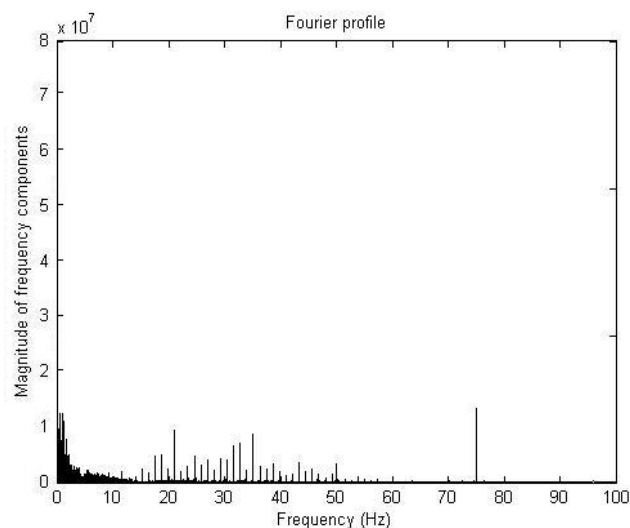
Figure 4.40 The Fourier profile acquired from a test with a diffuse reduction in response amplitudes. Both the peak at the stimulus frequency and the discrete peaks can be seen; it should be noted that the magnitude of these is decreased relative to that in figure 4.38.

A 75Hz component was seen in addition to discrete peaks within a similar frequency range to those of the healthy control (figure 4.38). The magnitude of each of these was however noticeably smaller than those seen in the control subject.

4.5.2.2 Case 2: diffuse amplitude reductions and moderate delays
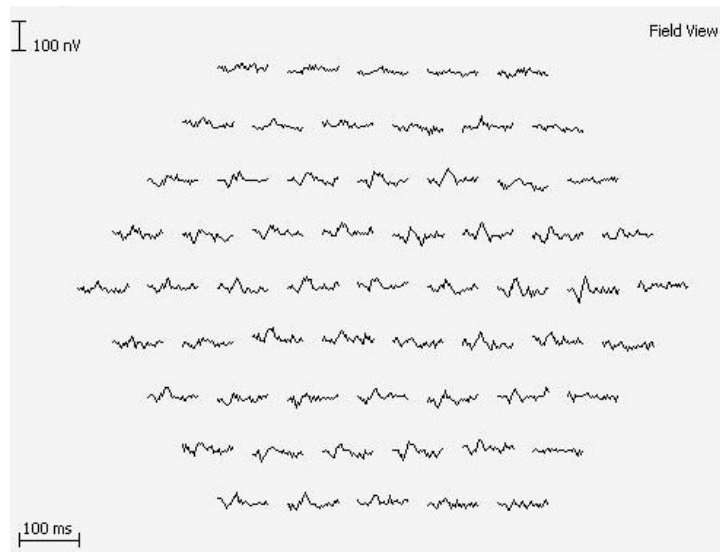
Figure 4.41A trace array comprising responses which are decreased in amplitude and moderately delayed.

Diffuse P1 reductions and moderate delays were noted. The Fourier profile of the raw data is shown in figure 4.42:
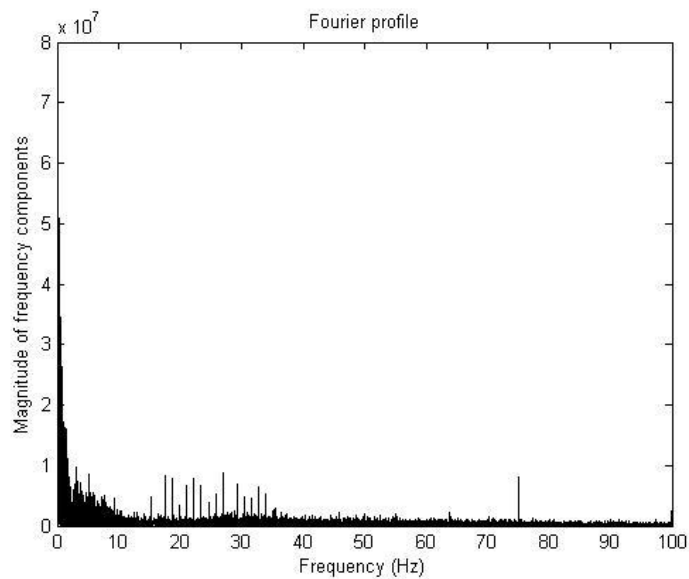
Figure 4.42 The Fourier profile obtained from a recording with diffuse amplitude reductions and moderate delays. The frequency range of the discrete peaks is decreased relative to that seen in figure 4.38. Both the discrete frequency peaks and the component at the stimulus frequency were decreased in amplitude relative to those in figure 4.38.

Discrete peaks were observed however those greater than approximately 35Hz were abolished. Again the peaks present were smaller than those seen in the healthy subject. A peak at 75Hz was evident.

4.5.2.3 Case 3: diffuse amplitude reductions and significant delays, decreased P1:N2
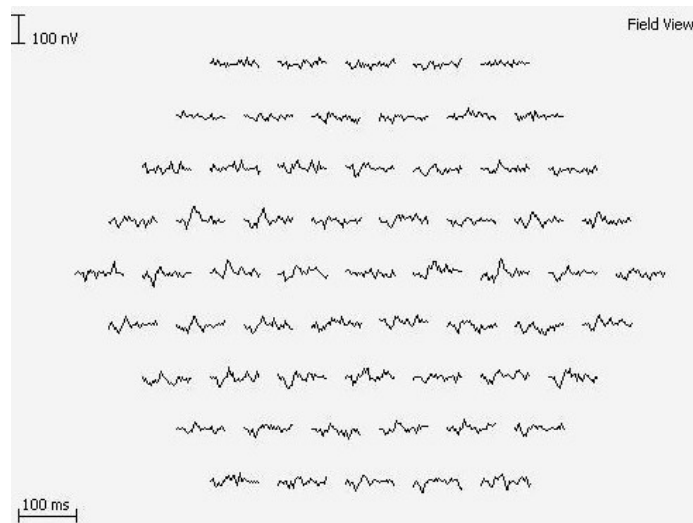
Figure 4.43 A trace array with diffuse reductions in amplitude and significant delays.

The P1 component of the mfERG responses was decreased and significantly delayed. Responses were delayed relative to those in case 2. The amplitude of the N2 component was also decreased. Figure 4.44 reveals the frequency spectrum of the uncorrelated data signal:
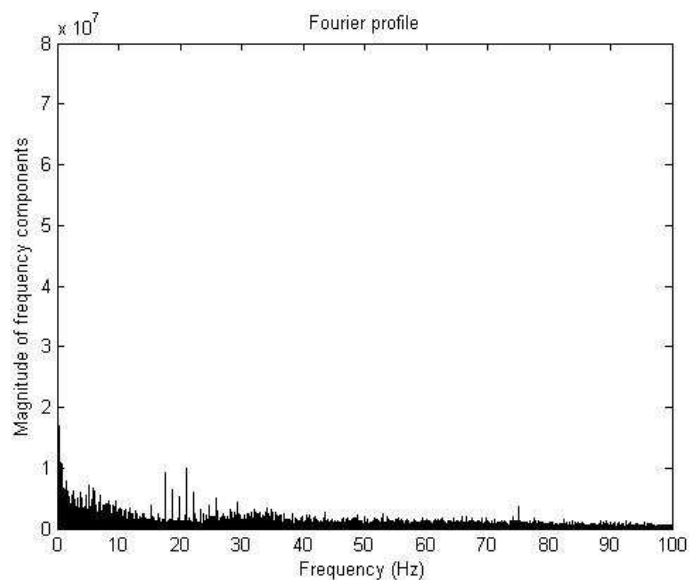
Figure 4.44 The Fourier profile acquired from a recording with diffuse amplitude reductions and significant delays. The frequency range of the discrete peaks has been further decreased; those greater than 30Hz have been abolished. The magnitude of both the discrete peaks and the peak at 75Hz are reduced relative to those acquired from the healthy subject.

Discrete frequency peaks were present however those greater than 30Hz were absent. A frequency component was evident at the stimulus frequency however this contribution was relatively small.

4.5.2.4 Case 4: absent N2 component
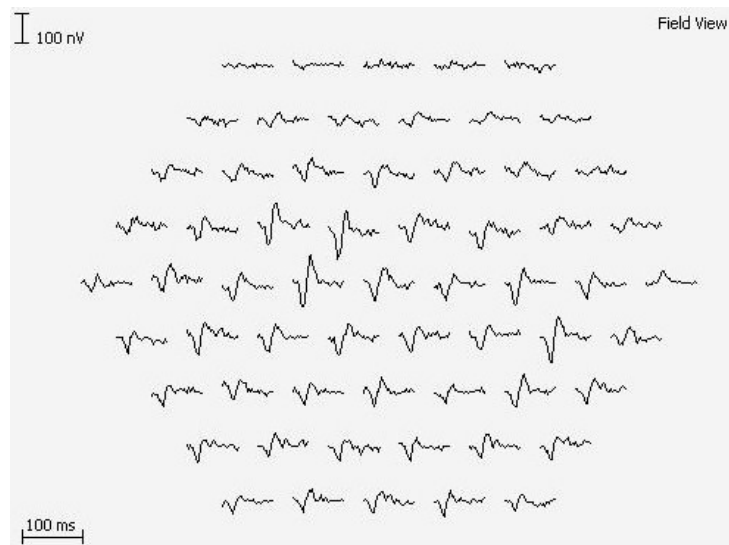
Figure 4.45 A trace array comprising responses with no N2 component. The P1 latency was within normal limits for all responses.

As illustrated in figure 4.45, the N2 component of the waveforms was absent. Responses were within normal P1 latency limits. When the raw data were transformed to the Fourier domain figure 4.46 was acquired:
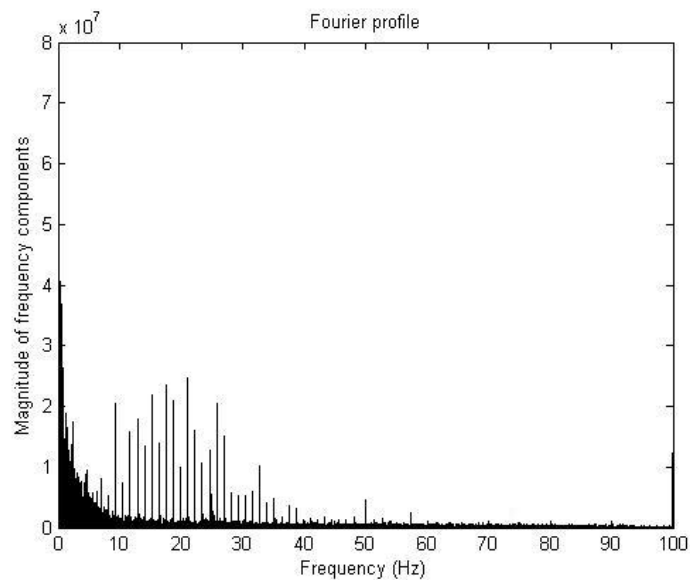
Figure 4.46 A Fourier profile recovered from a recording for which the N2 component was absent in the correlated responses. The peak at the stimulus frequency is not present in this case. The range of the discrete frequency peaks was similar to that of the normal subject.

No peak was seen at the stimulus frequency, 75Hz. Discrete frequency peaks were present within a frequency range similar to that seen in the healthy control.

4.5.2.5 Case 5: no significant retinal function
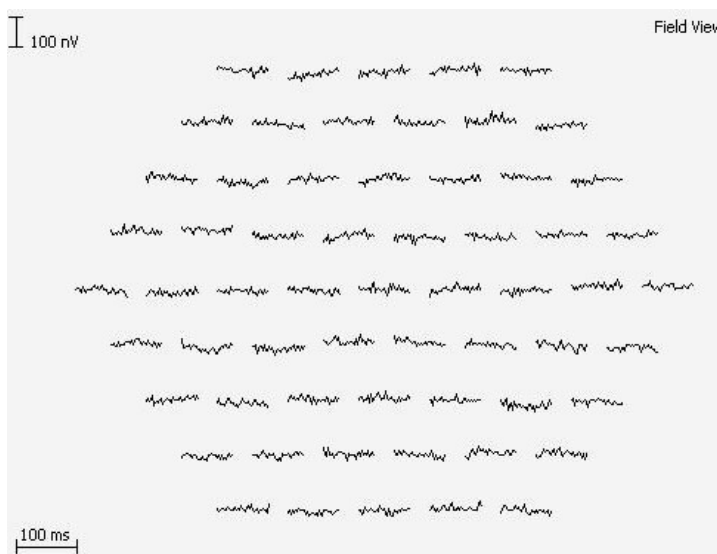
Figure 4.47 A trace array comprising waveforms with no significant retinal function.

No significant responses were observed on the trace array. The Fourier profile of the corneal electrode data is shown in figure 4.48:
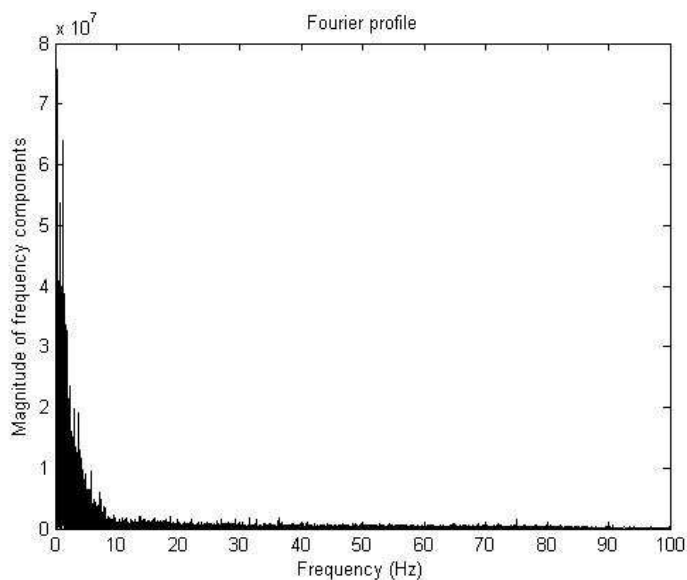
Figure 4.48 The Fourier profile acquired in the case of no significant retinal function. Neither the discrete peaks nor the peak at the stimulus frequency were evident.

In this instance no peaks associated with the mfERG stimulus were seen.

4.5.2.6 Case 6: localised area with no significant response
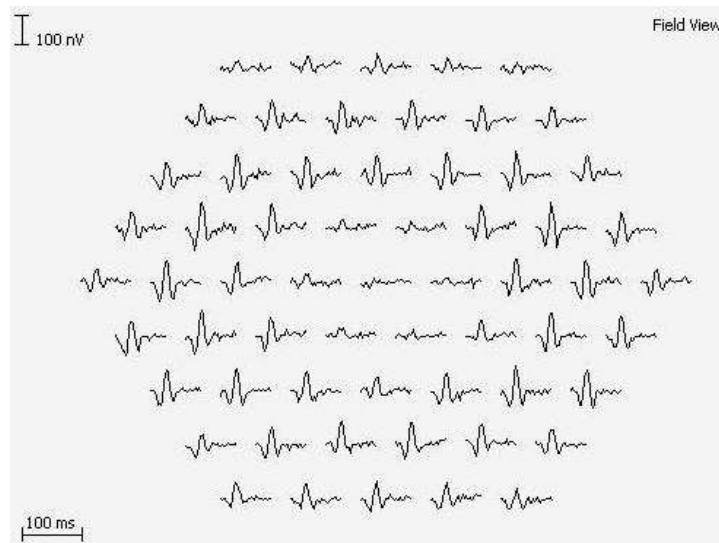
Figure 4.49 A trace array with poor central function and normal surrounding responses. Both the P1 amplitude and latency were within normal limits for the peripheral responses.

The central responses were significantly decreased in amplitude, however peripheral responses were within normal amplitude and timing limits. When transformed to the Fourier domain the following profile was obtained:
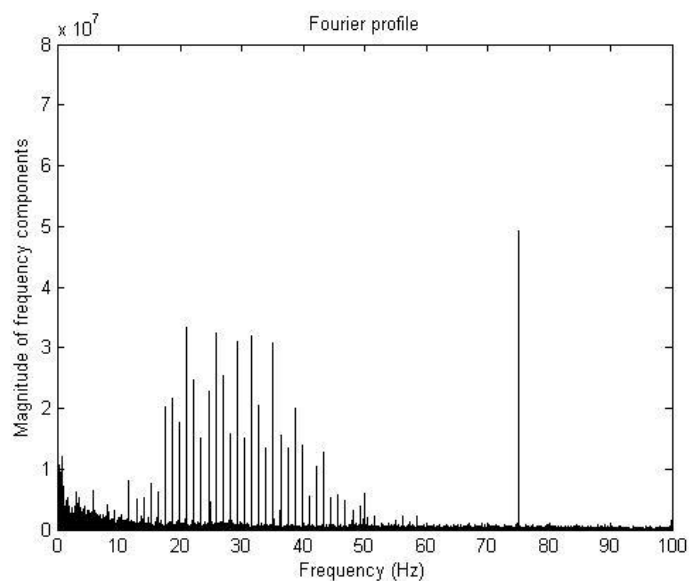
Figure 4.50 The Fourier profile obtained from a recording with a localised abnormality. The Fourier profile has a similar appearance to that acquired from the normal subject in figure 4.38.

A similar profile to that seen from the healthy control was observed: a dominant peak at 75Hz; and discrete peaks in a similar frequency range to those obtained from a healthy retina.

In summary it was found that:

- when there were diffuse delays the discrete peaks in the upper frequency range of that observed in the control subject were abolished. The more severe the delays the greater the loss of the upper frequencies;

- the magnitude of the peaks was decreased in the case of diffuse P1 reductions;

- the peak seen at the stimulus frequency was linked to the N2 component of the correlated responses. For those waveforms with an absent N2 no 75Hz peak was seen while it was decreased for those with a compromised N2;

- recordings obtained from those with no significant retinal function contained no stimulus-associated peaks;

- when a localised area of dysfunction was present the Fourier profile had a similar appearance to that acquired from a healthy retina.

4.5.3 Discussion

Earlier in the chapter it was established that the frequency profile provides information on the retinal stimulation frequencies and retinal response frequencies. It is interesting that for those recordings containing diffuse delays the upper discrete frequency peaks were absent. Furthermore it was found that the frequency range of the peaks was decreased when more severe delays were reported. This implies that the retina was unable to respond to these particular stimulating frequencies.

Another point of note was the abolishment of the peak at the stimulus frequency in recordings with an absent N2 component. This indicates that the retina was unable to respond when stimulated at 75Hz. In a study by Keating *et al.* (37) it was shown that N2 is dominated by the interaction between consecutive stimuli and the P1 component of the preceding stimulus (figure 2.6). When N2 is absent this implies that the retina could not respond to successive stimuli, in other words, those presented at the stimulus frequency. This would explain why the peak at 75Hz was missing in recordings with no N2 component and why it was decreased when N2 was compromised.

For those recordings with no significant cross correlated responses, peaks associated with the mfERG stimulus were absent in the Fourier profile suggesting that the retina could not respond to any of the principal stimulation frequencies.

When there was a gross depression of P1 amplitudes in the time domain a reduction in the magnitude of the frequency peaks was seen. The frequency range of the peaks was however comparable with that of a healthy retina demonstrating that the retina could respond to all of the stimulation frequencies. It is likely that the smaller magnitude can be attributed to the reduced strength of the evoked retinal signal.

The Fourier profile obtained from a recording with compromised central function was similar to that acquired from a normal retina. This is unsurprising as the uncorrelated data are a global signal containing information from all areas of the retina. If the majority of the retina is capable of responding to all stimulation frequencies the Fourier profile will reflect this.

4.5.4 Conclusions

The mfERG responses are conventionally analysed by measuring the amplitude and latency of the three main turning points, P1, N1 and N2. This method of analysis is however insufficient to extract all of the information embedded within the mfERG signal, for example the retina's ability to respond to the different stimulation frequencies. By studying the Fourier profile of the uncorrelated data temporal information can be extracted in a highly simple and visual manner, enabling knowledge of the micro adaptive profile of the retina to be obtained.

4.6 Assessing if a significant mfERG response is present

Thus far the global retinal function and overall recording quality have been studied. In addition to grading the integrity of a recording, one of the principal aims of this thesis was to develop an objective method for differentiating a physiological response from a waveform with no significant response. Cross correlated responses were therefore studied in the frequency domain to investigate if the Fourier profile could be used to reduce the subjectivity of this distinction.

4.6.1 Transforming correlated data from the time to the frequency domain

Prior to transforming the cross correlated waveforms from the time domain to the frequency domain it was important to establish if aliasing or leakage would be a problem; each cross correlated waveform comprised 256 data points with a length of 0.21333s. The stimulus frequency and sampling rate were 75Hz and 1200Hz respectively therefore aliasing was not an issue. Leakage was also irrelevant as the analysis interval contained 16 stimulus periods. The resolution in the Fourier domain, Δf, was $\Delta f = 1/(0.21333) = 4.69$Hz. A program was written using Matlab 2007a to transform the cross correlated mfERG signals from the time domain to the frequency domain using the built in FFT function based on the Cooley and Tukey algorithm (refer to appendix 2). All phase information was discarded and frequencies greater than the Nyquist frequency were ignored due to the symmetrical nature of the data in the Fourier domain. To test that this produced the correct output the following signal (75Hz, sampled at 1200Hz over a period of 0.21333s), was transformed from the time to the frequency domain:
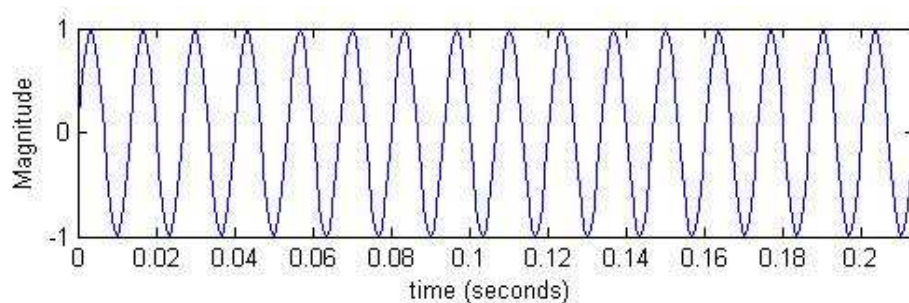
Figure 4.51 The signal used to test the program designed to transform the correlated data from the time to the frequency domain. The signal has a frequency of 75Hz and has been sampled at 1200Hz; the analysis interval is 0.21333s.

The output can be seen in figure 4.52:
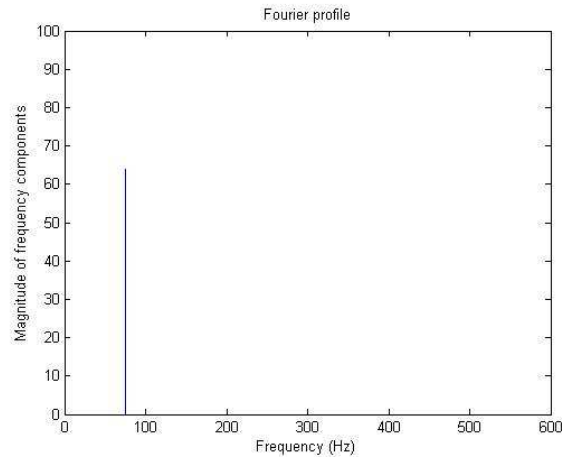
Figure 4.52 The output from the program designed to transform data from the time to the frequency domain. A clear peak can be seen at 75Hz therefore the program is functioning as expected.

A frequency component at 75Hz was recovered, thus the program was functioning as expected.

4.6.2 Differences between 'response' and 'no response' in the frequency domain

As a preliminary investigation two correlated mfERG waveforms were transformed to the Fourier domain: a normal response; and a waveform with no significant function. The Fourier profile of each was studied and differences were noted; this was to establish if this approach had the potential to be used. The following waveform, representing an area of normal retina, is shown in addition to its Fourier profile:
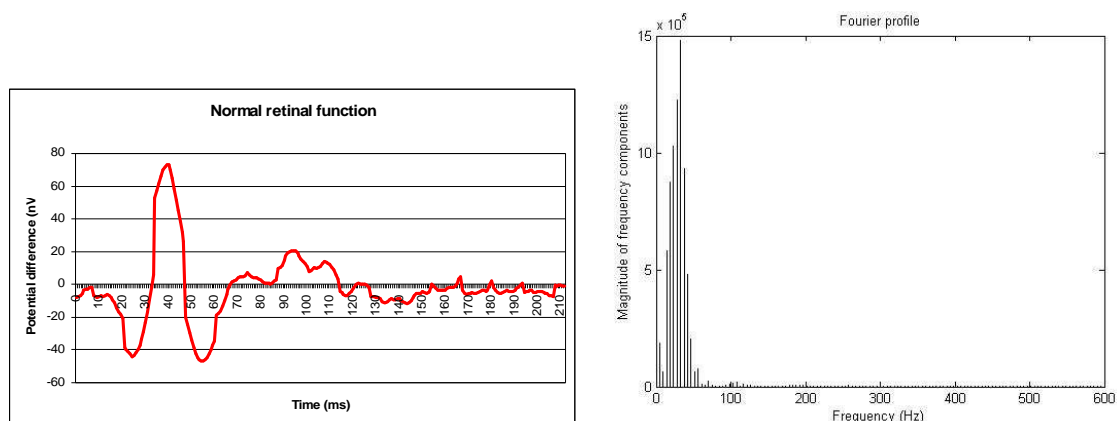
Figure 4.53 A normal mfERG response (left) and its corresponding Fourier profile (right). A normative-like frequency distribution is evident in the Fourier profile. The principal contributions are seen in the 0-100Hz range.

A normative-like frequency distribution can be seen. The discrete nature of the plot should not be confused with that observed when the uncorrelated data were visualised in the frequency domain; in this instance the peaks are simply a reflection of the frequency resolution achieved in the Fourier domain. It is evident that there was no

significant contribution from frequencies greater than 100Hz therefore all subsequent plots are shown from 0 to 100Hz. When a waveform containing no significant retinal response (left, figure 4.54) was visualised in the frequency domain, the profile shown on the right of figure 4.54 was obtained:
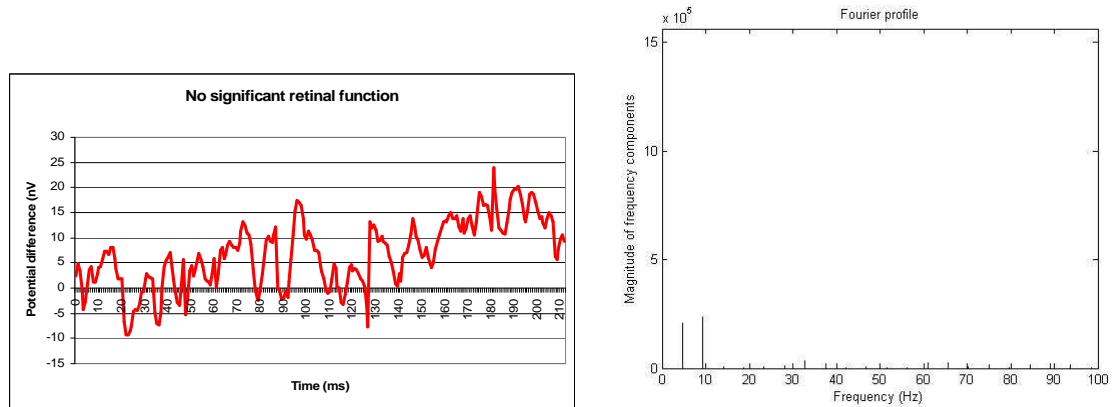
Figure 4.54 A mfERG waveform with no significant retinal function (left) and its corresponding Fourier profile (right). No significant frequency contributions were noted above 10Hz. This profile is distinct from that recovered from a normal response (figure 4.53).

No significant peaks were noted at frequencies higher than 10Hz. It was evident that the frequency profile of a normal response is distinct from one with no significant function.

4.6.3 Automating findings: correlated data

In light of the differences seen in section 4.6.2 it was decided to investigate the possibility of using the frequency domain to develop an automated, objective method for stating if a waveform should be analysed.

4.6.3.1 Methods

1500 mfERG waveforms were taken from mfERG recordings on 200 patients, each of whom was tested using the mfERG protocol (section 3.4.1). Waveforms were selected from different locations in the trace arrays and ranged from a clear retinal response to no significant function. Each waveform was analysed by a human expert and was classified as 'response' or 'no response'. A waveform was defined as a retinal response when there was an intact P1 component. This category therefore included waveforms which were significantly decreased and delayed. All 1500 waves were transformed into the frequency domain and were utilised to define the limits for 'response' and 'no response'.

The Fourier profile for each waveform was compared with that of an ideal response, which was defined as the global response taken from a healthy and compliant individual. The magnitude of all frequency components was normalised. The frequency profiles of the two waves used in section 4.6.2 to demonstrate differences between a clear response and no retinal function in the Fourier domain can be seen below. In each case the normalised ideal response is shown in green while the normalised frequency profile of the wave is shown in blue:
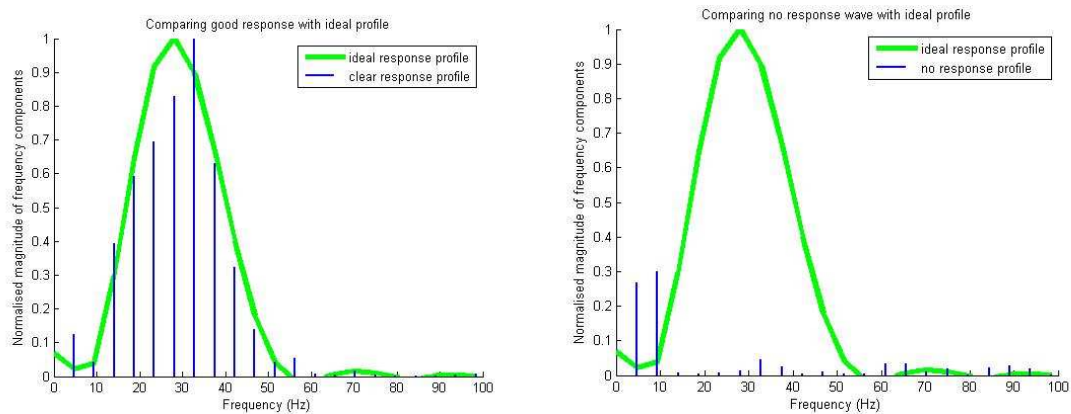
Figure 4.55 A comparison of the Fourier profile from a normal response (left) and 'no response' (right) with that of an ideal response. The frequency profile of the normal mfERG waveform is comparable with that of the ideal response. The Fourier profile of the waveform with no significant function differs considerably to that of the ideal response.

The frequency profile of each waveform was compared with that of the ideal response by calculating the cumulative difference in the y-coordinate between the two profiles at each data point. A single value was therefore obtained for each of the 1500 waveforms. As the waveforms had been categorised as 'response' or 'no response' the 5th-95th percentile was calculated for the group of waveforms said to represent an area of function. The same was done for the group of waves with no significant function. There would ideally be a separation of these ranges.

4.6.3.2 Testing the system

It was important to test the system to assess its potential for classifying the mfERG waveforms. An additional 1000 mfERG waveforms were therefore chosen, taken from 100 different patient trace arrays. Again the quality of these recordings varied, the location of the waveforms in the trace array differed and they represented a wide range of function. Each of the 1000 waveforms was classified as having a response or representing an area of no function by an expert. In cases where this distinction was difficult, comment was made. Each waveform was subsequently transformed from the time to the frequency domain and normalised. The frequency profile of each response

was compared with that of the ideal normalised Fourier profile by summing the difference in the y-coordinate between the two profiles at each data point. Based on the limits defined in section 4.6.3.1 the waveform was classified; this was compared with that of the expert in each case.

4.6.3.3 Results

For the 1500 waveforms used to define the limits for 'response' and 'no response' a 5th-95th percentile for the deviation from the ideal Fourier profile was calculated for each classification. This range was 2.29-6.30 and 5.99-6.92 for the 'response' and the 'no response' groups respectively. A small region of overlap therefore existed. This can be seen in figure 4.56:
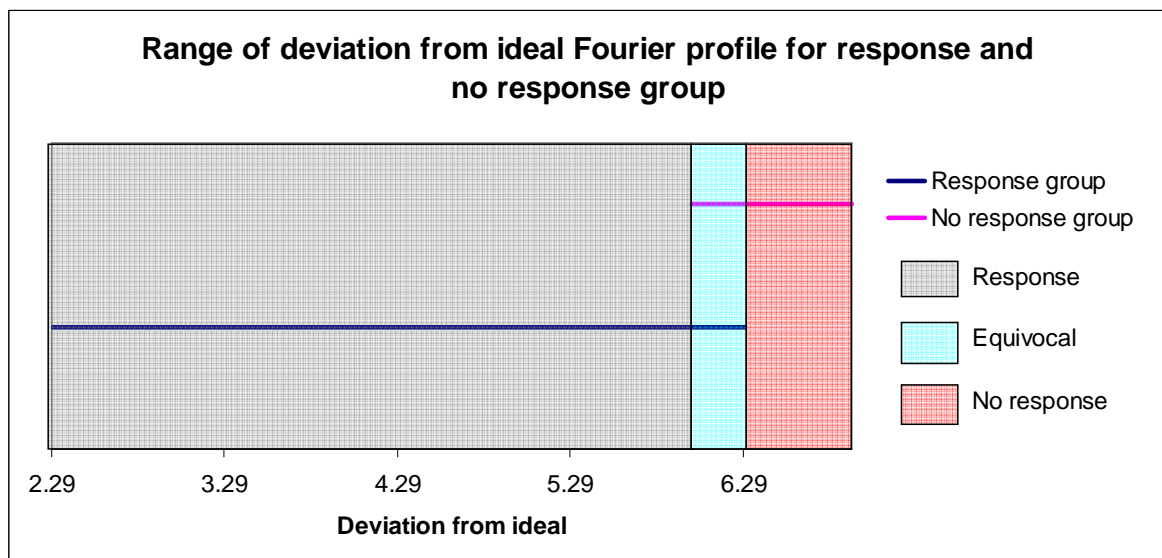
Figure 4.56 Comparing the deviation from the ideal Fourier profile for all waveforms said to be 'response' with that of those classed as 'no response'. Those waveforms with a deviation value of less than 5.99 (shown in grey) can be classified as 'response' while those with a deviation value greater than 6.30 (orange) can be categorised as 'no response'. A region of overlap exists (5.99 to 6.30), shown in blue, therefore these waveforms are classed as 'equivocal'.

Using these limits it could be said that a response with a deviation from the ideal profile of less than 5.99 has a response, those with a value greater than 6.30 have no response and those between these values are equivocal. These limits were subsequently utilised to classify the additional 1000 waveforms; the resulting classifications were then compared with those of the expert. The results can be seen in table 4.3:

		Computer's classification		
		Response	No response	Equivocal
Expert's classification	Response	**426**	82	151
	No response	17	**215**	109

Table 4.3 A comparison of the expert's and the system's classification when categorising the cross correlated waveforms as 'response' or 'no response'. There was a 64% agreement between them, with 26% said to be equivocal; the expert and the system disagreed for 10% of the 1000 waveforms.

Of the 1000 waveforms, 64% were classified correctly by the system while the expert and the system disagreed for 10% of the waveforms. 26% were said to be equivocal by the system. Of those waveforms classified incorrectly by the computer (not including the responses said to be equivocal) 50% were reported as being difficult to categorise by the expert.

4.6.4 Discussion

When the deviation from ideal was calculated for the 'response' and the 'no response' groups a region of overlap was found; this comprised 7% of the total range of values. Ideally there would have been no such region as it required the creation of a third classification, 'equivocal', thus limiting the potential use of this technique. The following examples demonstrate why a clear separation did not exist between the two groups. The first waveform was classified as having a response by the expert while the second was said to have no response. Each was however categorised as equivocal by the computer. The normalised Fourier profile, in addition to that of an ideal response, is shown for each waveform:
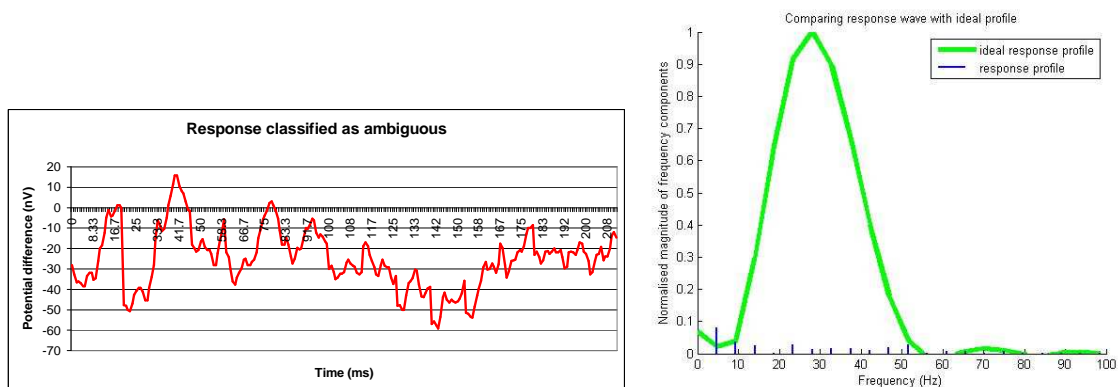
Figure 4.57 A mfERG waveform classified as 'response' by the expert (left) and its corresponding Fourier profile (right). This example was categorised as equivocal by the system. Small frequency contributions are evident across the frequency spectrum in the Fourier profile.
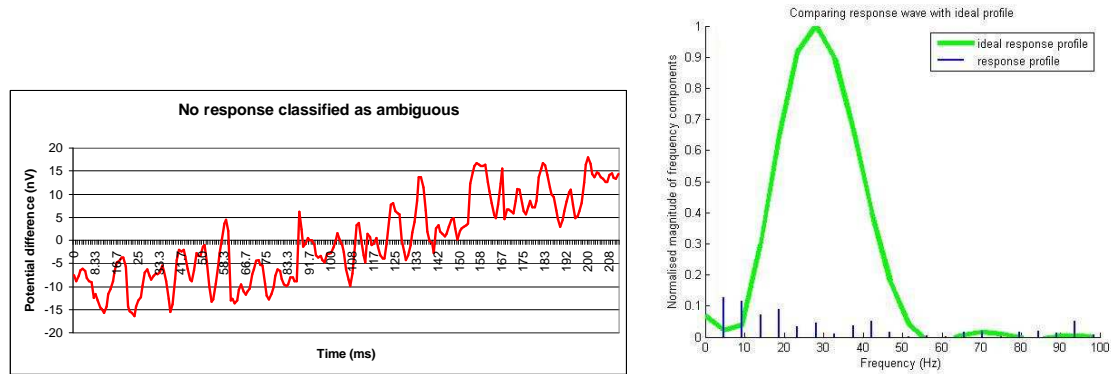
Figure 4.58 An example of a waveform classified as 'no response' by the expert (left) and its Fourier profile (right). This was categorised as equivocal by the system. Small frequency contributions are evident across the frequency spectrum in the Fourier profile. These are relatively similar to those seen in figure 4.57.

It can be seen that there is very little to distinguish between the two Fourier profiles, thus explaining why there was a region where it was not possible to differentiate one from the other. Although discrepancies were observed between the expert and the system, this method has the potential to be utilised to state if a waveform should be analysed or represents an area of no physiological response as only 10% of the waveforms were categorised into the wrong group.

4.6.5 Conclusions

The expert and the system agreed for 64% of the waveforms analysed, which is relatively low; the system's classification was however only completely incorrect for 10% of the testing set, with the remainder said to be equivocal. It is apparent that analysis of the correlated mfERG waveforms in the frequency domain cannot be used in isolation to assess if a waveform contains a physiological response however it could potentially be utilised as part of a multilayered system; its classification could be used in conjunction with a number of additional approaches, with the importance of each of being weighted.

4.7 Chapter summary and conclusions

The two main aims of this chapter were to develop a method for grading the calibre of a recording, both during and after testing, and to assess if the Fourier domain can be used to differentiate a waveform with a physiological response from one with no retinal function.

It was shown that by viewing the uncorrelated data in the frequency domain good recordings were highly distinct from poor tests, both when viewing the entire data set and when studying a small section of it. A method for grading the recording quality into one of four categories was developed for the complete recording to emulate grading upon completion of the test. When tested, a high agreement was seen between the experts and the computer when defining if a recording should be reported. Differences were however observed when sub-classifying the recording quality and should be the subject of further investigation. A technique was also developed for categorising the recording quality of one data segment into one of two groups (acceptable or unreportable) to simulate grading the integrity in a 'live' manner. Again a relatively high agreement was observed between the system and the expert. This system therefore has the potential to grade the integrity of the recording quality, both during and after a test, thus helping the operators and those analysing the results.

Cross correlated waveforms were analysed in the frequency domain with a view to distinguishing between waveforms with a 'response' and those with 'no response'. Differences were observed in the Fourier domain between normal physiological responses and those with no function however in the case of compromised retinal function this distinction was less clear. It is apparent that this method could not be used in isolation to state if a waveform should be categorised as 'response' or 'no response', but it could have a role to play as part of a multilayered system.

When investigating the Fourier profile of the raw uncorrelated data the principal mfERG stimulus frequencies were established. By conducting a series of experiments it was found that these were determined by the fundamental frequencies of the orthogonal sequences utilised, and the superposition of these frequencies when using a multi-stimulus element. By viewing this profile the ability of the retina to respond to the different stimulation frequencies could be visualised with ease. The selective abolition of different frequencies in the case of compromised retinal function was noted. Temporal information can therefore be extracted in a highly simple manner.

As it has been shown that the Fourier domain cannot be used as the sole technique to distinguish a physiological response from no significant function, a number of additional methods have been investigated to make this distinction. These include artificial neural networks (ANNs) and the use of the signal to noise ratio. The potential role of ANNs to classify the mfERG waveforms is investigated in the following chapter.

5 The artificial neural network

The previous chapter studied the potential of the Fourier (frequency) domain to distinguish a physiological response from a waveform with no significant retinal function. When the classifications offered by this approach were compared with those of a human expert an agreement of 64% was realised, which is relatively low. Only 10% of the waveforms were however misclassified, with the remainder said to be equivocal based upon their Fourier profiles. Although this technique showed potential, it could not be used in isolation to analyse the mfERG data. This chapter therefore investigates an alternative method: artificial neural networks (ANNs). The aim was to train a network to categorise the mfERG responses into one of three classes: responses within normal timing limits; responses which are delayed; or no significant retinal function. These classifications were based on the presence of a P1 peak, and when evident, its latency.

Real data, in this case clinical mfERG waveforms, are typically used to train networks; however acquiring sufficient amounts of data can be problematic. Furthermore the analysis process can be very time consuming. One approach used to minimise these issues is to form a training set from synthetically generated data, created by performing a series of manipulations on a small number of examples; only the original data requires analysis. Two data sets were created to train the networks: one comprised solely clinical mfERG waveforms while the other was made from artificially generated data. The ability of each data set to teach an ANN was compared.

5.1 Introduction to neural networks

ANNs are an attempt to emulate biological neural networks. They comprise many highly interconnected processing elements, analogous to neurons in the biological system, working in parallel to solve a particular problem. Representations of a biological and an artificial neuron are shown in figure 5.1:
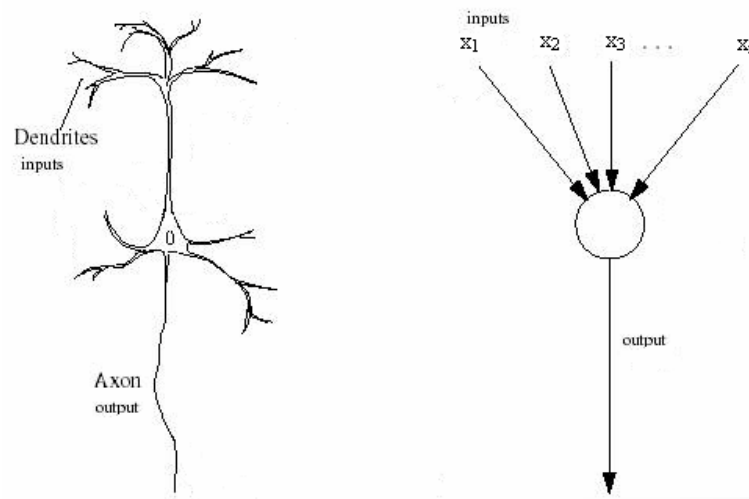
Figure 5.1 Comparison of a biological (left) and an artificial neuron (right) (adapted from www.hemming.se).

It can be seen that the artificial neuron's inputs and output correspond to the biological neuron's dendrites and axon, respectively. The axon of the biological neuron splits into thousands of branches, synapsing with the dendrites of other neurons; learning is achieved by adjusting these synaptic connections. Similarly, the weighting value between each of the artificial neurons is changed during the training process to enable an ANN to learn patterns. The ANN's output, when presented with an input, is dependent on the weights between each of the elements (artificial neurons). The method by which these weights are updated depends on the learning paradigm, the type of network utilised and parameters such as the learning rule.

5.2 Learning paradigms

Two of the main learning paradigms are supervised and unsupervised learning. When using the former paradigm the network is provided with a set of training examples and the target answer for each input. In contrast only the inputs are presented to the ANN for unsupervised learning; underlying patterns and trends in the data are searched for in order to categorise it. Supervised learning was utilised throughout this chapter as the target classification was known for the mfERG waveforms therefore all subsequent discussions refer to this paradigm.

5.3 The learning process

For supervised learning each example has a corresponding answer. The aim is to design a network which yields the same output as this target for all training examples. At the start of training the weighting value between each element is randomised

therefore the outputs generated by the network are likely to differ from the desired answers. This discrepancy, termed the error, is used to modify the weights until the error is minimised for the entire training set. This process is illustrated in figure 5.2:
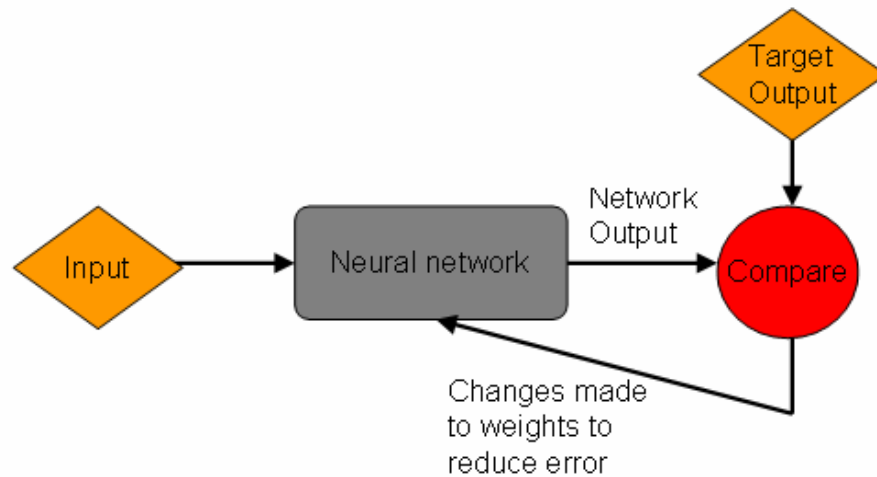
Figure 5.2 An overview of the training process used during supervised learning (modified from Mathworks Neural Network Tool Box). The objective is for the network's output and the target output to be the same for the complete training set, however a difference (error) normally exists. This difference is used to change the weights during training, to reduce the error.

A network may successfully learn the training data however it is essential that it performs well when presented with previously unseen data; this is defined as its ability to generalise. The network is therefore tested with a data set not used during the learning process. The ANN's output is compared with that of the desired answer in each case, enabling the network's performance to be assessed. Good generalisation is dependent on the network parameters and the data presented to the network during training. If these are not selected properly the network can overtrain in which case it acts more as a memory, thus reducing its ability to analyse new data.

5.4 Selecting network parameters

When designing an ANN the type of network, the structure of the network and the learning rule must all be selected. Many types of network exist including Hopfield, Radial Basis Function, self organising maps and feed-forward. The aim of the work presented in this chapter was to categorise the mfERG waveforms. Multilayer feed-forward ANNs were therefore chosen as these have been applied successfully to classification and pattern recognition problems similar to this (102;128). This type of network typically comprises three layers: an input; a hidden; and an output layer:
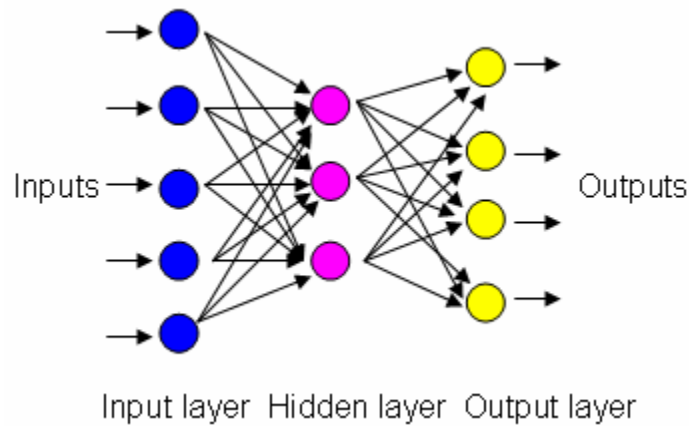
Figure 5.3 Structure of a multilayer feed-forward network. The input, the hidden and the output layer are all shown.

It can be seen that there are no connections between elements in the same layer or to elements in the preceding layer, i.e. they are unidirectional. For the majority of problems one hidden layer is sufficient to train a network effectively (129), although two hidden layers can be used. This can however compromise the ability of the network to generalise. One layer ANNs were therefore chosen throughout this study.

The type of error correction algorithm (the means by which the error is decreased), the method of presenting the data to the network during training and the number of times the training data are presented to the network can be chosen, as can the number of elements in the hidden layer and the transfer function utilised. Each of these parameters affects the ability of the network to perform the task required of it.

5.4.1 Error correction learning algorithms

The aim of learning is to decrease the error; the error surface is utilised to achieve this. The error surface is a graphical representation of the total error (i.e. the error for each training pattern) as a function of the network weights. As the weights are updated during training, the error surface should eventually descend to a minimum; the lowest point on the surface corresponds to the optimal network solution (i.e. the minimum error). The objective during training is therefore to find the combination of weights which locate the minimum error for the entire training set. Different methods are utilised to achieve this. It would be extremely time consuming and hence inefficient to try every possible combination of weights to obtain this minimum therefore the gradient descent method was developed to provide a quicker approach. The most common error correction algorithm using the gradient descent method is back propagation. This assesses the gradient of the error surface during the learning process;

the slope of the error surface indicates the sensitivity of the error to a change in the network's weights. The network weights are therefore updated in the direction where the error is decreasing most rapidly until the optimal solution is found. This can however be very slow to find an optimal solution. Furthermore the network solution can stop at local minima, occurring because the algorithm always updates the weights to decrease the error; in many instances the error must however rise before it will decrease again. If the local minimum is very different from the global minima the performance of the network will be very poor. A number of methods have been developed in an attempt to overcome these limitations, one of which is the inclusion of momentum when updating the network weights. This encourages movement in a particular direction; if multiple steps are taken in the same direction the speed of the algorithm increases, reducing the risk of a network becoming trapped in a local minimum, and decreasing the training time (130).

Many different learning rules exist, three of which will be described as they, in addition to back propagation with momentum, have been applied to the analysis of the mfERG in this chapter. The first of these is the quickprop algorithm, proposed by Fahlman. This is a rule based on back propagation which utilises the second order derivate of the error surface, corresponding to its curvature, in addition to the first order derivative of the error surface (i.e. the gradient) to locate the minimum. This learning rule has been shown to find a solution more quickly than back propagation however it can sometimes fail to converge (131). The second algorithm utilised was the delta-bar-delta rule, an alternative modification of the back propagation algorithm. Unlike the original rule for which each weight has the same learning rate, this changes each weight's learning rate during training. The learning rate is increased if the error decreases in the same direction for several steps whereas it is decreased if the direction changes. This offers greater flexibility and improved speed (131). Finally, the conjugate gradient descent is a second order algorithm which utilises only the second derivative of the error surface to determine the direction in which the weights are updated. This algorithm constructs a series of line searches across the error surface and locates the minimum along each line. There is an underlying assumption that the error surface is quadratic, which is not always the case. However, when close to the minimum the quadratic assumption is more appropriate, allowing the algorithm to locate the minimum very quickly (131).

5.4.2 Method of presenting the input data to the ANN

The input data can be presented to the network in one of two ways: incremental; or batch. When the incremental mode is chosen the network weights are updated after each training example is presented whereas the weights are only changed after the presentation of the entire training set when using the batch method (131).

5.4.3 Number of training iterations

During the learning process the training data are presented to the network many times. If the number of iterations (number of presentations) is insufficient the ANN may be unable to detect patterns in the data, however if too large the network can become over trained resulting in poor generalisation.

5.4.4 Number of elements in each layer of the network

The number of elements in the input and the output layers is determined by the nature of the problem being solved. The number of elements used in the hidden layer is however chosen by the designer and is crucial to the final performance of the network. If too few neurons are used, the network may not be able to solve the problem whereas if too many neurons are selected the ability of the network to form generalisations may be diminished. To establish the optimal number of hidden elements a process of trial and error is required.

5.4.5 Transfer function used by the elements

Each element has the following structure:

$$\text{Output} = \text{function } (w_{1,1}{}^{*}x_1 + w_{1,2}{}^{*}x_2 + w_{1,3}{}^{*}x_3)$$
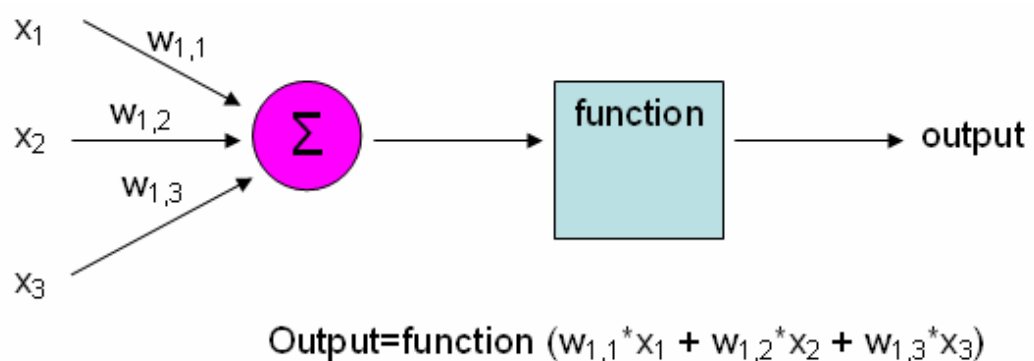
Figure 5.4 The structure of a processing element, including the function which determines its behaviour.

The function used by the element governs its behaviour. Three general categories of transfer function exist: threshold; linear; and sigmoid. For threshold elements the output is restricted to 0 or 1. The output is dependent on the sum of the weights and inputs; if less than the threshold value the output is 0, if greater the output is 1. The output from a linear element is proportional to the summation of the weighted inputs while the output from a sigmoid processing element varies continuously, but not linearly as the input changes. Sigmoid elements were utilised throughout this study as they enable non-linear problems to be solved (130) and can decrease training times (131).

It is evident that there are a number of network parameters which can be varied when training an ANN; a process of trial-and-error is required to establish the optimal network.

5.5 Training and testing data sets

Prior to training an ANN it is necessary to create two data sets: one to teach the network; and another to assess its performance once training is complete. The training data should be representative of the problem being solved, exposing the network to all possible examples during the learning process. The number of examples used must also be considered as this can impact on the network's ability to solve the problem; if too few examples are utilised, the network can overfit to the data. The number of training examples should be greater than the number of weights in the network to avoid overtraining the ANN. The recommended ratio of training examples to weights varies between researchers with values ranging from 30 (132) to 5-10 (133). It has also been reported that a ratio of 2 can be utilised (134). An insufficient number of training examples is often available thus a number of techniques have been developed with a view to preventing overtraining, one of which is early stopping. When using this approach the data available for training the network are divided into two data sets: a training set; and a validation set. The former is utilised to update the weights during the learning process while the error on the latter data set is monitored during training. The error on both the training and the validation sets typically decreases at the start of training however if the network starts to overfit to the data the error on the training set continues to decrease while that on the validation set increases. Training is stopped at this point. There is no consensus regarding the optimal ratio of the validation set size to the training set size, with ratios ranging from 1:3 (135) to 1:10 (132). A ratio of 1:5

was selected throughout this thesis as it was a compromise between these. Wang *et al.* stated that early stopping offers no significant advantage when the ratio of training examples to network weights exceeds 20 (132) therefore this technique was only utilised in this chapter when the ratio was less than 20.

A testing set is also required to assess the generalisation of the network once training is complete. This previously unseen data set is presented to the network, which provides an output; this is compared with that defined by an expert, allowing the performance of the network to be established.

As stated earlier the aim of the study was to investigate the potential application of ANNs for classifying mfERG waveforms as being within normal timing limits, delayed or as having no significant response. Ideally the network would also state if a response is decreased in amplitude or within normal limits. It was however shown in section 3.4.2, when establishing the normal range for the mfERG, that the amplitude of responses decreases with eccentricity. Five different neural networks (one for the central response and one for each of the four concentric rings) would therefore have been required, which was impractical at this stage. It was however possible to employ a single network when classifying responses as delayed or within normal limits, as the 95th percentile for the P1 latency was 42ms for all four concentric rings and the central response (refer to section 3.4.2).

For classification problems similar to this, real data are typically used to train the ANN as they contain the diversity seen in practice (104;107;136;137). A series of networks was therefore trained and tested using clinical data. Working with real responses does however have a number of limitations associated with it including the laborious nature of classifying the data, the difficulty of obtaining large amounts of clinical data and the lack of knowledge of the true underlying signal prior to the addition of noise artefacts. An alternative method, utilised by Fisher *et al.* is to form the data set from synthetic data; this data set is generated by performing a series of manipulations on the original data set (108). This offers the possibility of creating very large data sets thus achieving the desired training set size. Furthermore, only the primary waveforms require classification by the expert; each of the secondary waveforms are categorised automatically by the program used to generate them as the processes used to create them are embedded in the program. Consequently the analysis time required by the expert is reduced. Additionally the interpretation process is consistent; this is not always the case with humans. The possibility of training

ANNs with synthetic mfERG data was therefore investigated; the performance of networks trained using synthetic and clinical data was compared. Those trained with artificial data are described initially.

5.6 Methods: training a network with synthetic data

The synthetic data set was created from 50 mfERG waveforms. These were selected from 50 different patient recordings, all of which were acquired using the mfERG protocol (section 3.4.1). Responses were chosen from a variety of locations in the trace array and represented a wide range of retinal function. 10 of the waveforms had no significant response while 40 represented an area of functioning retina. They were selected from recordings obtained from compliant patients therefore the signals were relatively undistorted by noise thus the location of P1 was easily identifiable for those 40 waveforms with a physiological response. Each response was analysed by an expert; waveforms with no significant function were classified while the latency of the P1 component was noted for all other responses. In chapter 3 normative data was presented, showing that responses with a P1 latency of 42ms or less were within normal limits (section 3.4.2). Each of the responses said to represent an area of functioning retina were therefore classified as delayed or not delayed based upon this limit. Target values of -0.9, 0 and 0.9 were utilised for the groups 'no significant response', 'delayed' and 'not delayed' respectively. 0.9 and -0.9 were chosen instead of 1 and -1 as the sigmoid transfer function used by the elements cannot attain values of 1 or -1 (130).

The aim was to generate a data set from these 50 responses which encompassed the amplitude and latency range seen in practice, in addition to some of the artefacts found clinically such as baseline drift and 50Hz electrical noise. Random noise was also added in an attempt to emulate patient noise. Each mfERG response comprises 256 data points therefore a network would require 256 processing elements in its input layer if presented with the full waveform. The main points of interest are contained within the initial 100ms therefore only the first 120 data points of each base wave (corresponding to approximately the initial 100ms of the response) were used to generate the data set. Fewer processing elements were therefore required in the input layer of the ANN and hence the complexity of the network was reduced. Matlab 2007a was used to create the artificial waves, all of which were written to Excel as they were created.

5.6.1 Creation of the synthetic data set

A number of different manipulations were performed on each of the 50 base waves to generate the data set, each of which are described in the following section.

5.6.1.1 Amplitude and latency range

It has been noted in the department that mfERG responses with a P1 latency range of 35-67ms are typically seen clinically when using the mfERG protocol described in chapter 3 (section 3.4.1) therefore the synthetic data set must comprise waveforms with P1 latencies within this range. To achieve this, each of the original waveforms was shifted, in increments of 2 data points, (corresponding to approximately 1.67ms) until reaching the upper and lower limits of this range.

The P1 amplitude of mfERG responses was observed to lie between 10nV and 150nV when using the mfERG protocol thus the original waveforms were multiplied by a scaling factor to encompass this amplitude range. 10 different scaling factors were utilised in each instance and were defined with knowledge of the original P1 amplitude.

The 10 waveforms defined as 'no response' by the expert could not be scaled and shifted based upon a P1 value. Each waveform was therefore shifted to the left and right in steps of 2 data points until a shift of 20 data points was achieved in either direction. They were also multiplied by 10 different scaling factors ranging from 0.5 to 2.0 (increments of 0.167). The shifting and scaling functions were performed simultaneously for each of the 50 waveforms, creating 10000 secondary waveforms. A detailed description of this process can be found in appendix 3.

5.6.1.2 Stretch

The shape of a waveform is unaffected by shifting it and changing its amplitude therefore the original waveforms were also stretched by varying amounts. The aim was to shift data points to the right, with those data points at the latter part of the waveform experiencing the greatest shift. To achieve this the first 30 data points were kept the same. The shift of the final data point (number 120) was varied from 5 to 25 data points, in increments of 5 while the shift of data points 31 to 120 varied linearly from 0 to the value of the maximum shift. The amplitude was also changed, again by

multiplying the waveforms by 10 different scaling factors, to achieve an amplitude range of approximately 10-150nV. As before, those waveforms with no physiological response were multiplied by 0.5 to 2.0 in steps of 0.167 (i.e. 10 different scaling factors). The scaling and stretching tasks were performed concurrently on the 50 primary waveforms, generating 2500 new waveforms. This process is discussed in more detail in appendix 4.

By performing the functions described in sections 5.6.1.1 and 5.6.1.2, 12500 waveforms were created.

5.6.1.3 Noise

It was also important to include waveforms affected by noise in the data set. 2000 waveforms were selected from the secondary waveforms generated in sections 5.6.1.1 and 5.6.1.2; noise artefacts were subsequently added to each of these. This group of 2000 waveforms comprised 1000 waveforms containing a response and 1000 waveforms with no significant function. The 1000 responses were selected from the 10000 secondary waveforms produced by stretching, scaling and shifting the 40 primary clinical responses. 20 shifted and scaled versions of each of the 40 primary waveforms were selected at random, in addition to 5 stretched and scaled versions of each of the 40 original responses (again selected at random); a group comprising 1000 responses was thus formed. The group of 1000 waveforms with no retinal function was formed by selecting, at random, 80 waveforms from the shifted and scaled version of each of the 10 original waveforms with no function. 20 examples were also chosen at random from the secondary waveforms created by stretching and scaling each of the 10 primary waveforms with no significant function. The process by which the examples were selected is explained in greater detail in appendix 5.

One of the noise artefacts added was baseline drift as this is a problem commonly encountered when testing patients. It is a low frequency artefact which arises from eye movement and blinking. The drift of the uncorrelated data from the baseline can result in the raw data moving out with the range of the electronics used for the signal acquisition, causing saturation of the signal. This results in a tilt of the final responses after the cross correlation process. To mimic this a line of the form $y = mx + c$, where m is the gradient and c is the y-intercept, was created. This was then added to the original waveforms. m was varied from -0.75 to 0.75 in steps of 0.5 while c was given

the values -20, 0 and 20. 12 different lines were thus created and added to the waveforms to mimic different severities of drift.

In addition to baseline drift, it is often the case that a 50Hz signal (electrical noise) is superimposed onto the final responses. This was therefore created in Matlab and added to the waveforms. Both the magnitude and phase of this signal were varied; the magnitude was defined as 10% or 20% of the P1 amplitude of the original waveform while phases of both 0^0 and 180^0 were utilised. The phase of the 50Hz sine wave was changed as the addition of a peak or a trough affects the final response differently. Patient noise can also degrade the quality of the cross correlated waveform therefore random noise was generated in Matlab in an attempt to emulate this. The maximum magnitude of this noise was defined as 10% or 20% of a waveform's P1 amplitude. Each of these three noise artefacts were added, resulting in the creation of 96 new responses for each waveform. Only drift was superimposed onto those waveforms with no significant physiological response as they already comprised solely of noise. 108000 new waveforms were thus generated. A detailed description of this process is included in appendix 5. In total, 120500 synthetically generated waveforms were therefore created (12500 in sections 5.6.1.1 and 5.6.1.2 and 108000 in section 5.6.1.3). An example of 6 responses produced from one of the original waveforms can be seen in figure 5.5:
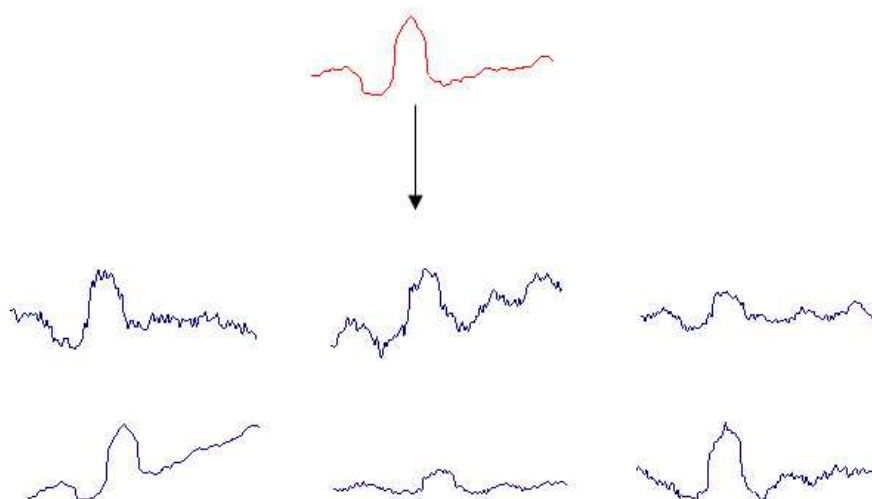
Figure 5.5 An example of six synthetically generated waveforms, shown in blue, along with the original clinical mfERG response, shown in red. These blue responses are shifted and stretched versions of the original waveform. 50Hz, random noise and drift have also been superimposed onto the responses.

Waveforms were classified into one of the three groups depending on the processes performed to generate them, thus each synthetic waveform had a target value of either -0.9 ('no significant response'), 0 ('delayed') or 0.9 ('not delayed') associated with it.

The method by which a target value was allocated to each waveform is detailed in appendices 3, 4, and 5.

5.6.2 Training and testing set

It was previously discussed in section 5.5 that the number of examples forming the training set must exceed the number of weights in a network. If an ANN with 10 elements in the hidden layer was chosen the network would comprise 1210 weights as there are 120 inputs and 1 output (120x10 + 10x1). Wang *et al.* stated that the number of examples in the training set should ideally be 30 times greater than this, which corresponds to 36300 examples. A training set was therefore formed by selecting 40000 waveforms at random from the synthetic data set. A further 1000 of the synthetic waveforms were randomly assigned to a testing set; in this instance a validation set was not required as a sufficient number of training examples were available. The method by which data were allocated to the training and the testing set is described in appendix 6.

It was also important to assess the performance of the ANNs on real clinical data therefore 1000 mfERG responses, representing a wide variety of retinal function and signal quality were taken from 100 patients. These were the same as those used to test the ability of the Fourier domain to distinguish between 'response' and 'no response'. Each clinical wave was classified into one of the three categories (no significant physiological response, a delayed response or a response within normal timing limits) by a human expert.

5.6.3 Training the ANN

As previously stated multilayer feed-forward networks with one hidden layer were chosen with all elements utilising the sigmoid transfer function. The learning rule and the method by which the data were presented during training were changed initially. Momentum, quickprop, conjugate gradient and delta learning rules were investigated, using both batch and incremental learning when appropriate. The number of elements in the hidden layer was then varied from 2 to 40 in increments of 2, and finally the momentum rate was changed from 0.3 to 0.9 in steps of 0.1. Training was stopped and the network was tested after every 50 epochs (i.e. 50 presentations of the data to the network); this was continued until 1000 epochs. In the first instance each ANN was trained using 10 elements in the hidden layer while the different learning methods and

learning rules were investigated. All inputs were normalised and the weights were randomised prior to starting each training session.

Upon completion of training, each network was tested on both the synthetic and the clinical testing set. The network's output was compared with that of the expert in each case, enabling its performance to be assessed. Its ability to accurately classify the waveforms into one of the three categories was studied initially.

It was also of interest to assess the network's ability to differentiate 'response' from 'no response' for the optimal networks. All waveforms categorised as being delayed or within normal latency limits (by the expert, and then by the ANN) were classified as 'response', while those said to be 'no response' were categorised so. The agreement between the expert and the network when distinguishing between these two categories was assessed. The sensitivity and specificity was also calculated using the following equations:

$$Sensitivity = \frac{truepositives}{truepositives + falsenegatives} \times 100\% \qquad (5.1)$$

$$Specificity = \frac{truenegatives}{truenegatives + falsepositives} \times 100\% \qquad (5.2),$$

where true positive = abnormal identified as abnormal;

false positive = normal identified as abnormal;

true negative = normal identified as normal;

false negative = abnormal identified as normal.

The accuracy, sensitivity and specificity of the network to differentiate delayed responses from those within normal limits was also calculated; all responses categorised as 'no response', either by the expert or the ANN were therefore ignored in this instance.

5.6.4 Training a network with a smaller data set

It was originally stated that ANNs would be trained using both synthetic and clinical data; however to directly compare their performance it was important to train them

with the same number of examples. It was impractical to individually analyse 40000 clinical waveforms. A smaller synthetic data set was therefore created to train the network; the aim was to establish if similar results could be achieved when training the ANN with fewer examples, and hence enabling a fair comparison of a synthetically trained network with a clinically trained ANN.

Masters reported that it was possible to train a network with a data set comprising twice as many examples as weights in the network (134). If for example there were 10 elements in the hidden layer the training set would require 2420 waveforms; a data set consisting of 2500 synthetic waveforms was therefore formed. These were selected at random from the larger synthetic training set; this was to ensure that the training and testing data were independent from one another.

In section 5.5 it was discussed that a validation data set is required to prevent overtraining of a network when the ratio of training examples to network weights is less than 20. It was therefore necessary to form a validation set in this instance. It was also stated that the ratio of examples in the validation set to those in the training set would be 1:5 throughout this chapter. 500 synthetic waveforms (20% of 2500) were therefore selected at random from the large synthetic training data set to form the validation data set. The same synthetic and clinical testing sets described in section 5.6.2 were utilised. The optimal network parameters established when teaching ANNs with the larger data set were selected originally to assess the possibility of teaching a network with a smaller data set.

5.7 Results: training a network with synthetic data

5.7.1 Large synthetic training set

The following table demonstrates the performance of each network when trained using the four learning rules. Both batch and incremental learning have been utilised, with the exception of the network being taught using the conjugate gradient learning rule, which can only learn using the batch mode of learning. In the first instance the hidden layer comprised ten elements. The number of epochs for which the optimal performance was achieved when tested with the clinical data is shown in each case:

Learning Rule	Type of training	Hidden elements	Number of epochs	% agreement with expert: **clinical set**	% agreement with expert: **synthetic set**
Momentum	Batch	10	750	62	96
Momentum	**Incremental**	**10**	**300**	**68**	**98**
Quickprop	Batch	10	900	59	91
Quickprop	Incremental	10	100	63	97
Conjugate gradient	Batch	10	450	53	92
Delta	Batch	10	300	64	97
Delta	Incremental	10	150	58	94

Table 5.1 Agreement between the expert and the ANN trained with synthetic data when tested on synthetic and clinical data: varying the learning rule and the mode of learning. Results are displayed for four different learning rules, in addition to incremental and batch learning.

It can be seen that the highest performance was yielded when using the momentum rule and the incremental learning mode; agreements of 68% and 98% were achieved for the clinical and synthetic data sets respectively. In each instance the results were significantly better when tested on the synthetic data. It should be noted that the ability of each network to classify the synthetic data set improved with an increasing number of epochs. This was not however the case when tested on the clinical data set; the performance initially increased to a maximum and either fell or stayed constant as the number of epochs was increased further.

Multi layer feed-forward ANNs, trained with the momentum learning rule using a momentum rate of 0.7 and one hidden layer were subsequently trained while the numbers of elements in the hidden layer was varied. A momentum rate of 0.7 was utilised as this was the default when using the momentum learning rule. The results achieved are shown in table 5.2:

Hidden elements	Number of epochs	% agreement with expert: **clinical set**	% agreement with expert: **synthetic set**
2	50	60	93
4	100	62	93
6	50	65	94
8	200	68	98
10	300	68	98
12	**300**	**69**	**99**
14	250	67	98
16	300	68	99
18	250	68	98
20	200	67	98
22	150	65	97
24	150	66	98
26	200	64	96
28	150	65	97
30	200	66	98
32	150	65	97
34	150	66	98
36	150	63	95
38	100	64	96
40	50	66	97

Table 5.2 Agreement between the expert and the ANN trained with synthetic data when tested on synthetic and clinical data: changing the number of elements in the hidden layer. The number of elements was varied from 2 to 40 in increments of 2, while the learning rule and learning mode (momentum and incremental respectively) remained constant.

The performance of the network was optimal when there were twelve elements in the hidden layer; the agreement between itself and the expert when presented with the clinical and the synthetic data sets was 69% and 99% respectively. The momentum rate was subsequently varied. Multi layer feed-forward ANNs, trained with the momentum learning rule and one hidden layer comprising twelve elements were utilised:

Momentum	% agreement with expert: **clinical set**	% agreement with expert: **synthetic set**
0.3	68	98
0.4	68	98
0.5	69	99
0.6	**70**	**99**
0.8	67	96
0.9	66	93

Table 5.3 Agreement between the expert and the ANN trained with synthetic data when tested on synthetic and clinical data: changing the momentum. The momentum value was varied from 0.3 to 0.9 in steps of 0.1, to optimise the performance, while the learning rule, the learning mode and the number of elements in the hidden layer remained constant.

It can be seen that the network correctly classified 70% of the clinical mfERG waveforms and 99% of the synthetically generated data. Thus the optimal network was a multi layer feed-forward network trained with the momentum learning rule using incremental learning with a momentum of 0.6 and twelve sigmoid neurons in its hidden layer.

When making the distinction between 'response' (i.e. all 'delayed' and 'not delayed' waveforms were grouped together) and 'no response', this network correctly classified 86% of the clinical data testing set, with sensitivity and specificity values of 80% and 87% respectively. The lower sensitivity indicates that misclassifying waveforms with no significant retinal function as 'response' was a greater problem than the reverse situation. When distinguishing delayed responses from those within normal timing limits (i.e. all waveforms said to have no physiological response were ignored) an accuracy of 81% was realised. The sensitivity and specificity values were 84% and 78% respectively therefore stating that a waveform was delayed when it had a P1 latency of 42ms or less was more problematic than the reverse situation. Of those responses incorrectly classified, 50% had a P1 latency, as stated by the expert, within 1ms of the timing boundary between normal and abnormal (i.e. 42–43ms). Furthermore, 81% had a P1 latency within 2ms of this threshold (i.e. 41–44ms), thus the majority of misclassifications occurred when close to this timing boundary.

5.7.2 Smaller synthetic training set

This section presents the results achieved when teaching the ANN with fewer examples. When an ANN was trained using the same parameters as those used by the optimal network in section 5.7.1, the network and the expert provided the same classification for 99% of the synthetic testing set. This is the same as that achieved when the larger training set was utilised. When tested on the clinical data an agreement of 69% was realised, which is comparable with that of the optimal network in section 5.7.1. No additional networks were trained as the results were similar to those achieved when using large volumes of training data.

5.8 Methods: training a network with clinical data

Having established that it was possible to train the ANN with 2500 waveforms when using a validation set, networks were trained using real mfERG waveforms.

5.8.1 Training, testing and validation set

2500 clinical mfERG waveforms were selected and classified into one of the three categories ('no significant response', 'delayed' or 'not delayed') by an expert. These were taken from 200 different patient recordings and represented a wide range of retinal function and quality. 200 recordings were utilised as opposed to selecting each of the 61 waveforms from 41 trace arrays; this was to increase the variety in the basic shape of the mfERG waveforms, as responses in a trace array can have a similar underlying shape to one another.

As when training the network using the smaller synthetic data set, a validation set was required in this instance to prevent overtraining of the network, as the ratio of training examples to weights was less than 20 (refer to section 5.5). Again the ratio of the validation set to the training set was 1:5 (see section 5.5); a further 500 (20% of 2500) mfERG waveforms were therefore selected. These were subsequently classified into one of the three groups by the expert to form the validation set. The testing set utilised to test the synthetically trained ANNs, comprising 1000 clinical waveforms, was used to assess the performance of the networks. Again only the first 120 data points were selected from each waveform.

5.8.2 Training the ANN

As in section 5.6.3 feed-forward multilayered networks with one hidden layer consisting of sigmoid elements were used. The learning rule, the method of presenting the data during training, the number of elements in the hidden layer, the duration of training and the momentum were all varied to achieve the optimal performance. All inputs were normalised and weights were randomised prior to training. Training was stopped and the network was tested after every 50 epochs; this was continued until 1000 epochs or the point at which the ANN stopped training due to overfitting, as determined by the error on the validation set. The overall classification accuracy was calculated, in addition to the sensitivity and specificity for the distinction between 'response' and 'no response' and 'delayed' or 'not delayed' using equations 5.1 and 5.2.

5.9 Results: training a network with clinical data

Table 5.4 demonstrates the agreement between the ANN and the expert when the different learning rules and methods of presenting the training data were investigated:

Learning Rule	Type of training	Hidden elements	Number of epochs	% agreement with expert
Momentum	Batch	10	850	72
Momentum	**Incremental**	**10**	**150**	**77**
Quickprop	Batch	10	200	63
Quickprop	Incremental	10	150	75
Conjugate gradient	Batch	10	50	60
Delta	Batch	10	100	65
Delta	Incremental	10	550	60

Table 5.4 Agreement between the expert and the ANN trained with clinical data when tested on clinical data: varying the learning rule and the mode of learning. Results are displayed for four different learning rules, in addition to incremental and batch learning.

As observed in section 5.7.1 the highest agreement was achieved when the momentum rule and the incremental learning mode were selected; 77% of waveforms were correctly classified into one of the three categories. When the number of elements in the hidden layer was varied no improvement upon this result was found. Finally the

momentum was changed as a default value of 0.7 had been utilised when teaching with the momentum learning rule; the results can be seen in table 5.5:

Momentum	% agreement with expert
0.3	70
0.4	75
0.5	**78**
0.6	76
0.8	76
0.9	73

Table 5.5 Agreement between the expert and the ANN trained with clinical data when tested on clinical data: changing the momentum. The momentum value was varied from 0.3 to 0.9 in steps of 0.1, to optimise the performance, while the learning rule, the learning mode and the number of elements in the hidden layer remained constant.

The ability of the network to classify the waveforms improved slightly from 77% to 78% by decreasing the momentum to 0.5. The optimal network found for the particular problem was therefore a multi layer feed-forward ANN with ten sigmoid elements in its hidden layer, taught with the momentum rule using incremental learning. An important decision to be made by the ANN is whether or not a waveform represents an area of functioning retina. In this instance the ANN agreed with the expert for 90% of the cases, with sensitivity and specificity values of 83% and 92% respectively. When differentiating delayed responses from those within normal timing limits, an agreement of 86% was realised. The sensitivity was 89%, while the specificity was 83%. It was observed that for those responses misclassified as 'delayed' or 'not delayed', 53% were within 1ms of the timing boundary between normal and abnormal while 85% were within 2ms of this P1 latency. In this case networks trained with real clinical data yielded superior results to those taught using synthetically generated mfERG waveforms.

5.10 Discussion

The aim of this study was to investigate the possibility of using ANNs to classify a mfERG waveform as delayed, within normal timing limits or as no significant response. Networks were initially trained using synthetically generated waveforms in an attempt to overcome some of the limitations associated with real clinical data. The highest performing synthetically trained ANN classified 99% of the synthetic test set

correctly however when presented with clinical data its performance fell to 70%. This implies that the synthetic testing set was very similar to the data used to train the network while the clinical data differed greatly. Although the training data incorporated waveforms with a wide range of P1 amplitude and latency values, in addition to different types of noise including baseline drift and 50Hz, the basic underlying shape of each artificially generated waveform was similar to that of one of the 50 primary waveforms. The synthetic testing set, although not used to train the network, was therefore familiar to the ANN, explaining why such a high performance was yielded. In contrast the clinical testing set comprised 1000 waveforms unrelated to the training set. It may therefore be possible to improve the generalisation of the synthetically trained ANN by increasing the number of primary waveforms utilised to generate the artificial data set, as greater variety in the underlying shape would be incorporated into the learning process.

The performance attained by the network trained using clinical data was higher than that of the ANN taught with artificial data, with an agreement of 78% between itself and the expert. It is likely that this improved generalisation can be attributed to the greater variety in the basic shape of the training mfERG waveforms. The ability of the network to classify the mfERG waveforms into one of three groups was however relatively low. The majority of networks reported in the literature are utilised to classify physiological signals into one of two groups, for example normal or abnormal. It was previously stated in chapter 3 that accuracies of 94%, 97% and 98% have been attained when categorising the EOG (136), the VECP (137) and the PERG (104) respectively into one of two classes. When the capacity of the clinically trained network to distinguish a mfERG with a 'response' from one with 'no significant response' (i.e. one of two categories) was assessed an accuracy of 90% was achieved while 86% of responses were correctly classified as delayed or within normal timing limits. These results are lower than those presented for other physiological data. It should however be noted that each of these three types of physiological signal are in an order of microvolts, compared with nanovolts in the case of the mfERG. Consequently the responses are in general easier to distinguish from noise thus easing their interpretation. Furthermore a single network was trained to categorise the EOG, the VECP and the PERG data into one of two categories as opposed to three in the case of the mfERG thus simplifying the task for these ANNs. Finally, the number of inputs used for the networks taught to classify the EOG, the PERG and the VECP was considerably less than that utilised by the ANN developed in this chapter thus

reducing the complexity of the problem in each case. For these reasons it is unsurprising that a slightly lower performance was achieved when analysing the mfERG data.

Ideally a higher accuracy would have been achieved. It would be assumed that increasing the size of the training data set would result in an increased performance as the ratio of training examples to weights when teaching the ANN was two, considerably less than that recommended by many investigators. However it was shown that the networks trained using the two synthetic data sets, one large and one small, yielded a similar performance to one another. As the use of the validation set in conjunction with the smaller training set was comparable with the larger data set it is unlikely that a superior ANN would be obtained by significantly increasing the number of training examples. One approach which could however be taken to improve the network would be to increase the number of waveforms in the training data set which have a P1 latency of 41-44ms (i.e. close to the threshold between normal and abnormal). This is likely to improve the network's capacity to state if a response is delayed or within normal timing limits, as 85% of this type of misclassification involved responses within this group. Additional examples of waveforms with no significant response could also be included in the training data to attempt to increase the sensitivity of the network, the principal source of error when differentiating 'response' from 'no response'. It is evident that the ANN has a potential role to play in the analysis of the mfERG but could not be used in isolation as a higher classification accuracy would be required; it could however be used in conjunction with other techniques as part of a multilayered system.

5.11 Conclusions

ANNs were trained to classify mfERG data into one of three categories: delayed; within normal timing limits; or no significant response. Both synthetically generated data and real clinical mfERG waveforms were utilised to train the networks. It was found that superior results were achieved when using clinical data, with a classification accuracy of 78%. When differentiating between 'response' and 'no response' the expert and the ANN concurred for 90% of the examples; an agreement of 86% was achieved when stating if the latency of a response was normal or abnormal, with 85% of misclassifications occurring close to the threshold between normal and delayed. Although it would not be possible to rely solely on this technique

to analyse the mfERG it may be possible to incorporate it into a multilayered system. Additional methods are investigated in chapter 6 with a view to increasing the objectivity of the mfERG analysis, including the use of the signal to noise ratio, digital and wavelet filtering and curve fitting techniques.

6 SNR, spline fitting and digital signal processing

The possibility of using the frequency domain to distinguish a physiological response from a waveform with no significant retinal function was studied in chapter 4, while chapter 5 investigated the ability of artificial neural networks to classify the mfERG data into one of three categories: no significant response; a delayed response; or a response within normal latency limits. Each approach showed potential but could not be utilised as the sole technique to automatically analyse the data, as a higher performance would be required. A number of additional methods were therefore studied.

This chapter initially investigates the efficacy of using the signal to noise ratio (SNR) to distinguish a waveform with no significant response from one representing an area of functioning retina. Two methods were described for calculating the SNR in chapter 3; that proposed by Zhang *et al.* (76) using a noise window and a signal window, and that introduced by Keating *et al.* (81) employing 'dead' (unused) m-sequences to represent the noise. The former has been used for the analysis of clinical mfERG data (77-79) whereas the dead sequence method has thus far only been applied to mfVECP recordings (138) and mfERG responses acquired from healthy and compliant individuals (81). This chapter therefore directly compares the ability of these two SNR techniques to classify clinical waveforms as 'response' or 'no response' with a view to finding an optimal method for making this distinction.

For those waveforms with retinal function the user must assess if the P1 component is delayed or within normal timing limits. This chapter investigates the possibility of using spline fitting, a technique used to fit curves to data, to classify responses as 'delayed' or 'not delayed'. In addition to assessing the latency of a response it is necessary to distinguish responses which are decreased in amplitude from those which are within normal limits. The ability of the spline fitting technique to make this distinction has therefore been studied.

Finally a number of digital signal processing techniques including digital filtering and wavelet analysis, both of which were discussed in chapter 3, were investigated with a view to reducing the noise present in the responses as many of the difficulties encountered when analysing mfERG waveforms are due to noise. Both the SNR and

the spline fitting experiments were then repeated after filtering the responses in the hope that the performance of each technique would be improved.

6.1 SNR for 'response' or 'no response' classification

The feasibility of using the SNR value to classify a waveform as 'response' or 'no response' is discussed initially.

6.1.1 Methods

1000 mfERG waveforms were chosen from 100 patient trace arrays. The recording quality of these varied from excellent to very poor and they incorporated a wide range of retinal function ranging from normal to no significant response. Each waveform was graded as 'response' or 'no response' by an expert. These were the same waves as those used in chapters 4 and 5 when testing the ability of the Fourier profile and the ANN to differentiate between 'response' and 'no response', thus enabling a fair comparison of the techniques.

6.1.1.1 Noise window and signal window (method 1)

The SNR value was initially calculated for each of the 1000 mfERG waveforms using the noise window and signal window method proposed by Zhang *et al.*. When using the windowing method a time period must be selected for both the signal window and the noise window. Previous studies have utilised the first 80ms of the mfERG waveform for the signal and the final 80ms for the noise (77) therefore these values were chosen. To compare the two approaches for calculating the SNR the following equation, based on that employed by the dead sequence method, was used:

$$SNR_i = 20\log_{10}\frac{RMS(signalwindow)_i}{RMS(noisewindow)_{average}} \qquad (6.1)$$

This differs to that detailed in the literature however it enabled a fairer comparison of the two methods, as this is the same as the equation utilised by the dead sequence method (refer to equation 3.3). 1000 SNR values were thus obtained. The mfERG responses were categorised into one of two classes ('response' or 'no response') based on their SNR value using a number of different thresholds: 0; 1; 2; 3; 4; and 5dB respectively. If a waveform had an SNR ratio of less than or equal to the threshold it

was classified as 'no response' while it was said to have a response if its SNR was greater than the threshold. The percentage of agreement between the expert and the SNR approach was calculated for each threshold value, allowing the optimum cut off SNR to be found.

6.1.1.2 Dead sequence (method 2)

The SNR was subsequently calculated using the dead sequence approach described by Keating *et al.*. It was stated in chapter 3 that the dead sequence method utilises an unused m-sequence when calculating the SNR. One such sequence was therefore selected and cross correlated against the raw, uncorrelated data used to recover the responses in the trace array. Figure 6.1 shows an example of a response containing only noise which was recovered using an unused sequence:

Figure 6.1 A waveform recovered by cross correlating the raw data against an unused (dead) sequence. It comprises only noise.

In order to compare methods 1 and 2 (the windowing and the dead sequence methods, respectively) the same time window was chosen to represent the signal for each approach; the first 80ms of both the response and the dead sequence were therefore chosen for the signal and the noise respectively. The SNR of each waveform was obtained using equation 3.3 (see chapter 3). Again the threshold value providing the best distinction between 'response' and 'no response' was established by varying it until the highest agreement between the expert and the computer was achieved.

6.1.2 Results

6.1.2.1 Noise window and signal window (method 1)

The percentage of waveforms which were given the same classification by both the expert and the windowing SNR method can be seen for the following 6 threshold values:

Cut off (dB)	**0**	1	2	3	4	5
% agreement with expert	**68**	67	63	60	56	53

Table 6.1 Percentage of waveforms correctly analysed as 'response' or 'no response' using the windowing SNR method. The cut off value defining the threshold between a physiological and no significant function was varied.

The highest level of agreement (68%) was found when a cut off of 0dB was chosen; all responses with a SNR of 1dB or greater were said to have a response. The sensitivity and specificity values for this threshold value were 52% and 75% respectively. Stating that a waveform with no significant response contained a physiological signal (i.e. false negative) was therefore a considerable problem. The level of disagreement, 32%, with the expert was reasonably high.

6.1.2.2 Dead sequence (method 2)

The performance achieved using the dead sequence method to calculate the SNR can be seen in table 6.2:

Cut off (dB)	0	1	2	**3**	4	5
% agreement with expert	76	77	77	**78**	77	74

Table 6.2 Percentage of waveforms correctly analysed as 'response' or 'no response' using the dead sequence SNR method. The value defining the threshold between a physiological and no significant function was altered.

The closest agreement with the expert, 78%, was observed when a 3dB cut off was chosen. Sensitivity and specificity values of 47% and 91% respectively were calculated, revealing that misclassifying a waveform which had no function as 'response' was the principal cause of error.

The following trace array, originally classified by an expert, is used to demonstrate pictorially the performance of the two approaches. All responses shaded with pink were classified as 'no response' by the expert while those in green were said to have a response:
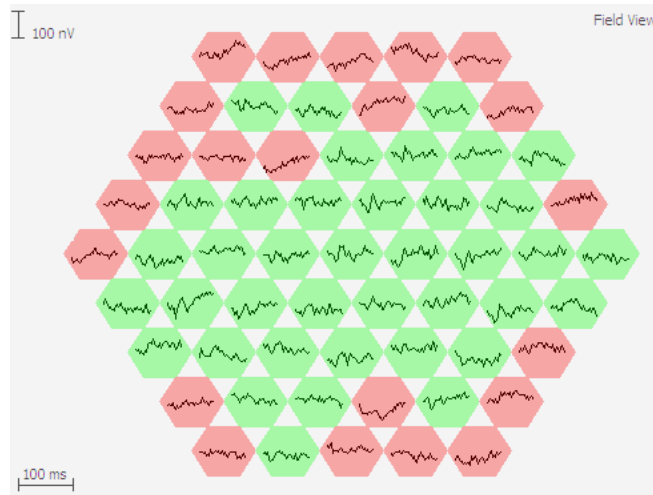
Figure 6.2 Trace array for which the expert classified each waveform as 'response' or 'no response'. Those said to have no significant retinal response are depicted in pink while all those with a physiological response are shown in green.

When the trace array in figure 6.2 was classified using method 1, the windowing approach, with a threshold value of 0dB, the results shown in figure 6.3 (left) were obtained. Again pink indicates no function and green states that there is a physiological response. The highlighted waves seen in figure 6.3 (right) represent the discrepancies between the expert and the classifications based on the SNR value:
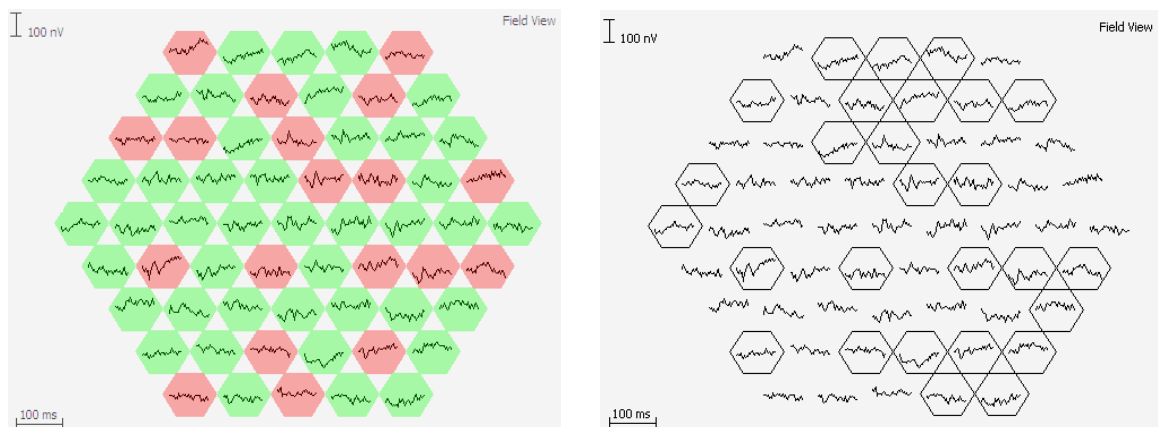
Figure 6.3 Classifications for the trace array in figure 6.2 when the windowing SNR method was utilised to categorise each waveform as 'response' or 'no response' (left). The classification differences between the expert and the SNR method are shown (right). Pink equates to 'no response' while green corresponds to 'response'. The right trace array highlights the discrepancies between the expert and the SNR approach.

27 of the 61 waveforms (44%) were classified differently from the expert when using method 1, which is a considerable difference. Figure 6.4 shows the classifications (left) and the discrepancies (right) when the SNR was calculated for each waveform in figure 6.2 using method 2, the dead sequence technique, with a threshold of 3dB:
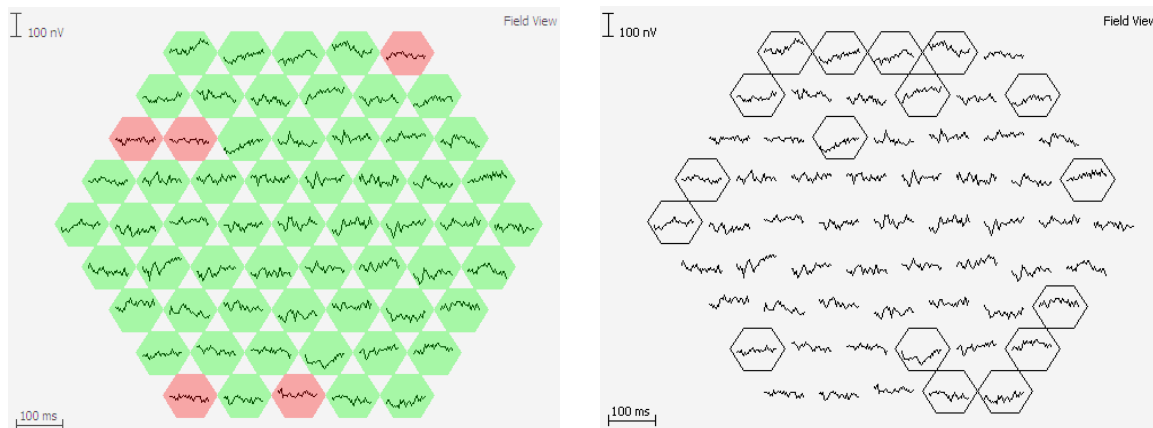
Figure 6.4 Classifications when the dead sequence SNR method was used to categorise each waveform in figure 6.2 as 'response' or 'no response' (left). The classification differences between the expert and the SNR approach are shown (right). Pink equates to 'no response' while green corresponds to 'response'. The right trace array highlights the discrepancies between the expert and the SNR approach.

In this instance 17 of the mfERG waveforms (28%) were classified differently by the expert and method 2. Classifying a waveform as a physiological response when there was no significant response was the greatest source of error. This is consistent with the relatively low sensitivity found when examining the 1000 individual waveforms.

It is evident when examining the trace array in figure 6.2 that this was a relatively difficult recording to analyse for two reasons. Firstly many of the responses were significantly attenuated, thus making the distinction between 'response' and 'no response' challenging. Secondly there was a considerable amount of baseline drift present. On closer inspection it was observed that many of the waveforms mistakenly identified as representing an area of functioning retina were affected by baseline drift. In the following trace array the mfERG waveforms labelled as A and B were both classified as 'no response' by the expert. It can be seen that the principal difference between the two waveforms is the amount of baseline drift present; it is significant on wave A while minimal on wave B:
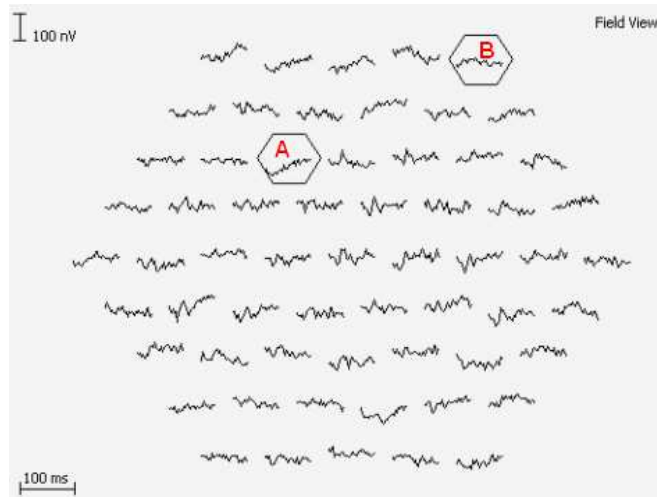
Figure 6.5 Trace array highlighting the potential problems arising from baseline drift. Two waveforms have been labelled as A and B, both of which represent areas of no significant retinal function. A is affected by baseline drift while B is relatively unaffected by drift. A was classified as 'response' by each SNR method while B was classed as 'no response'.

When the SNR values were calculated for the waveforms A and B using method 1 they were 8dB and -5dB respectively. Waveform A was thus categorised as 'response'. Similarly, when using method 2, SNR values of 15dB and -1dB were calculated for A and B respectively, again misclassifying A.

It was found both when studying the 1000 waveforms and the trace array in figure 6.2 that superior results were achieved when the noise was calculated using a dead sequence rather than the latter part of the mfERG wave. The performance was however relatively low for each approach, with each method suffering from poor sensitivity. By removing baseline drift from the response it may be possible to improve the ability of the SNR to discriminate a physiological response from one with no significant function. This will be investigated later on in the chapter.

6.2 Spline fitting

Section 6.2 explores the possibility of using spline fitting to obtain the P1 latency of a mfERG response and hence its ability to classify a response as 'delayed' or 'not delayed'. A spline is a common technique used to fit curves to a data set which changes in shape along its x-axis. Knot points are defined along the horizontal axis, splitting the data into a number of different sections; a polynomial of degree n is then fitted to each segment of the data. The optimal fit is found using the least squares method, a procedure utilised to establish the curve most appropriate to a data set. This is done by calculating the difference between the fitted curve and the actual data at each data point, squaring each difference and then summing the results. This process

is carried out for each curve; the optimal solution is that with the smallest summation value.

The mfERG system utilised throughout this thesis enables the user to fit a spline to the waveforms. The operator dictates both the resolution and the order of the spline. The former defines the number of segments into which the data are divided while the latter determines the order of polynomial used. To locate P1, the maximum value of the spline between data points 42 and 80 (corresponding to 35-67ms when using a stimulus frequency of 75Hz) is found. These values were chosen as they encompass the range of P1 latencies seen clinically when using the particular experimental protocol used for this thesis. The time value at this maximum point was therefore defined as the latency of P1. The following image demonstrates a spline fit to a mfERG waveform: the blue image is the physiological response while the red curve is the spline. Two red circles can be seen, the second of which is where the spline has located P1. It can be seen that this is the maximum point on the spline. The first circle, the estimation for N1, is simply the minimum value of the mfERG waveform in the first 50 data points (41.5ms when the stimulation frequency is 75Hz). The spline places cursors for N1 and P1 at the data points on the mfERG response (as opposed to on the spline) corresponding to the latencies found by the spline. The latency of P1 will be concentrated on in this section.
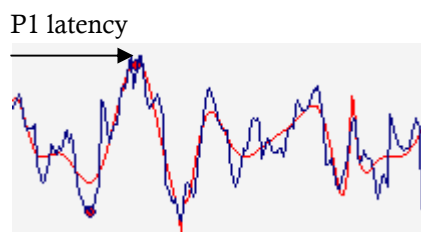
Figure 6.6 The method by which the spline locates P1. The mfERG response is shown in blue while the spline fit is red; the first circle, seen at the first trough, is where the spline has located N1. The second circle, at the peak of the spline, corresponds to the P1 latency, as found by the spline. It has located P1 accurately; in this instance the maximum value of the mfERG waveform would have been inappropriate for locating P1.

The aim was to establish the most effective spline parameters (polynomial order and resolution) for locating P1 and hence distinguishing responses which are delayed from those which are within normal timing limits.

6.2.1 'Delayed' or 'not delayed' classification

6.2.1.1 Methods

The 1000 waveforms utilised in section 6.1 to study the potential of the SNR to distinguish between 'response' and 'no response' were used. 694 of these waveforms had been classified as 'response'; the expert therefore stated the P1 latency for each of these. In chapter 3 the normative range was presented for the mfERG, showing that responses with a P1 latency of 42ms or less were within normal limits. Each of the 694 responses was therefore classified as 'delayed' or 'not delayed' based upon this normal range.

The spline parameters were varied to ascertain the spline which would yield the closest agreement with the expert. The mfERG system allows both the order of the polynomial and the resolution to be varied from 1 to 15. Upon inspection it was found that when using a polynomial with an order of four or less the spline was unable to form a suitable fit to the data while a spline using a polynomial with an order of ten or greater fitted too closely to the data. The latter therefore lacked the generalisation required in cases when P1 was not the maximum value. This can be seen in figure 6.7 where polynomials of degree four (left) and ten (right) were chosen:
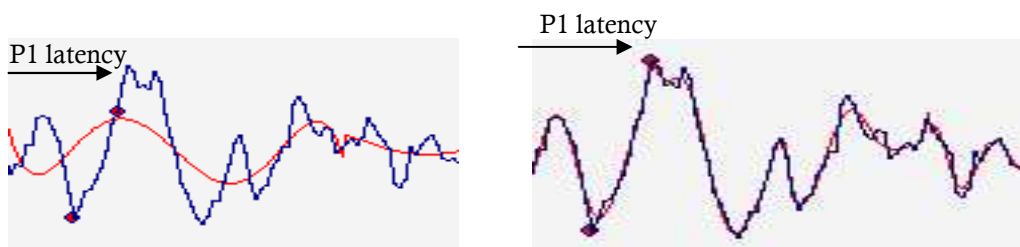
Figure 6.7 Locating P1 with a spline using a fourth order polynomial (left) and a tenth order polynomial (right). Again the mfERG response is shown in blue while the spline fit is red. Neither the fourth nor the tenth order spline located P1 accurately.

Polynomials with an order of 5, 6, 7, 8 and 9 were therefore investigated. Similarly it was observed that when a resolution of less than 6 or greater than 10 was chosen a poor fit to the data was obtained therefore resolution values of 6, 7, 8, 9 and 10 were considered. 25 different spline parameters were thus evaluated (5 different degrees of polynomial, each with one of the 5 resolutions). For each set of parameters a P1 latency value was determined for a response, enabling it be classified as 'delayed' or 'not delayed'. The classifications provided by the expert and the spline were compared for each of the 694 waveforms allowing the agreement between the expert and the

spline to be calculated. This was done for all 25 spline parameters to establish those most effective for stating if a waveform is delayed or within normal limits.

6.2.1.2 Results

The following table details the results obtained when using the 25 different spline parameters:

Polynomial Order	Resolution	% agreement with expert
5	6	80
	7	81
	8	81
	9	82
	10	**85**
6	6	82
	7	82
	8	83
	9	84
	10	84
7	6	83
	7	79
	8	83
	9	84
	10	84
8	6	78
	7	84
	8	84
	9	83
	10	82
9	6	82
	7	82
	8	80
	9	82
	10	77

Table 6.3 Agreement between the expert and the different splines when classifying responses as 'delayed' or 'not delayed', based on their P1 latency. The order and the resolution of the spline was altered to find the optimal parameters; the highest agreement was found when using a 5th order polynomial and a resolution of 10.

It is evident that the most promising results were achieved when using a spline with a 5th order polynomial and a resolution of 10 (O5_R10); an agreement of 85% between the spline and the expert was achieved. The sensitivity and specificity were 90% and 74% respectively when this particular spline was utilised implying that classifying a response as 'delayed' when it had a P1 latency of less than 43ms was the greatest problem. When those responses which were incorrectly categorised were inspected it

was noted that for 59%, the P1 latency (as stated by the expert) was within 1ms of the boundary between normal and delayed (i.e. 42-43ms), while 76% were found to be within 2ms of this threshold (i.e. 41-44ms). There was very little difference in performance between many of the different splines investigated however as O5_R10 yielded the optimal results it will be discussed for the remainder of the chapter.

6.2.2 Ability to locate P1 accurately

This section investigates the ability of the spline fitting technique to locate P1 accurately.

6.2.2.1 Methods

In section 6.2.1 the expert and the spline O5_R10 stated a P1 latency for each of the 694 waveforms. These values were exported to Excel where the timing difference between the spline's and the expert's assessment of P1 was calculated for each response. The median discrepancy was calculated for the set of mfERG responses, as was the maximum timing difference and the 95th percentile for the discrepancy.

6.2.2.2 Results

The median, the maximum and the 95th percentile for the timing difference were 2ms, 30ms and 15ms respectively. A maximum timing difference of 30ms is significant, as is a 95th percentile value of 15ms. These values imply that spline fitting could not be used with confidence for stating the actual P1 latency and hence for comparing the latency of responses within one trace array or from sequential visits. It was of interest to dissect the results further to reveal the cases where spline fitting performed well and those for which it made mistakes. Methods of increasing the accuracy of the technique could then be investigated. The following table details the errors made, showing the percentage of waveforms correctly analysed to within specified time periods:

Difference (ms)	0	≤ 1	≤ 2	≤ 3	≤ 4	≤ 5	≤ 10	≤ 15	≤ 20
% of waveforms correctly analysed to within each time difference	12	37	65	75	85	86	93	95	96

Table 6.4 Examing the difference in P1 latency defined by the spline and the system – the percentage of responses correctly analysed to within different time periods are shown. It can be seen that the expert and the spline stated the same P1 latency for 12% of the responses. They were within 1ms of each other for 37%, while for 65% of responses analysed they were within 2ms of one another etc.

It can be seen that only 12% of the waveforms were given the same P1 value by the expert and the spline, while the spline and the expert were within 2ms of each other for 65% of the responses. The percentage of responses for which a significant error was made was considerable with 7% of the mfERG waves having an error of greater than 10ms and 4% with a difference of more than 20ms. When studying the data it was observed that the ability of these splines to locate P1 accurately was dependent on the quality of the mfERG response. When the response was clearly defined and relatively noise free the spline performed well. This can be seen in figure 6.8, with the expert's estimate of P1 shown first, followed by that of the spline O5_R10:
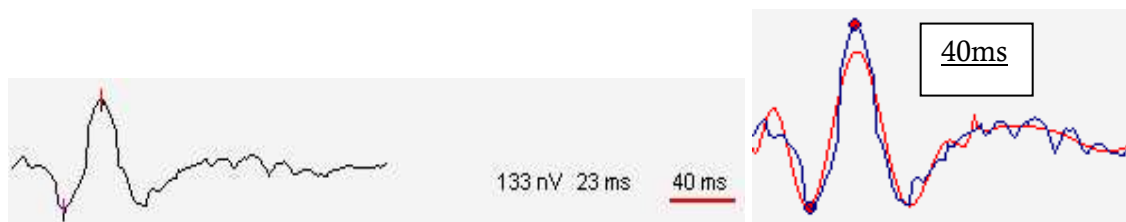
Figure 6.8 Example 1: comparing the P1 located by the expert (left) with that stated by the spline (right). The P1 latency defined by the expert has been underlined in red, while that found by the spline is shown inside the box. The spline and the expert agreed in this instance.

The spline also performed well when P1 was not at the maximum value. Figure 6.9 demonstrates this, again with the expert's cursor shown first followed by that identified by O5_R10:
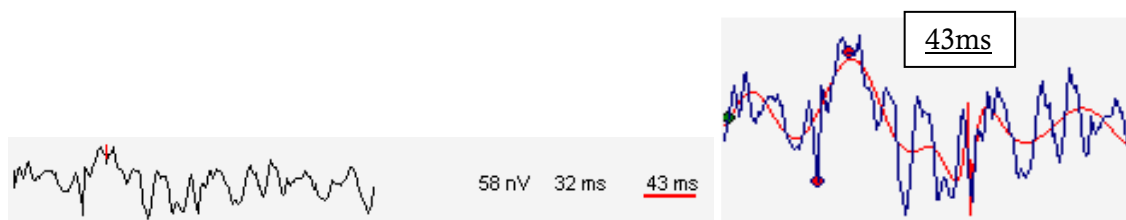
Figure 6.9 Example 2: the P1 latency as stated by the expert (left) and the spline (right). Both the spline and the expert located P1 at 43ms.

It was however seen in table 6.4 that for 7% of the responses an error of more than 10ms was made by the spline. It was found that baseline drift and noise was a problem for this group of waveforms. The following waveform, which has been contaminated by both noise and baseline drift has been used to demonstrate this. The location of P1 as stated by the expert (left) and the spline (right) can be seen:

Figure 6.10 Example 3: the location of P1 as stated by the expert (left) and the spline (right). There was a 27ms discrepancy between the spline and the expert when analysing this noisy mfERG response.

A significant mistake was made by the spline fit; 27ms in this instance, thus misclassifying it as 'delayed' when it was within normal timing limits. It is evident that in the case of high quality recordings this technique has the potential to provide an accurate assessment of the P1 latency but that large errors were often made when baseline drift or noise was present. The performance of this approach may be improved if the baseline drift was removed from the waveform and if the noise was decreased. Various methods commonly used to minimise noise were discussed in chapter 3 including digital filtering and wavelet analysis. Each of these will be considered with a view to improving the signal quality of the mfERG waveforms. The efficacy of using spline fitting to classify responses as 'delayed' or 'not delayed' and to locate P1 successfully will then be tested to ascertain if it is improved by the application of digital signal processing techniques.

6.3 Removal of drift

The removal of baseline drift is studied initially. Baseline drift is a low frequency artefact caused by patient factors such as blinking and eye movement, resulting in a skewing of the waveforms. The following trace array shows an example of such a recording:
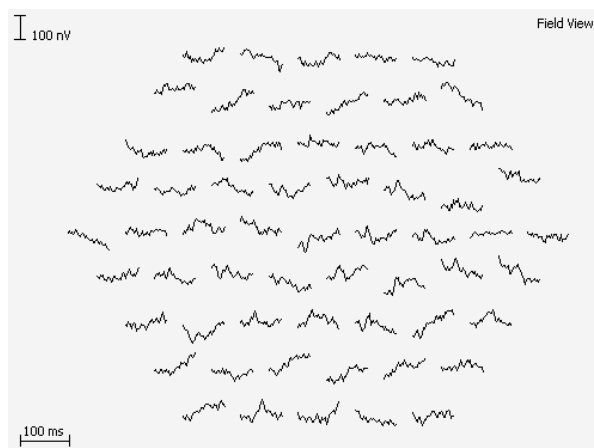
Figure 6.11 An example of a trace array affected by baseline drift.

High pass filters, those designed to attenuate frequencies lower than a predefined frequency can be used to minimise the baseline drift. Figure 6.12 shows the trace array in figure 6.11 after the application of a 3Hz high pass filter (left) and the equivalent after using a 10Hz high pass filter (right):
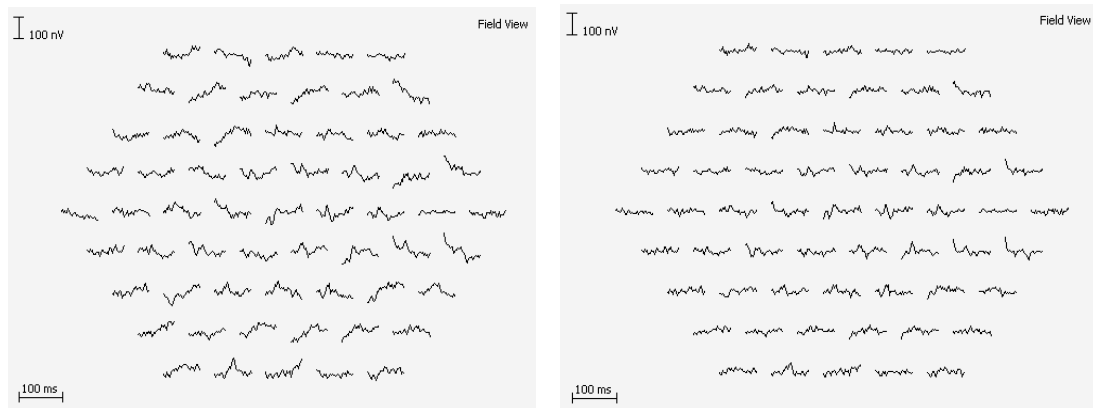
Figure 6.12 The effect of high pass filtering on the trace array shown in figure 6.11. The left trace array was obtained after the application of a 3Hz high pass filter while that shown on the right was acquired by using a 10Hz high pass filter. The baseline drift is still present when using the 3Hz high pass filter but it is less evident when using a cut off frequency of 10Hz.

It can be seen that the baseline drift was decreased considerably by using the 10Hz high pass filter but that it remained a problem when using the 3Hz filter. There is however a limitation associated with this technique; the signal can be distorted by the filter, resulting in a shift of P1. To demonstrate this problem an individual waveform with a P1 latency of 40ms (top waveform in figure 6.13) is utilised. The same waveform after the application of a 3Hz high pass filter (middle) and a 10Hz high pass filter (bottom) is shown:
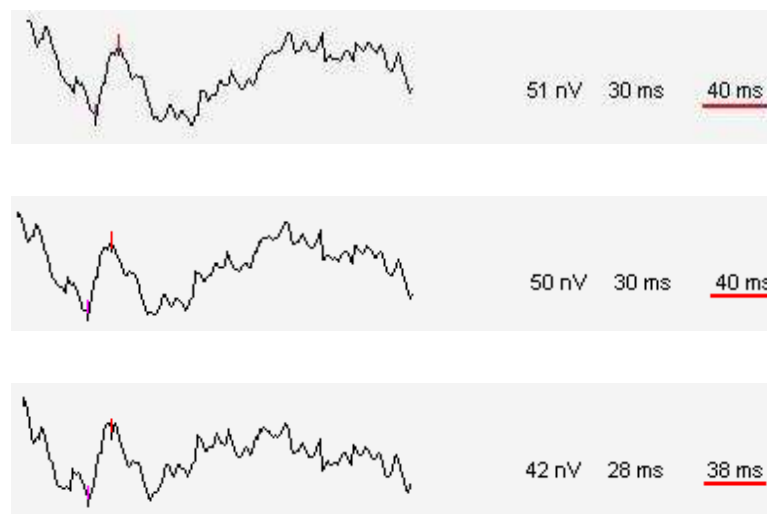
Figure 6.13 The effect of filtering on an individual mfERG waveform. A waveform before filtering is shown first (top). The same waveform is shown after the application of a 3Hz high pass filter (middle) and finally after applying a 10Hz high pass filter (bottom). The P1 latency was unaffected by using the 3Hz high pass filter but was shifted by 2ms when the 10Hz high pass filter was applied.

P1 remained at 40ms when using a 3Hz high pass filter but was shifted to 38ms when using the 10Hz filter. As the timing of the P1 component is of particular importance when analysing mfERG waveforms a latency shift such as this is unacceptable. A decrease in the P1 amplitude was also noted when utilising the 10Hz high pass filter; ideally a change in amplitude would be avoided as this measurement is used clinically when analysing the responses. An alternative method was therefore investigated to remove this low frequency artefact: the use of spline fitting.

6.3.1 Methods

The aim was to fit a spline to the mfERG waveform and to subtract the y-value of the spline from the y-value of the response at each data point. To demonstrate this a waveform affected by baseline drift was selected from the trace array shown in figure 6.11 This can be seen in figure 6.14 (left). The mfERG waveform is depicted in blue while the red curve shows the spline fit; a 2nd order polynomial with a resolution of five was chosen in this case. When the value of the spline was subtracted from that of the mfERG response at each data point the waveform shown on the right of figure 6.14 was obtained:

Figure 6.14 Demonstrating the removal of baseline drift from an individual waveform using the spline fitting technique. The mfERG waveform affected by drift (left) is shown in blue while the 2nd order spline is shown in red. It can be seen from the waveform on the right that the drift has been successfully removed.

It is evident from figure 6.14 that the baseline drift originally present on the waveform is now absent. In order to select the most appropriate spline for this purpose, polynomials with an order of one and two were used while a resolution of five was chosen. Only low order polynomials were considered as the aim was to model the general trend of the data (e.g. an upwards tilt) rather than to obtain a close fit to the data. 20 responses were then selected to ascertain if this technique distorts the response; the P1 latency of each waveform was assessed before and after the removal of drift to determine if the responses were shifted in time. 15 of these were contaminated by baseline drift while 5 had minimal drift. Responses both with and

without drift were chosen to ensure that this technique could be applied to all waveforms, irrespective of the amount of baseline drift present, if successful.

6.3.2 Results

When splines with a polynomial of order 1 (left) and 2 (right) were applied to the trace array shown in figure 6.11 the following results were obtained:
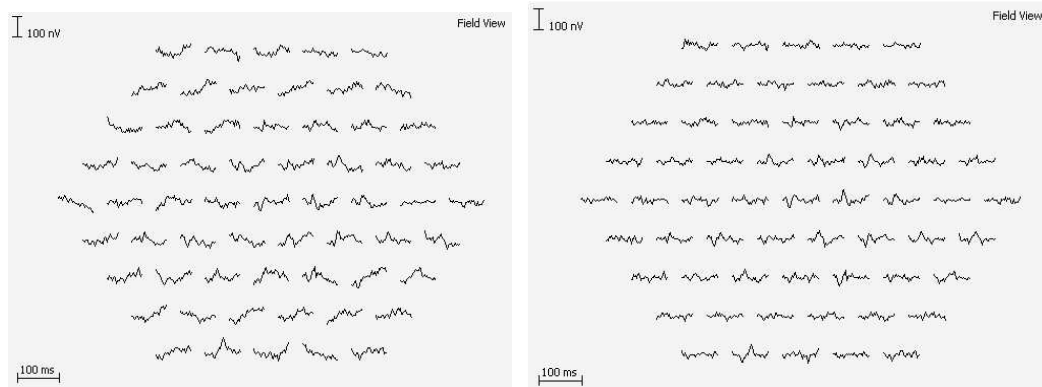
Figure 6.15 The removal of baseline drift using a spline with a first order (left) and a second order (right) polynomial. Baseline drift is still evident when a first order polynomial was employed whereas it is minimal on the right trace array, for which a second order polynomial was utilised.

It can be seen that using a first order polynomial was insufficient to completely remove the baseline drift while a spline with a polynomial of degree two removed the drift from the signals. When the latter was applied to 15 responses affected by baseline drift, the drift was successfully removed from the waves while the latency of P1 was unaffected. This can be seen in table 6.5:
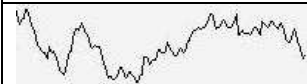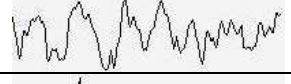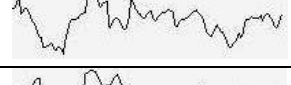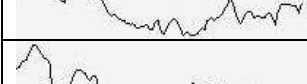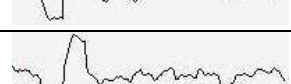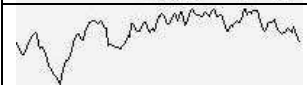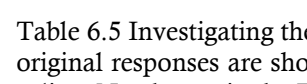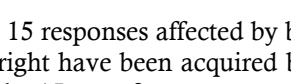
Original waveform	P1 latency (ms)	Change in P1 latency (ms)	New waveform
	40	0	
	56	0	
	43	0	
	45	0	
	51	0	
	50	0	
	40	0	
	43	0	
	39	0	
	41	0	
	43	0	
	40	0	
	41	0	
	48	0	
	49	0	

Table 6.5 Investigating the effect of subtracting a spline from 15 responses affected by baseline drift. The original responses are shown on the left while those on the right have been acquired by subtracting the spline. No change in the P1 latency was observed for any of the 15 waveforms.

No significant difference in the amplitude of P1 was noted for any of the 15 waveforms. The equivalent results when applied to responses with negligible drift can be seen in table 6.6:
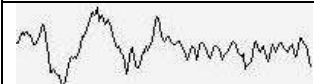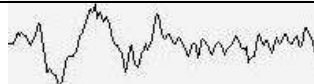
Original waveform	P1 latency (ms)	Change in P1 latency (ms)	New waveform
	39	0	
	55	0	
	50	0	
	40	0	
	44	0	

Table 6.6 Studying the effect of subtracting a spline from five responses unaffected by baseline drift. The responses before (left) and after (right) subtracting the spline are shown. No change in the P1 latency was observed for the five responses.

Again none of the responses were shifted in time and the amplitude of the signals was unaffected. This technique can therefore be applied effectively to remove the drift from the mfERG waveforms without distorting the signal; it will be used for the remainder of the chapter.

6.4 Digital filtering

In addition to baseline drift, noise on signals caused problems when analysing the data therefore section 6.4 investigates the applicability of digital filtering techniques. The mfERG system used throughout this thesis offers a number of different digital filtering options. These include selecting the filter response, the frequency range of the filter and the type of filter. Each of these will be described in the following sections.

6.4.1 Filter response

The four types of filter response available to the user are low pass, high pass, bandpass and bandstop. A low pass filter is one which passes low frequencies while attenuating high frequencies. In contrast a high pass filter passes high frequencies and attenuates lower ones. A bandpass filter stops all frequencies out with a frequency range defined by the user while a bandstop filter blocks frequencies within a specified range. For each type of filter a cut off frequency must be defined. This is the frequency at the boundary between a passband and a stopband. For an ideal filter, signals in the passband are unaffected while all those in the stopband are attenuated. In practice this

is not the case as a real filter decreases the input power by a factor of two at the cut off frequency as opposed to eliminating it.

The order of a filter determines the rate of attenuation of the signal at frequencies in the stopband. In the case of a low pass filter, for example, this applies to all frequencies greater than the cut off frequency. A first order low pass filter halves the signal amplitude for every doubling of the frequency while a second order low pass filter attenuates the frequencies at a greater rate; the signal amplitude is quartered each time the frequency doubles. The same applies to high pass filters with the exception that the amplitude is halved when the frequency is halved for first order filters and quartered when the frequency is halved when using a second order filter. The filters incorporated into the mfERG system are all first order therefore all subsequent work utilises these.

6.4.2 Filter Type

The three types of digital filter available in the system are the Bessel, the Butterworth and the Chebyshev filter, each digitised versions of these types of analogue filters. The Butterworth filter has a frequency response which is flat (theoretically) in the passband and then tapers off to zero. Chebyshev filters have a sharper transition from the passband to the stopband however the gain in the passband varies. Bessel filters neither have an optimally constant gain in the passband nor a sharp transition from the passband to the stopband. This linear filter however preserves shapes well when filtering signals (139). Bessel filters were chosen for all subsequent analysis as they cause the least distortion to the signals.

6.4.3 Methods

To design a suitable filter for improving the mfERG signal it was important to determine the frequencies associated with signal and those related to noise; digital filtering could then be applied to attenuate those specific to noise. A number of cross correlated responses ranging in signal quality were therefore selected and transformed from the time to the frequency domain using the method described in appendix 2. Frequency components associated with both noise and signal were thus identified.

6.4.3.1 Frequency profile of mfERG responses

Three examples have been selected to present the findings, the first of which is a normal physiological signal with very little noise:
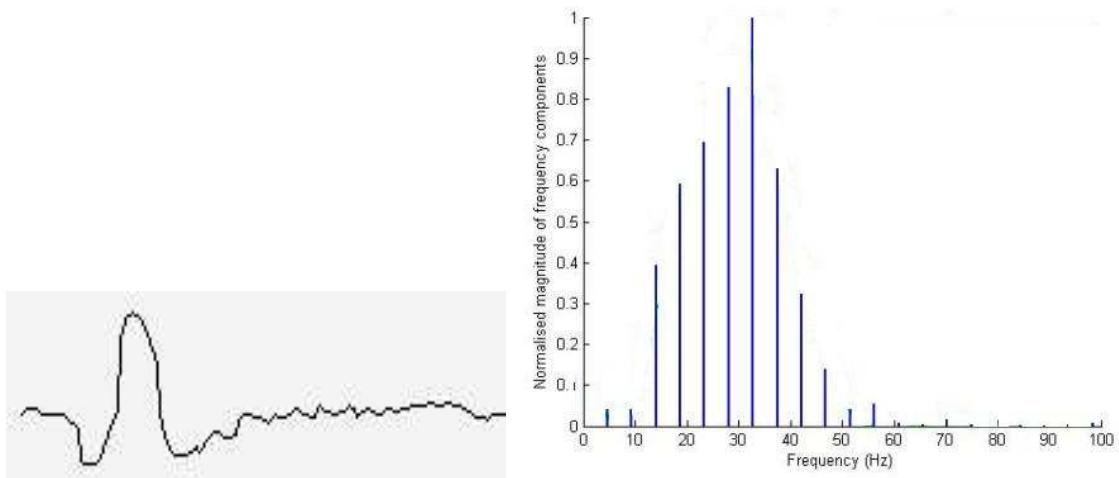
Figure 6.16 A normal waveform unaffected by noise (left) and its corresponding frequency spectrum (right). The Fourier profile predominantly comprises frequencies between 5 and 60Hz.

It is apparent that the most prominent frequency components lie between approximately 5 and 60Hz. The following example is taken from a noisy recording:
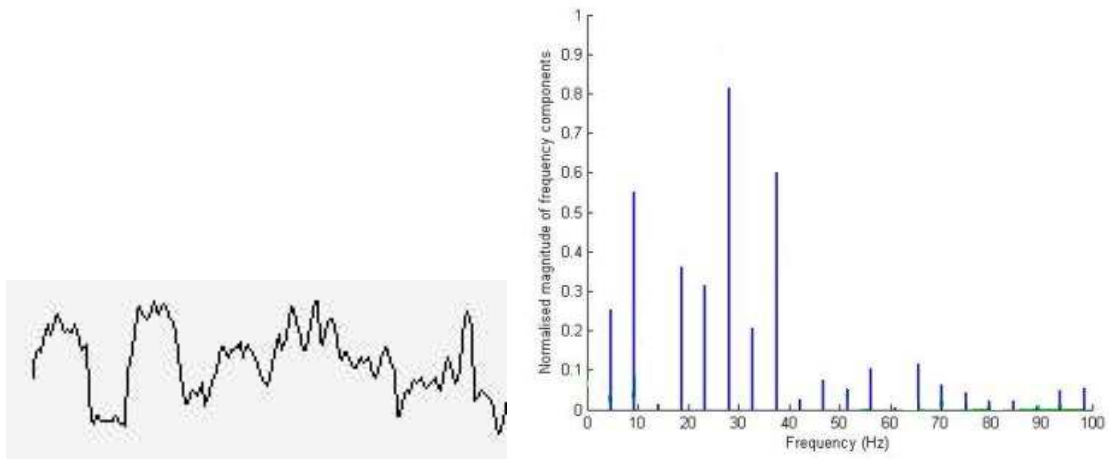
Figure 6.17 An example of a noisy waveform (left) and its Fourier profile (right). The magnitude of the frequency components less than 5Hz and greater than 60Hz was greater than those seen in figure 6.16.

Frequencies within the 5-60Hz range were again seen however those less than 5Hz and greater than 60Hz were significantly increased in magnitude relative to those observed in the good recording. A 2nd example of a noisy waveform is shown in figure 6.18:
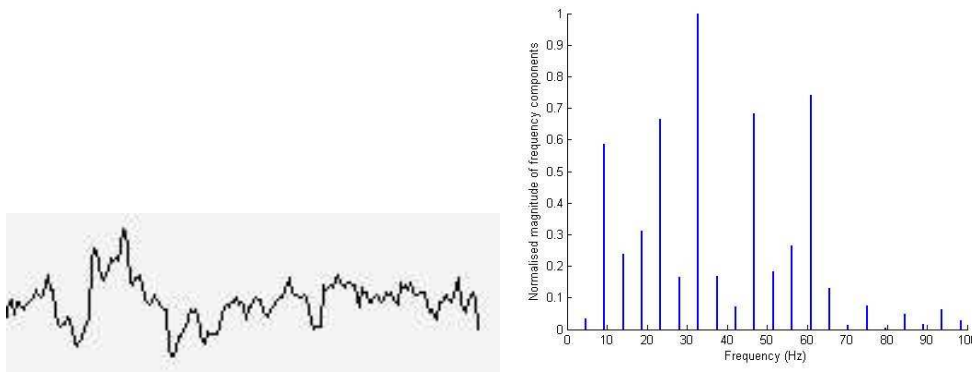
Figure 6.18 A 2nd example of a noisy waveform (left) and its frequency spectrum (right). In this instance the magnitude of frequencies at 60Hz and above was significant.

Again strong frequency components were seen above 60Hz. These findings were thus used as an initial basis when designing the different filters.

6.4.3.2 Filter responses and frequency ranges

It was evident that frequencies less than 5Hz and greater than 60Hz were predominantly associated with noise. A bandpass filter was therefore the most appropriate type of digital filter as it would attenuate both high and low frequencies. The frequency range chosen for the passband can distort a signal therefore 20 mfERG waveforms, ranging in both recording quality and retinal function were selected. The P1 latency was reported by an expert for each response prior to filtering. A number of filters, each with a different passband frequency range were then applied to the waveforms. The P1 latency was stated by the expert and compared with that prior to filtering. The upper frequency of the filter was varied from 60Hz to 100Hz in increments of 10Hz while both 3Hz and 5Hz were utilised for the lower cut off frequency.

6.4.4 Results

The following table details the impact the different filters had on the P1 latency for each of the 20 mfERG responses. The magnitude of the P1 latency change after applying the filter has been shown; '+' indicates a shift to the right. Those values shaded in orange have been shifted while those which are white have been unaffected by filtering:
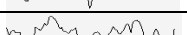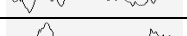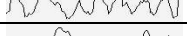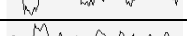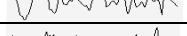
Wave	P1 lat (ms)	Bandpass frequency range (Hz)									
		3-60	3-70	3-80	3-90	3-100	5-60	5-70	5-80	5-90	5-100
		Change in P1 latency after digital filtering (ms)									
	40	+3	+3	+2	+1	0	+2	+2	+2	+1	0
	51	+1	+1	+1	0	0	+1	+1	+1	0	0
	48	+2	+2	+1	+1	0	+2	+2	+1	+1	+1
	40	+2	+2	+2	+1	0	+2	+2	+2	+1	+1
	40	+2	+2	+2	+1	0	+2	+2	+2	+1	+1
	38	+2	+2	+2	+1	0	+2	+1	+1	0	0
	54	+2	+1	+1	+1	0	+2	+1	+1	+1	+1
	45	+3	+3	+2	+2	+1	+2	+2	+2	+1	+1
	42	+2	+2	+2	+1	0	+2	+2	+1	+1`	+1
	51	+3	+2	+2	+1	+1	+2	+2	+1	0	0
	47	+4	+3	+2	+1	+1	+3	+2	+2	+1	+1
	41	+2	+2	+1	+1	0	+1	+1	+1	0	0
	40	+2	+2	+1	+1	0	+2	+2	+1	+1	0
	51	+3	+2	+1	+1	0	+2	+2	+2	+1	+1
	39	+1	+1	+1	0	0	+1	+1	+1	+1	0
	41	+3	+2	+1	+1	+1	+2	+2	+1	+1	+1
	37	+2	+2	+1	+1	0	+2	+2	+2	+2	+1
	45	+1	+1	+1	0	0	+1	+1	+1	0	0
	43	+2	+2	+1	0	0	+1	+1	+1	0	0
	40	+3	+2	+1	+1	0	+3	+3	+2	+1	+1

Table 6.7 The effect of different digital filters on the P1 latency of 20 mfERG responses. Those shaded in orange have been shifted to the right whereas those unaffected by filtering are white.

All 20 waves were shifted when using upper cut off frequencies of 60Hz, 70Hz and 80Hz. The filter using the largest passband (3-100Hz) had the least impact on the P1 latency of the responses; four of the mfERG waveforms were delayed by 1ms. These responses were however noisier than those filtered using a narrower passband as fewer of the frequencies associated with noise were attenuated. Figure 6.19 shows the original waveform (top) in addition to this response after the application of 3-100Hz (middle) and 5-60Hz (bottom) bandpass filters:
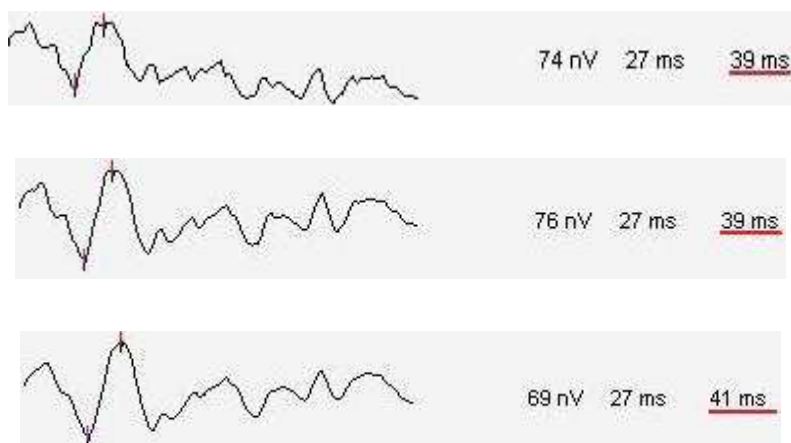
Figure 6.19 Demonstrating the effect of filtering on the P1 latency of a response. The mfERG waveform prior to filtering is shown first (top), and then after being filtered by a 3-100Hz bandpass filter (middle) and a 5-60Hz bandpass filter (bottom). The P1 latency of the middle response has been unaffected by filtering whereas the P1 latency of the bottom response has been shifted to the right by 2ms.

The waveform obtained after applying the 5-60Hz filter has a cleaner appearance than that from the 3-100Hz filter however it is delayed by 2ms relative to the original response. The latency of P1 is important when reporting the responses therefore a shift of 2ms is unacceptable. Given that the least distortion was observed for the 20 responses when using the 3-100Hz filter it was decided to utilise this filter, despite the slightly noisier appearance of the responses.

6.5 Wavelet filtering

Prior to applying the drift removal and digital filtering to the mfERG waveforms a final method was investigated to improve signal quality: wavelet filtering. It was discussed in chapter 3 that the advantage of wavelet filtering is that it decreases the noise present in the signal while maintaining the original shape. The mfERG system utilised for this thesis allows both the decomposition level and the order of the wavelet to be defined.

6.5.1 Decomposition level

When using wavelet filtering the time domain signal is passed through a series of high pass and low pass filters. At the first stage the signal is split into two parts using a high and a low pass filter. Two versions of the signal then exist: one contains the low frequencies within the signal which corresponds to an approximation, while the other contains the high frequencies, the fine detail. The same process is repeated on the low pass portion of the signal, again resulting in two versions of the signal; an approximation and fine detail. The user defines how many times this process is performed. Figure 6.20 illustrates this process:
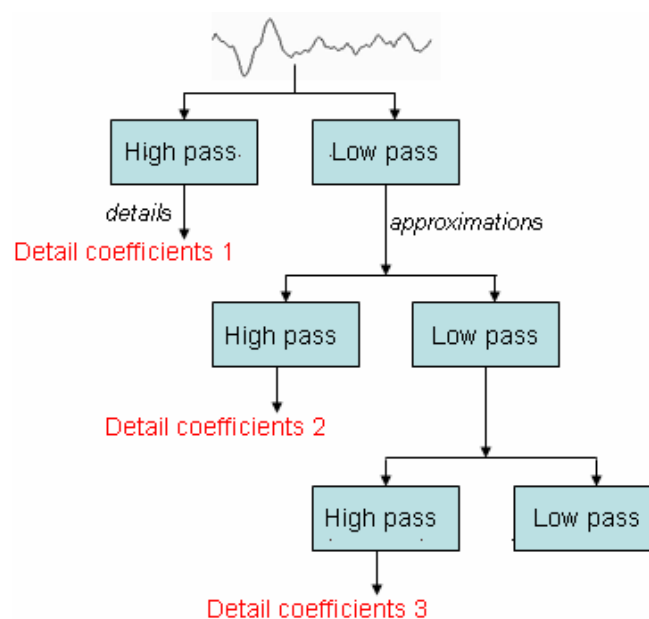
Figure 6.20 An overview of the wavelet decomposition process. The signals are filtered by both high and low pass filters; the high pass part represents the detail while the low pass part corresponds to approximations.

The multifocal ERG system allows the decomposition level (i.e. the number of times the signal is split using a high and a low pass filter) to be varied from one to four.

6.5.2 Wavelet Order

Wavelet analysis is similar to Fourier analysis in that it breaks a signal into its constituent parts for analysis. The Fourier transform decomposes the signal into a series of sine waves, each of different frequencies whereas the wavelet transform breaks the signal into a series of wavelets. While a sine wave is of infinite length and is smooth, a wavelet is more irregular in shape and has a finite length. As a result wavelet analysis performs well when there are sharp discontinuities in the data being

studied. The mfERG system enables the user to select the wavelet function used. Ten different wavelet functions are available, each of which is from the Daubechie wavelet family. These are translated and dilated versions of one another and are shown in figure 6.21:
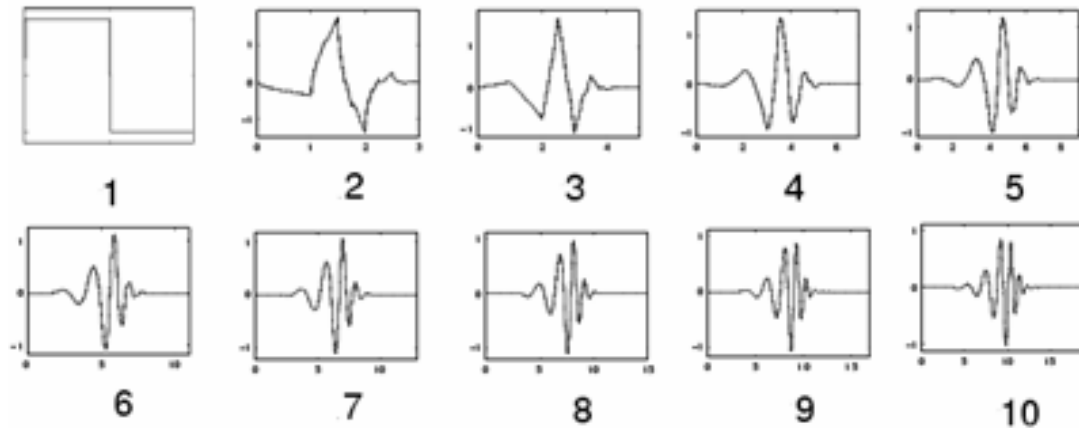
Figure 6.21 The Daubechie wavelet family (adapted from the Mathworks wavelet toolbox). 10 wavelet functions can be seen, each of which is a dilated and translated version of the others.

6.5.3 Methods

The 20 mfERG responses utilised in section 6.4.3 were once again selected to study the efficacy of wavelet filtering. Each waveform was filtered, varying both the decomposition level (i.e. from one to four) and the wavelet used (i.e. the ten waves shown in figure 6.21). The P1 latency of the waveforms after wavelet filtering was assessed by an expert and compared with the original value to determine if they had been distorted by this type of filtering. Upon initial inspection it was noted that using a decomposition level of one did not impact greatly on the signal whereas a decomposition level of four distorted the waveform significantly. All results presented therefore used decomposition levels two and three. Similarly when using the first wavelet in the family the response had a much less smooth appearance than the original wave as the wavelet utilised was a step function. This was not the case when using higher orders therefore wavelets two to ten were studied.

6.5.4 Results

The following table illustrates the shift of P1 latency after filtering the individual waveforms; a minus indicates a shift to the left while a plus is a shift to the right. Values have been shaded as grey, pink or white. These represent a shift to the left, the right and no shift, respectively:

Wave	P1 lat (ms)	Decomposition level 2									Decomposition level 3								
		Order 2	3	4	5	6	7	8	9	10	2	3	4	5	6	7	8	9	10
		Change in P1 latency after wavelet filtering (ms)																	
	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	51	-1	0	-1	-1	-1	-1	-1	-1	+1	+1	+1	+1	-1	-1	0	0	0	-1
	48	+1	+1	+1	+1	+1	+1	+1	+1	+1	0	0	0	+1	+2	+1	0	0	+2
	40	+1	0	0	0	0	0	0	0	0	+1	0	0	0	-1	0	0	0	0
	40	+1	+1	+1	0	+1	+1	+1	0	0	+1	0	0	0	0	+1	0	0	0
	38	0	0	0	0	0	0	0	0	0	+1	+1	0	0	0	0	0	0	0
	54	+1	+1	0	0	+1	+1	+1	0	0	+3	+1	+1	+2	+2	+1	+1	+1	+2
	45	0	0	0	0	0	0	0	0	0	0	0	+1	0	0	0	+1	+1	+1
	42	-1	0	0	0	0	-1	-1	0	0	0	0	-1	0	+1	+1	-1	0	0
	51	+1	0	0	0	0	0	0	0	0	+1	0	0	-1	0	0	0	-1	-1
	47	+1	+1	0	0	0	0	0	0	0	+4	+2	+2	+2	+3	+2	+1	+1	+2
	41	0	0	+1	-1	0	0	0	0	-1	0	0	-1	0	0	0	0	0	0
	40	0	0	0	0	0	0	0	0	0	+1	0	0	0	0	0	0	0	0
	51	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0
	39	0	0	0	0	-1	0	0	0	-1	+1	0	0	0	-1	+1	0	0	-1
	41	-1	0	0	0	0	0	0	0	0	0	-1	0	+1	+1	0	0	0	+1
	37	0	0	+1	+1	+1	+1	0	0	+1	+2	-1	+1	+2	+1	0	0	+2	+3
	45	0	0	0	+1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	43	0	0	0	0	0	0	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1
	40	0	0	0	0	0	0	0	0	0	+1	+1	+1	+1	+1	0	0	+2	+2

Table 6.8 The effect of different wavelet filters on the P1 latency of 20 mfERG responses. Those unaffected by the wavelet are shown in white, while those shifted to the right are shown in pink. Shifts to the left are depicted in grey.

The greatest distortion was noted when using a decomposition level of three and an order of two (i.e. the second wavelet). When using a decomposition level of two and the ninth wavelet only two of the 20 responses were shifted; neither of these was distorted by more than 1ms. Upon observation the waves obtained when using a decomposition level of three had a much cleaner appearance than those acquired when a decomposition level of two was chosen. This is due to the response being passed through an additional layer of high and low pass filters thus removing more fine detail (i.e. noise) from the signal. When a decomposition level of three was selected the least distortion was achieved when using the eighth wavelet; five of the responses were shifted, with a maximum distortion of 1ms. It was thought that these responses may be easier to analyse when using spline fitting due to their cleaner appearance. Subsequent discussion will therefore concentrate on the best wavelets when using both the second and the third decomposition levels.

The following example shows an unfiltered waveform (top), followed by the same waveform filtered using firstly a wavelet with a decomposition level of two and an order of none (D2_O9) (middle) and secondly a wavelet with a decomposition level of three and an order of eight (D3_O8) (bottom). The P1 latency has been highlighted in each case:
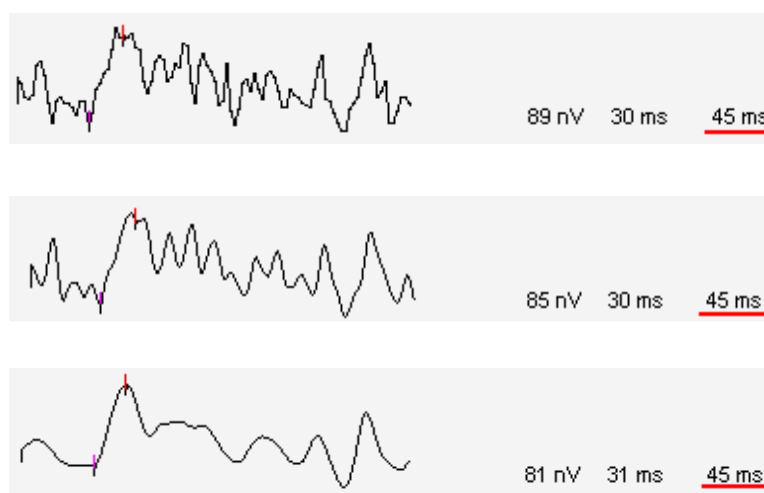
Figure 6.22 The impact of wavelet filtering on a mfERG response. The original waveform prior to filtering is shown initially (top), while that after applying the wavelet filter D2_O9 is shown second (middle). The third response shows the same mfERG waveform after applying the wavelet filter D3_O8 (bottom). The P1 latency of each filtered response is the same as that of the original waveform. Note that the response filtered using D3_O8 has a much cleaner appearance than that obtained using D2_O9.

No shift of P1 latency was observed in this case when using either of the wavelet parameters. That acquired when using the higher decomposition level has a much cleaner appearance than the other two responses, with a clear maximum point. It is

evident that wavelet analysis has the ability to decrease the noise present in the mfERG responses while the shift of the waveform is minimal.

6.6 Combining filters, wavelets and drift removal

Having investigated the removal of baseline drift, the use of digital filtering and wavelet analysis it was decided to combine these techniques to decrease the noise present in the mfERG responses. Again it was important to examine the distortion when combining the three techniques to ensure that it did not result in a significant shift of the responses. The same 20 mfERG responses utilised in previous sections were chosen. A bandpass filter with a frequency range of 3-100Hz was selected as this was shown to produce the least distortion in section 6.4.4. The baseline drift was removed using spline fitting with a second order polynomial while two wavelet filters were chosen; D2_O9 and D3_O8. D2_O9 was selected as it was shown to cause the least distortion to the signal while D3_O8 produced a significant improvement to the appearance of the waveforms which may be useful when locating P1. The original responses are shown along with the filtered responses and the changes in P1 latency. As before a '+' indicates a shift to the right and '- 'denotes a shift to the left.
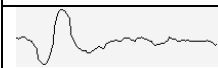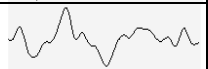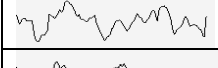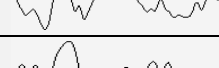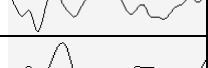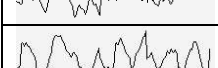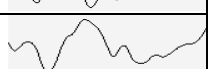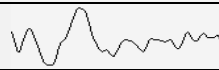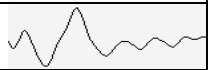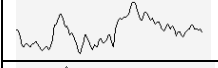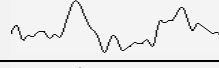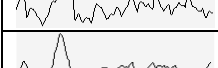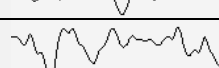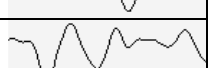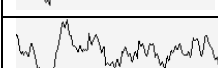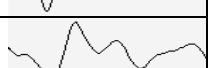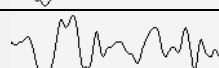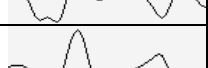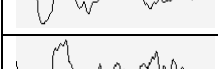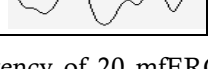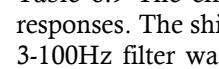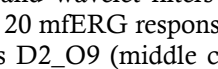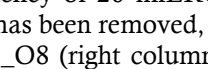
Wave	P1 latency (ms)	3-100Hz; D2_O9 Latency change (ms)	Filtered wave	3_100Hz, D 3_O8 Latency change (ms)	Filtered wave
	40	0		0	
	51	0		0	
	48	+1		0	
	40	0		0	
	40	0		0	
	38	0		0	
	54	0		0	
	45	0		0	
	42	0		-1	
	51	0		+1	
	47	+1		+1	
	41	+1		+1	
	40	0		0	
	51	0		0	
	39	0		0	
	41	0		0	
	37	0		0	
	45	0		0	
	43	0		0	
	40	0		+1	

Table 6.9 The effect of combining digital filters and wavelet filters on the P1 latency of 20 mfERG responses. The shift of the P1 latency is shown for 20 mfERG responses when drift has been removed, a 3-100Hz filter was applied and the wavelet filters D2_O9 (middle column) or D3_O8 (right column) were utilised. Those acquired using D3_O8 have the cleanest appearance however a greater number were shifted than when D2_O9 was applied.

As was seen when using only the wavelet filtering, fewer responses were distorted by D2_O9 than D3_O8; three responses were shifted to the right when using D2_O9 while five were affected by applying D3_O8. It can however be seen that those waveforms obtained when D3_O8 was chosen had a much less noisy appearance. It was observed that the P1 amplitude of the responses was affected by using each of the

filtering options. An average change of 3nV and 7nV was noted when using the bandpass filter in combination with the D2_O9 and D3_O8 wavelets respectively.

Different methods have been investigated to decrease the noise present in a trace array with a view to easing the interpretation of the test. It has been shown that the appearance of the mfERG waveforms can be significantly improved by combining filtering and wavelet techniques in addition to removing the baseline drift using spline fitting. It was therefore of interest to repeat the experiments conducted in sections 6.1 and 6.2 to establish if better results could be achieved when applied to less noisy signals.

6.7 SNR for 'response' or 'no response' classification using filtered signals

The ability of the SNR to distinguishing between a physiological response and no retinal function is discussed first.

6.7.1 Methods

The 1000 responses for which the SNR was calculated in section 6.1 were used. Baseline drift was removed in addition to filtering using a 3-100Hz bandpass filter and a D2_O9 wavelet filter. D3_O8 was not investigated as the cleaner signal appearance it provided was only thought to be advantageous when locating P1. As before, two methods were applied to calculate the SNR: that using a signal window and a time window; and that employing an unused sequence to represent the noise.

6.7.1.1 Noise window and signal window after signal processing (method 3)

The SNR was calculated using the method described in section 6.1.1.1. As before threshold values of 0, 1, 2, 3, 4 and 5dB were chosen to distinguish between 'response' and 'no response'; the value yielding the highest agreement with the expert was considered the optimum threshold value for stating if a waveform contained a physiological response.

6.7.1.2 Dead sequence after signal processing (method 4)

In addition to minimising the drift and filtering the individual mfERG responses, the baseline drift was also removed from the dead sequence. The SNR was then calculated using equation 3.3, again with a time window of 0-80ms. As before the best threshold value was defined by assessing the classification accuracy achieved using a number of different cut off values.

6.7.2 Results

6.7.2.1 Noise window and signal window after signal processing (method 3)

Table 6.10 shows the level of agreement between the expert and the classification defined by the SNR value for the different cut off values:

Cut off (dB)	0	**1**	2	3	4	5
% agreement with expert	81	**84**	82	80	73	65

Table 6.10 Percentage of waveforms correctly analysed as 'response' or 'no response' using the windowing SNR method after the application of signal processing techniques. The value defining the threshold between a physiological response and no significant function was altered.

This approach achieved an accuracy of 84% when using a threshold value of 1dB, with sensitivity and specificity values of 68% and 91% respectively. It is evident that removing the drift from the signal and filtering the waveforms improved the performance; an accuracy of 68% was achieved prior to applying signal processing techniques.

6.7.2.2 Dead sequence after signal processing (method 4)

The ability of the SNR to differentiate between a response and no significant function when the SNR was calculated using a dead sequence can be seen in table 6.11:

Cut off (dB)	0	1	2	**3**	4	5
% agreement with expert	76	80	84	**89**	87	83

Table 6.11 Percentage agreement between the expert and the dead sequence SNR (after the application of signal processing techniques) when categorising waveforms as 'response' or 'no response'. The performance for different SNR cut off values is shown.

89% of waveforms were given the same classification by both the expert and the computer when a threshold of 3dB was chosen. A sensitivity of 72% was found while the specificity was 95%. Again an increase in the ability of this technique to classify the waveforms into 'response' and 'no response' was achieved (78% using original signals) by applying filtering techniques and removing the drift. The sensitivity of this method did however remain relatively poor. Of those waveforms mistakenly identified as 'response', 40% had a SNR of 4dB; i.e. their SNR was very close to the threshold value. It was also noted that 17% of the false negative results were contaminated by large amounts of 50Hz noise.

In section 6.1.2 a trace array was utilised to demonstrate pictorially the performance achieved using the two SNR methods. The same trace array was selected to demonstrate the ability of methods 3 and 4 to categorise the signals. The classification of responses using the noise window and signal window SNR technique with a threshold of 1dB is shown in figure 6.23 (left), with pink depicting 'no response' and green a 'response'. The discrepancies between these classifications and those of the expert are also shown in figure 6.23 (right):
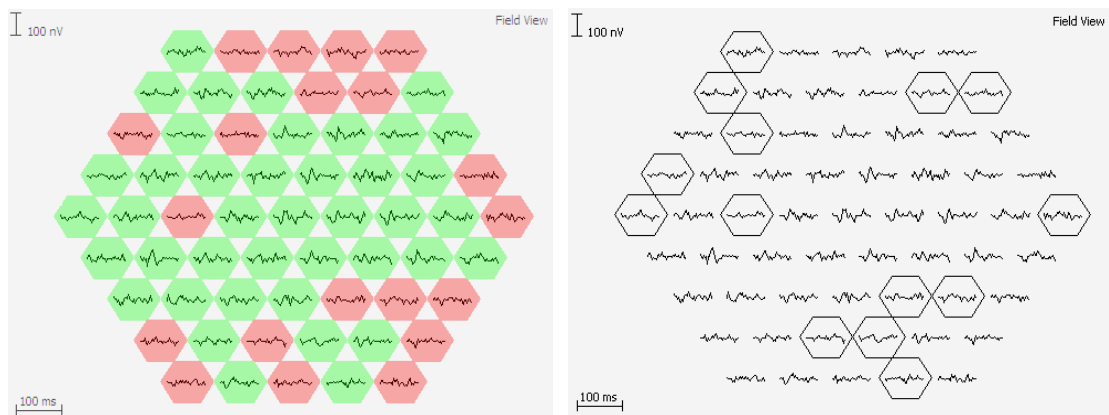
Figure 6.23 Classifications for trace array in figure 6.2 when the windowing SNR method (after the application of signal processing techniques) was utilised to categorise each waveform as 'response' or 'no response' (left). The classification differences between the expert and the SNR method are shown (right). Pink equates to 'no response' while green corresponds to 'response'. The right trace array highlights the discrepancies between the expert and the SNR approach.

14 differences of classification were observed between the expert and the SNR approach (23% of responses in the trace array). Finally the equivalent plots can be seen when using method 4 with a 3dB cut off:
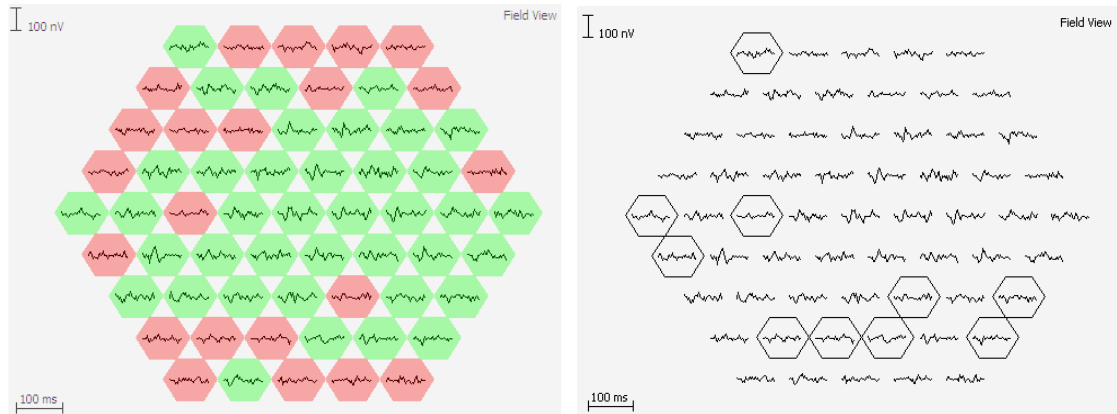
Figure 6.24 Classifications when the dead sequence SNR method (after the application of signal processing techniques) was used to categorise each waveform in figure 6.2 as 'response' or 'no response' (left). The classification differences between the expert and the SNR approach are shown (right). Pink equates to 'no response' while green corresponds to 'response'. The right trace array highlights the discrepancies between the expert and the SNR approach.

In this instance ten responses were classified incorrectly (16%). This represented the fewest errors of the four approaches studied, although the error rate remained relatively high. Of those ten waveforms classified incorrectly six had a SNR value very close to the threshold; three of those classified as 'no response' by the expert had a SNR of 4dB while three which were said to be 'response' by the expert had a SNR of 3dB.

The largest improvement in performance was seen for waveforms originally affected by baseline drift. In section 6.1.2 two responses, both representing areas of no retinal function were extracted from a trace array. Only one of these was affected by drift. It was shown that the SNR calculated for the waveform with drift was high using both methods (8dB and 15dB for methods 1 and 2 respectively) thus misstating that the waveform represented an area of functioning retina. The same trace array is shown below, after the removal of drift and the application of digital and wavelet filtering:
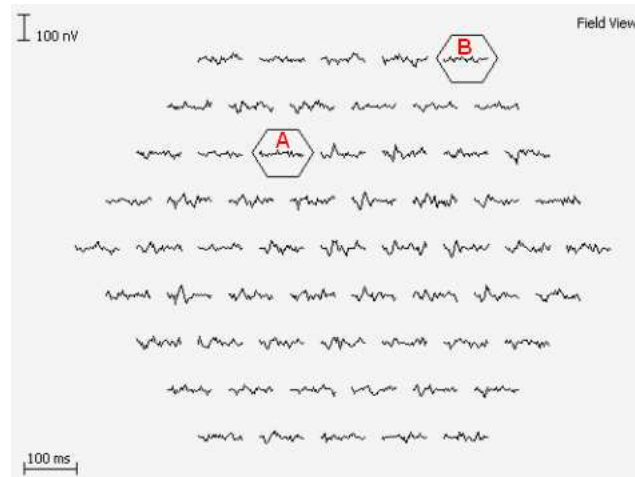
Figure 6.25 The trace array shown in figure 6.5 after the removal of drift; individual waveforms are highlighted. After the removal of drift both A and B were correctly identified as 'no response' by each SNR method.

In this instance the SNR values for A and B were 0dB and -3dB respectively when utilising method 3 and 0dB and -1dB respectively when using method 4. Each wave was therefore classified correctly as having no physiological response, which was not the case prior to applying the various signal processing techniques.

It has been demonstrated that removing the baseline drift from the mfERG responses and applying digital and wavelet filtering techniques improved the ability of the SNR to distinguish waves with a physiological response from those with no significant function. This was observed both when using the dead sequence method and the noise window and signal window approach. As was found when using the original signals, the dead sequence technique yielded superior results when compared with those achieved using the method proposed by Zhang *et al.*, with 89% of responses being correctly classified by the computer.

6.8 Spline fitting using filtered signals

This section examines the ability of the spline fitting technique to classify responses as 'delayed' or 'not delayed' and to locate the P1 component accurately after reducing the noise present in the mfERG responses.

6.8.1 'Delayed' or 'not delayed' classification

6.8.1.1 Methods

The 694 mfERG responses utilised in section 6.2 were used. The noise removal techniques employed when calculating the SNR value were applied; these included the removal of baseline drift using spline fitting, a Bessel filter with a passband frequency range of 3-100Hz and wavelet filtering using a decomposition level of two and an order of nine (D2_O9). In section 6.5 it was shown that the wavelet filter with a decomposition level of three and an order of eight (D3_O8) shifted a greater number of the responses but that the signals produced had a less noisy appearance. It was therefore decided to investigate this type of wavelet filter in addition to D2_O9 as the recovered responses may be easier to cursor when using spline fitting. Consequently two experiments were conducted with a view to finding the best possible solution:

1) drift removal; 3-100Hz Bessel filter; D2_O9 wavelet;

2) drift removal; 3-100Hz Bessel filter; D3_O8 wavelet.

For each of these experiments the filtered waveforms were analysed by the spline O5_R10 and were subsequently categorised as 'delayed' or 'not delayed' based on the P1 latency provided by the spline. This classification was compared with that of the expert for each of the 694 responses allowing the ability of the spline to analyse the responses using the two different filtering options to be assessed.

6.8.1.2 Results

The following table details the agreement between the expert and the spline when using the different filtering parameters:

	Baseline removal 3-100Hz Bessel filter	
	Wavelet D2_O9	**Wavelet D3_O8**
% agreement with expert	89	**90**

Table 6.12 Agreement between the expert and the spline for different filtering parameters when classifying responses as normal or abnormal based upon their P1 latency.

It is evident that the ability of the spline to categorise responses as 'delayed' or 'not delayed' was improved by reducing the noise present on the waveforms. It increased from 85% (when using the original data) to 90% when using D3_O8. Sensitivity and specificity values of 96% and 81% respectively were calculated. The agreement between the spline and the expert was slightly better when using the wavelet D3_O8 however there was very little to distinguish between the results. It was observed that of those waveforms which were misclassified 72% had a P1 value (as stated by the expert) within 1ms of the threshold between delayed and normal, while 94% were within 2ms of this timing boundary, implying that the majority of responses which were misclassified were close to boundary between delayed and normal.

6.8.2 Ability to locate P1 accurately

The ability of the spline to provide a reliable value for the P1 latency when applied to filtered mfERG waveforms is examined in section 6.8.2.

6.8.2.1 Methods

In the previous section the spline stated a P1 latency for each of the 694 waveforms, first when filtered using the wavelet D2_O9 and secondly when using the wavelet D3_O8 (for each experiment the baseline drift was removed and a 3-100Hz Bessel filter was applied). These values were exported to Excel where the timing difference between the spline and the expert was calculated for each response for the two filtering experiments. The median discrepancy was established for the set of mfERG responses, as was the maximum timing difference and the 95th percentile for the timing difference.

6.8.2.2 Results

P1 latency differences between the spline and the expert	Baseline removal 3-100Hz Bessel filter	
	Wavelet D2_O9	Wavelet D3_O8
Median (ms)	1	1
Maximum (ms)	24	22
95th percentile (ms)	5	5

Table 6.13 P1 latency differences between the spline and the expert for different filtering parameters. The median, the maximum and the 95th percentile for this timing difference are shown for the testing data set.

The ability of the spline to state the latency of P1 was improved by reducing the noise and the drift present on the responses. The median was decreased from 2ms to 1ms while the 95th percentile decreased from 15ms to 5ms using each of the filtering options. The maximum timing discrepancy between the expert and the spline was however considerable. The results achieved using the two different wavelet filters were similar. It was of interest to inspect the results more closely. As the results achieved using the different wavelets were similar, only those using the wavelet D3_O8 will be considered, as they were slightly superior. The following table depicts the percentage of responses analysed correctly to within a specified time period:

Timing difference (ms)	0	≤1	≤2	≤3	≤4	≤5	≤10	≤15	≤20
% of waveforms correctly analysed to within each time difference	17	52	78	88	93	96	98	99	99.4

Table 6.14 Examing the difference in P1 latency defined by the spline and the system after filtering the responses – the percentage of responses correctly analysed to within different time periods are shown. It can be seen that the expert and the spline stated the same P1 latency for 17% of the responses. They were within 1ms of each other for 52%, while for 78% of responses analysed they were within 2ms of one another etc.

17% of the responses were given the same P1 value by the expert and the spline. This is slightly better than that achieved when analysing the waveforms prior to the removal of drift and the application of filtering, when an agreement of 12% was found. Previously an error of greater than 10ms was made for 7% of the responses; this has been decreased to 2% by filtering the data. Similarly the percentage for which there was a disagreement of more than 20ms was decreased from 4% to 0.6% after the application of noise reduction techniques. It is evident that reducing the noise present on the waveforms and minimising the baseline drift improved the ability of the splines

to locate P1. To demonstrate one such waveform for which this was the case, the response selected in section 6.2.2.2 to highlight the problems resulting from baseline drift is used. The expert reported a P1 latency of 40ms for this response whereas the spline O5_R10 located P1 at 67ms prior to the application of noise removal techniques. The original signal, analysed by the expert can be in figure 6.26 (left). On the right, the P1 value as found by the spline after the application of baseline drift removal, digital filtering and the wavelet D3_O8 is shown:

Figure 6.26 Example 3: the location of P1 as stated by the expert (left), and the spline after filtering the response. The expert and spline both identified a P1 latency of 40ms. Prior to the removal of the baseline drift and filtering there was a discrepancy of 27ms between the expert and the spline.

The spline located P1 accurately; the error was therefore decreased from 27ms to 0ms by applying the various signal processing techniques. For a small number of waveforms gross mistakes were however made; the following figure demonstrates one such example. The original response, as classified by the expert is shown (left). On the right, the location of P1 as stated by the spline after the waveform was filtered using the Bessel filter, the wavelet D3_O8 and the baseline drift was removed is depicted:

Figure 6.27 Example 4: the location of P1 as stated by the expert (left) and the spline after filtering the response (right). In this instance there was a 22ms discrepancy between the spline and the expert when analysing this noisy mfERG response.

It can be seen that the P1 latency defined by the spline is 22ms later than that stated by the expert. Despite the reduction of the noise present on the waveform the later part of the response remained raised relative to the start thus the spline made a significant error when locating P1.

6.8.3 'Decreased' or 'not decreased' classification

When analysing the mfERG it is also important to establish if a response is decreased in amplitude or if it is within normal amplitude limits. This section therefore examines the possibility of using spline fitting to classify waveforms as 'decreased' or 'not decreased'. When using spline fitting both the N1 and P1 components are located by the spline (refer to section 6.2). The P1 amplitude is calculated by the difference in the y-values (on the mfERG waveform as opposed to the spline) at each of these data points. This can be seen in figure 6.28:
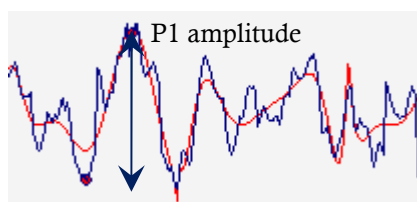
Figure 6.28 The method employed by the spline to calculate the P1 amplitude. The spline is shown in red while the mfERG response is blue. The difference in y-value (on the mfERG response) between N1 and P1 defines the P1 amplitude.

6.8.3.1 Methods

The expert measured the P1 amplitude for each of the 694 mfERG waveforms said to have a response. The location of the waveform in the trace array was noted, as it was shown in chapter 3 when establishing the mfERG normative range that the P1 amplitude of responses decreases with eccentricity. The responses were therefore classified as 'decreased' or 'not decreased' based on their P1 amplitude and their position in the trace array. The spline O5_R10 was then applied to each waveform, calculating its P1 amplitude using the method described above; each response was thus categorised as being decreased in amplitude or as being within normal limits based upon this value and its location in the trace array. This classification was compared with that of the expert for each of the 694 responses, allowing the ability of the spline to categorise the responses to be assessed. Different signal processing techniques were applied to the mfERG responses prior to analysing them with the spline to establish the optimal paramters; these included:

1) baseline drift removal;

2) baseline drift removal; 3-100Hz filter; D2_O9 wavelet;

3) baseline drift removal; 3-100Hz filter; D3_O8 wavelet.

6.8.3.2 Results

The following table details the agreement between the expert and the spline when using the different filtering parameters:

	Drift removal	Drift removal 3-100Hz filter D2_O9 wavelet	Drift removal 3-100Hz filter D3_O8 wavelet
% agreement with expert	**93**	91	90

Figure 6.29 Agreement between the expert and the spline for different filtering parameters when classifying responses as normal or abnormal in terms of their P1 amplitude.

The highest level of agreement was found when drift removal alone was used; the spline and the expert concurred for 93% of responses. This was decreased when filtering and wavelet techniques were applied. The sensitivity and specificity values for experiment 1 (drift removal only) were 86% and 97% respectively. Of those incorrectly classified by the spline, 80% had a P1 amplitude (as defined by the expert) within 5nV of the boundary between decreased and normal.

6.9 Discussion

The first aim of the experiments described in this chapter was to investigate the possibility of using the SNR value to discriminate mfERG waveforms containing a physiological response from those with no significant function. Two different approaches for calculating the SNR were studied and compared, with a view to achieving the optimal results. The first of these, the windowing method proposed by Zhang *et al.* (76), (method 1) has been used by various researchers when analysing clinical mfERG data (77-79), while the second technique, the dead sequence approach, (method 2) has thus far only been utilised to analyse normal mfERG responses (81). It was therefore of interest to establish if the latter technique could be

applied successfully to clinical data and if it would improve the ability of the SNR to classify waveforms.

When the ability of the two methods to classify a waveform as 'response' or 'no response' was compared it was found that method 2 produced superior results; 78% of the 1000 waveforms were categorised correctly as opposed to 68% when using method 1. The principal difference between the two methodologies was the data used to calculate the 'noise' component of the SNR; method 1 utilised the last 80ms of the mfERG waveform while method 2 employed a waveform recovered from an unused m-sequence. That used by method 2 can be assumed to comprise only noise (i.e. no physiological signal) as it was obtained from an m-sequence which had not stimulated the eye. A similar presumption cannot however be made for method 1 as the same mfERG waveform was used to calculate both the signal and the noise; the noise component may therefore contain some physiological response. The more representative nature of the 'noise' data when using method 2 is likely to explain why it achieved better results.

The accuracy of each method was however relatively low. By removing the baseline drift and filtering the waveforms the performance of each SNR method improved; accuracies of 84% and 89% were realised by the windowing and the dead sequence methods respectively. The improvement in performance for the dead sequence method was primarily due to a significant increase in the sensitivity; it changed from 47% to 72%, while the change in specificity was less, with an increase from 91% to 95%. In contrast both the sensitivity and the specificity improved by 16% for the windowing method; the sensitivity changed from 52% to 68% while the specificity increased from 75% to 91%. The elevated performance of each technique can predominantly be attributed to the removal of baseline drift from the data. This low frequency artefact can cause either the start or the end of the mfERG waveform to be raised, with each resulting in different problems when calculating the SNR.

If, for example the latter part of a true physiological response is elevated it can cause a deceptively low SNR to be calculated when using the windowing method; the value of the data points in the noise window are increased as a result of the upwards tilt causing the value of the 'noise' RMS to be larger than it should be. This decreases the SNR, potentially causing a waveform to be mistakenly classified as 'no response'. By removing the drift, the RMS of the noise window is not falsely inflated, thus increasing the SNR. This explains why the specificity of method 1 was increased by

removing the low frequency artefact. Equally, when a waveform containing no significant response is tilted so that the initial part of the wave is elevated, it can be misclassified as 'response', as the values in the signal window are misleadingly high. This causes the RMS value for the signal to be higher than it should be, potentially resulting in the SNR value being above the threshold value. Removing the tilt from the waveform results in a smaller, more representative value for the signal RMS, thus reducing the calculated SNR value. The sensitivity was thus improved by removing the baseline drift from the responses.

Only the sensitivity of method 2 was noticeably improved when the drift was eliminated. In the case of a waveform with no response, the RMS value of the 'signal' was deceptively high, both when the start or the end of the wave was raised, although this affect was more evident when the start of the wave was elevated. By removing the tilt the SNR was decreased to a more representative value, thus reducing the rate of misclassifying these waveforms as 'response'. When waveforms containing a physiological response were affected by baseline drift, again the RMS value of the signal window was increased; the resulting increase of the SNR value did not however cause a misclassification of the response. Consequently the specificity of method 2 was not significantly affected by the removal of the baseline drift.

Removing the baseline drift from the signals and calculating the SNR value using the dead sequence with a threshold of 3dB achieved the most accurate results, with an agreement of 89% between the expert and the system. However, misinterpreting a waveform with no significant function as 'response' remained a problem. On inspection it was found that 40% of these waveforms had an SNR value of 4dB. In other words it was very close to the threshold value. The user could potentially be warned that the grading of waveforms with an SNR value of 4dB may be incorrect to account for this problem. It was also noted that 17% of the responses misclassified as 'response' were contaminated by 50Hz noise. This was because the 50Hz signal was superimposed on the waveforms resulting in an increase in the value of the data points. The RMS value calculated for the 'signal' component of the SNR was therefore inflated, thus producing a falsely high SNR value. If this method were to be used in practice caution would have to be taken when analysing mfERG waveforms contaminated by 50Hz noise.

An accuracy of 89%, achieved using the dead sequence method after removing the baseline drift from the data, is relatively good; it is comparable with that achieved

using the ANN (90%) and is superior to that acquired when using the Fourier profile (64%). If this technique were to be used clinically it could not however be used in isolation to distinguish between 'response' and 'no response', as a higher success rate would ideally be required. It does however have potential and could be used as part of a multilayered approach to the analysis of the mfERG test.

The second aim addressed in this chapter was to investigate the viability of using spline fitting to classify responses as 'delayed' or 'not delayed' based upon their P1 latency. The spline found to yield the highest concurrence with the expert used a fifth order polynomial and a resolution of ten; 85% of responses were given the same classification by the expert and the spline. This spline provided a more generalised fit to the mfERG waveforms than those using higher ordered polynomials, which was particularly useful when P1 was not the maximum data point. As this is often the case when analysing mfERG data this spline produced the best results. A number of different noise reduction techniques were subsequently applied to the responses, after which the ability of this spline to classify responses as 'delayed' or 'not delayed' increased to 90%. This increase in performance can predominantly be attributed to the cleaner appearance of the waveforms, thus making it easier for the spline to locate P1. This is better than that achieved using the ANN in chapter 5, for which the agreement with the expert was 86%.

Of interest when investigating the different digital filtering parameters, was that those filters which attenuated frequencies greater than 60Hz, 70Hz and 80Hz, and less than 5Hz, caused distortion to the responses, leading to a bandpass filter with a passband of 3-100Hz being chosen. These findings were however in contrast to results presented by Seeliger *et al.* who used a bandpass filter with a frequency range of 9.4 -56.4Hz, reporting that it did not systematically distort the shape of the signals (63). This was ascertained by performing an inverse Fourier transform on the attenuated frequencies. No N1-P1-N2 complex was seen when these frequencies were viewed in the time domain, therefore it was assumed that the mfERG responses were not distorted. The timing of the filtered responses was not however examined.

It was also of interest that similar results were achieved when the spline was presented with data filtered by the two different wavelets. It was shown earlier in the chapter that D3_O8 produced responses with a much cleaner appearance than D2_O9, therefore it had been assumed that the spline would find it easier to locate P1 and hence yield more accurate results. However it was also found when studying 20 mfERG

waveforms that a greater number of the responses were shifted in time when using D3_O8, thus an error would have been introduced before applying the spline fitting techniques in many instances. This may explain why significantly better results were not achieved when using D3_O8.

It is useful to know if a waveform is delayed since delays are associated with dysfunction. However it is also beneficial to know the value of the P1 latency, for example when assessing the severity of the delay or when comparing responses from different visits to determine if there has been any change in function. The P1 latency values stated by the expert and the spline were therefore compared to establish the ability of this technique to locate P1 accurately. Their agreement was relatively low; 12%. This was increased slightly to 17% after the application of noise reduction techniques but this is still very low. It was however observed that the timing discrepancy was within 2ms for 78% of the responses which is more promising. A large timing difference was noted for a small percentage of responses; greater than 10ms for 2% of the waveforms and more than 20ms for 0.6% of the responses after being filtered. Given the poor agreement between the spline and the expert it would not be possible to rely on the exact latency stated by the spline. However, gross mistakes were only made for a small number of the responses and the timing difference was within 2ms for the majority of responses therefore it may be possible to sub-classify the delayed waveforms, providing the user with more information than simply 'delayed' or 'not delayed'. This will be investigated in the following chapter.

The third and final aim addressed in this chapter was to determine if spline fitting could be used to classify a response as 'decreased' or 'not decreased'. Promising results were achieved, with an agreement between the spline and the expert of 93%. 80% of those misclassified were within 5nV of the boundary between normal and decreased; thus large errors were only made for 1% of the 694 responses. It was noted that poorer results were achieved when the waveforms were filtered. It was however stated in section 6.6 that filtering affected the P1 amplitude of the responses, explaining why the classification accuracy of the spline was decreased when analysing filtered waveforms.

Unlike approaches such as template fitting, this curve fitting technique does not rely on the waveform having a similar shape to the original template; each response is treated individually thus widening the scope of the technique. It is evident that spline fitting has the potential to classify mfERG responses, both in terms of their P1 latency and their P1 amplitude but caution must be taken if stating actual P1 latency values.

6.10 Conclusions

It has been shown that the SNR has the potential to classify a mfERG waveform as 'response' or 'no response'. Superior results were attained when using the dead sequence method as opposed to that using a noise and a signal window. Optimal results were found when the baseline drift was removed prior to calculating the SNR, an accuracy of 89% being achieved. Promising results were obtained when spline fitting was used to classify responses as 'delayed' or 'not delayed'; the expert and the spline agreed for 85% of the waveforms. This was further improved to 90% by removing the drift from the responses and applying digital and wavelet filtering techniques. Finally, encouraging results were obtained when spline fitting was utilised to categorise responses as decreased or within normal amplitude limits based on their P1 amplitude. In this instance the expert and the spline concurred for 93% of the waveforms. Both the SNR and the spline fitting technique have a potentially important role to play in the interpretation of the mfERG. They will each be used as part of a multilayered approach to the analysis of this test. Chapter 7 combines each of the techniques discussed in chapters 4, 5 and 6 and assesses the ability of this final system to interpret mfERG data.

7 Development of a multilayered expert system

The principal aim of this thesis was to develop a means of reducing the subjectivity of the mfERG analysis process. The system should be capable of accurately analysing the individual cross correlated waveforms, reporting if they contain a physiological response or have no significant function. Furthermore it should state if a response is within normal P1 amplitude and latency values. An overview of the process which will be employed by the final system to classify the mfERG waveforms is shown in figure 7.1:
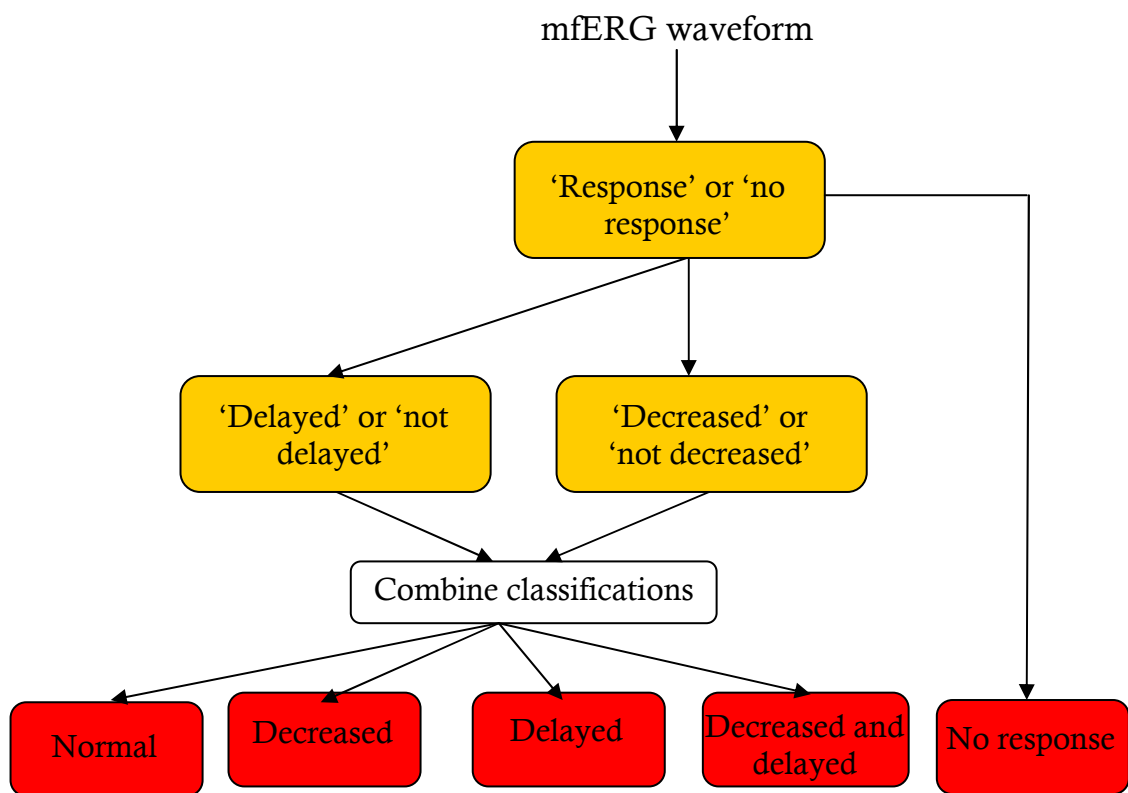
Figure 7.1 Overview of final system used to classify mfERG waveforms

Waveforms will first be classified as 'response' or 'no response', with those said to contain no physiological response classified so. All other waveforms will subsequently be categorised as 'delayed' or 'not delayed', and finally they will be classed as 'decreased' or 'not decreased' in amplitude. Responses will hence be categorised as normal, delayed, decreased, or decreased and delayed.

A number of methods have been investigated in the previous three chapters to perform each of these classifications, with each showing potential. However they could not be relied upon individually, as a higher performance was expected of the final system.

This chapter therefore investigates the possibility of combining the techniques studied in chapters 4, 5 and 6 with a view to increasing the accuracy of each classification step performed by the system.

7.1 'Response' or 'no response'

The distinction between 'response' and 'no response' is examined first. Three methods have been studied to distinguish a waveform containing a physiological response from one with no significant function. These were:

1) analysis of waveforms in the frequency domain in chapter 4;

2) the use of an artificial neural network (ANN) in chapter 5;

3) calculation of the signal to noise ratio (SNR) in chapter 6.

Each approach was tested on the same set of mfERG waveforms, which comprised 1000 responses taken from 100 different patient trace arrays. The classification provided by each technique was compared with that of an expert to assess its potential. Table 7.1 summarises the results achieved using each of these three techniques:

Technique applied	% agreement with expert
Fourier profile	64
ANN	90
SNR	89

7.1 Summary of performance achieved by each technique when distinguishing between 'response' and 'no response'. Results are shown when using the Fourier profile (chapter 4), an individual ANN (chapter 5) and the SNR (chapter 6).

When waveforms were studied in the frequency domain the agreement between the expert and the system was 64%, which is relatively low. Only 10% were however misclassified, with the remainder categorised as equivocal (refer to section 4.6).

The optimal ANN achieved an agreement with the expert of 90% when distinguishing waveforms with a physiological response (i.e. all of those classed as being delayed or within normal latency limits) from those with no retinal function. The sensitivity and

specificity values were 83% and 92% respectively. This particular network was a multilayer feed-forward ANN with 10 elements in its hidden layer, trained using the momentum learning rule with a momentum of 0.5. The clinical data were presented to it using incremental learning (refer to section 5.9). This network will be termed ANN 1 throughout the chapter.

A classification accuracy of 89% was seen when using the SNR to define a waveform as 'response' or 'no response' (sensitivity = 72%; specificity = 95%). This was obtained when the SNR was calculated with the dead sequence method; all waveforms with an SNR of 4dB of greater were classed as 'response' with the remainder labelled as 'no response'. Prior to calculating the SNR value the baseline drift was removed from both the waveform and the 'noise waveform' acquired by cross-correlating the uncorrelated data with an unused m-sequence; this was achieved by applying a spline with a second order polynomial. The waves were also filtered with a 3-100Hz Bessel bandpass filter and a wavelet (decomposition level of two and an order of nine) before calculating the SNR (refer to section 6.7). This approach will be termed SNR 1 for the remainder of the chapter.

Each technique, particularly the ANN and the SNR value, showed potential to make the distinction between 'response' and 'no response' however they could not be used in isolation, as a higher level of agreement was ideally required for the final system. Upon closer inspection of the data it was noted that for 96% of the 1000 mfERG waveforms at least one of the three techniques concurred with the expert; it was therefore decided to combine the techniques with a view to improving the performance of the system. An ANN was utilised for this purpose. The aim was to train the ANN to provide a final classification for a waveform based upon the classes stated for it by the Fourier approach, ANN 1 and SNR 1.

7.1.1 Methods

In chapter 5 it was stated that two data sets are required when designing ANNs: a training; and a testing set.

7.1.1.1 Training set

1500 mfERG waveforms were selected from 200 patient trace arrays. Each of these was analysed by an expert and grouped as 'response' or 'no response'. Subsequently they were:

1) classified as 'response' or 'no response' by ANN 1;

2) categorised as 'response', 'no response' or 'equivocal' based on their frequency profile;

3) classed as 'response' or 'no response' by SNR 1.

Three values were thus obtained for each of the 1500 mfERG waveforms. These were the data presented to the ANN during training, along with the desired answer (i.e. the classification previously defined by the expert).

For those networks trained using clinical data in chapter 5 it was necessary to use a validation set as there were an insufficient number of training examples relative to the size of the network. The number of elements in the input layer is considerably less in this instance (3 as opposed to 120); if, for example there were 15 elements in the hidden layer the network would comprise 60 weights (3*15 + 15*1). The ratio of training examples to weights would therefore be 25, thus it is not necessary to include a validation set.

7.1.1.2 Testing set

The 1000 waveforms used to assess each technique in isolation were utilised to test each network after training. This was to determine if agreement with the expert was improved by combining the techniques. The 1000 waveforms were classed using the Fourier profile approach, ANN 1 and SNR 1. The three values for each waveform, along with its classification as defined by the expert formed the testing set.

7.1.1.3 Training the ANN

Multilayer feed-forward networks with one hidden layer were chosen; all elements utilised a sigmoid transfer function. The learning rule and the method by which the data were presented during training were changed initially. Momentum, quickprop,

conjugate gradient and delta learning rules were investigated, using both batch and incremental learning when appropriate. The number of elements in the hidden layer was then varied from 2 to 20 in increments of 2, and finally the momentum rate was changed from 0.3 to 0.9 in steps of 0.1. Training was stopped and the network was tested after every 25 epochs (i.e. 25 presentations of the data to the network); this was continued until 1000 epochs. In the first instance each ANN was trained using 15 elements in the hidden layer while the different learning methods and learning rules were investigated.

7.1.2 Results

The following table details the ability of each ANN to classify the 1000 testing examples into 'response' or 'no response' when the learning rule and the learning method were changed. The number of epochs for which the best result was achieved is shown for each network trained:

Learning Rule	Type of training	Elements in hidden layer	Number of epochs	% agreement with expert
Momentum	Batch	15	75	91
Momentum	**Incremental**	**15**	**100**	**94**
Quickprop	Batch	15	100	90
Quickprop	Incremental	15	50	91
Conjugate gradient	Batch	15	25	91
Delta	Batch	15	25	90
Delta	Incremental	15	50	92

Table 7.2 Agreement between the expert and the combined ANN when differentiating between 'response' and 'no response': varying the learning rule and the mode of learning. Results are displayed for four different learning rules, in addition to incremental and batch learning.

The ANN trained using incremental learning and the momentum learning rule yielded the most promising result, with an agreement of 94% between itself and the expert; this was superior to that achieved by each of individual techniques. In each case the performance of the ANN stayed the same or fell when trained using a higher number of epochs than those presented in table 7.2.

Multi layer feed-forward ANNs, trained with the momentum learning rule using a momentum rate of 0.7 and one hidden layer were subsequently trained using different numbers of elements in the hidden layer. A momentum rate of 0.7 was utilised as this was the default when using the momentum learning rule. The results achieved are shown in table 7.3:

Elements in hidden layer	Number of epochs	% agreement with expert
2	50	89
4	75	89
6	75	92
8	75	91
10	100	90
12	125	94
14	**100**	**95**
16	125	93
18	125	94
20	100	92

Table 7.3 Agreement between the expert and the combined ANN when differentiating between 'response' and 'no response': varying the number of elements in the hidden layer from 2 to 20 in steps of 2, while keeping the learning rule and learning mode (momentum and incremental respectively) constant.

A high level of agreement between the expert and the ANN was achieved (95%) when 14 elements were present in the hidden layer. The following table demonstrates the effect of varying the momentum rate. In this instance multi layer feed-forward ANNs, trained with the momentum learning rule and one hidden layer comprising 14 elements were used. A momentum rate of 0.7 was utilised previously, therefore 0.3, 0.4, 0.5, 0.6, 0.8 and 0.9 were investigated:

Momentum	% agreement with expert
0.3	89
0.4	91
0.5	90
0.6	93
0.8	**94**
0.9	93

Table 7.4 Agreement between the expert and the combined ANN when differentiating between 'response' and 'no response': varying the momentum from 0.3 to 0.9 in increments of 0.1. The learning rule, learning mode and the number of hidden layer elements remained constant.

94% of the waveforms comprising the testing set were classified correctly when using a momentum rate of 0.8, however this was slightly lower than that achieved when using a momentum of 0.7. It was therefore established that the optimal ANN for this particular problem was a multi layer feed-forward network, trained with the momentum learning rule with a momentum of 0.7, and 14 sigmoid neurons in its hidden layer. The training data were presented to it 100 times during training using incremental learning. An agreement with the expert of 95% was realised using this ANN; the sensitivity and specificity values were 89% and 97% respectively thus misclassifying a waveform with no physiological response as a response was more problematic that the reverse situation. The ability of the system to classify a mfERG waveform as 'response' or 'no response' was thus improved by combining the outputs from the Fourier profile, ANN 1 and SNR 1. This is summarised in table 7.5:

Technique applied	% agreement with expert
Fourier profile	64
ANN 1	90
SNR 1	89
ANN 2 (combining above 3 techniques)	**95**

Table 7.5 Summary of performance achieved when distinguishing between 'response' and 'no response' using each individual technique, and a combined approach. Results are shown when using the Fourier profile, an individual ANN, the SNR and an ANN combining these three techniques.

The network successfully designed to combine these techniques will be referred to as ANN 2 for the remainder of the chapter.

7.2 'Delayed' or 'not delayed'

Section 7.2 explores the possibility of combining the techniques investigated in chapters 5 and 6 to classify a response as 'delayed' or 'not delayed'. Two approaches were studied to make this distinction. These were:

1) the use of an ANN in chapter 5;

2) the application of spline fitting, a curve fitting technique, in chapter 6.

Of the 1000 mfERG waveforms used to test the system's ability to distinguish retinal function from no significant response, the expert classified 694 as 'response'. Each of these was subsequently categorised as 'delayed' or 'not delayed' by the expert and these classifications were compared with those of the ANN 1 and the spline. Table 7.6 summarises the results achieved using each of these approaches:

Technique applied	% agreement with expert
ANN 1	86
spline	90

7.6 Summary of performance achieved by each technique when distinguishing between 'delayed' and 'not delayed'. Results are shown when using an individual ANN (chapter 5) and the spline fitting technique (chapter 6).

An agreement of 86% was achieved using ANN 1; the sensitivity and specificity were 89% and 83% respectively (refer to section 5.9).

When spline fitting was utilised the spline and the expert concurred for 90% of the responses analysed, with a sensitivity of 96% and a specificity of 81%. The optimal spline for classifying the response as delayed or within normal timing limits had a resolution of ten and used a fifth order polynomial. Prior to presenting the mfERG responses to the spline the baseline drift was removed from each waveform. A bandpass Bessel filter (3-100Hz) and a wavelet with a decomposition level of three and an order of eight were also utilised to minimise the noise present on the waveforms (refer to section 6.8.1). This method will be referred to as spline1 throughout the chapter.

The results obtained from each of these approaches were promising, however a higher performance was ideally required. As it was noted that at least one of the methods agreed with the expert for 94% of the responses it was decided to combine the output of the two approaches, with the aim of increasing the ability of the system to classify a response as 'delayed' or 'not delayed'. As in section 7.1 an ANN was designed; the objective was to train the network to provide a final classification for the waveform using the results provided by ANN 1 and spline1.

7.2.1 Methods

The data used to form the training and the testing data sets are described in the following section.

7.2.1.1 Training set

The same data set used in section 7.1, comprising 1500 mfERG waveforms, was utilised to train the network. For each of the waveforms said to represent an area of functioning retina the expert stated if the responses were delayed or within normal timing limits. These were subsequently analysed by ANN 1 and spline1, again categorising the responses as 'delayed' or 'not delayed'. These two classifications were presented to the network during training, in addition to the desired answer (i.e. the expert's assessment of whether or not a response was delayed). Again a validation set was not required as there were a sufficient number of training examples when compared with the size of the network.

7.2.1.2 Testing set

In order to compare the combined system with the individual techniques the same testing set as was used in chapters 5 and 6, composed of 1000 mfERG waveforms, was selected. Each of the waveforms reported as 'response' by the expert was classified by ANN 1 and spline1; these outputs were subsequently presented to the new network after each training session to assess the performance of each ANN. The new ANN's output was compared with the expert's answer.

7.2.1.3 Training the ANN

As before, multilayer feed-forward networks, with one hidden layer comprising sigmoid elements were trained. The network parameters were varied until the optimal performance was achieved. This was done using the same methodology as that utilised in section 7.1.1.3. Initial networks were trained with 15 elements in their hidden layer while the four different learning rules and two learning methods were studied.

7.2.2 Results

Table 7.7 details the performance of the ANNs trained using different learning rules when presented with the testing set. As before the number of epochs at which each individual network's optimal performance was achieved is shown:

Learning Rule	Type of training	Elements in hidden layer	Number of epochs	% agreement with expert
Momentum	Batch	15	200	90
Momentum	Incremental	15	100	89
Quickprop	Batch	15	75	90
Quickprop	**Incremental**	**15**	**150**	**92**
Conjugate gradient	Batch	15	250	91
Delta	Batch	15	100	89
Delta	Incremental	15	200	91

Table 7.7 Agreement between the expert and the combined ANN when distinguishing between responses which are delayed or not delayed: varying the learning rule and the mode of learning. Results are displayed for four different learning rules, in addition to incremental and batch learning.

The highest agreement with the expert (92%) was achieved when the network was trained using the quickprop learning rule and when the network weights were updated after the presentation of each training example (i.e. incremental learning). The number of elements in the hidden layer was then varied with each network trained using the quickprop learning rule and incremental learning. A momentum of 0.5 was initially chosen for each ANN as this was the default value for the quickprop learning rule. The ability of each network to classify responses as 'delayed' or 'not delayed' can be seen in table 7.8:

Elements in hidden layer	Number of epochs	% agreement with expert
2	50	91
4	125	89
6	150	92
8	250	92
10	150	92
12	**200**	**93**
14	175	92
16	200	90
18	200	92
20	225	91

Table 7.8 Agreement between the expert and combined ANN when differentiating between 'delayed' and 'not delayed': varying the number of elements in the hidden layer from 2 to 20 in increments of 2. The learning rule and learning mode (quickprop and incremental respectively) were kept constant.

The optimal network was that with 12 elements in its hidden layer, with an agreement of 93% between it and the expert. Finally the momentum was changed and the impact on the ability of the network to train was studied. The results are shown in table 7.9:

Momentum	% agreement with expert
0.3	92
0.4	92
0.6	**93**
0.7	92
0.8	90
0.9	91

Table 7.9 Agreement between the expert and the combined ANN when differentiating delayed responses from those within normal limits: momentum adjusted from 0.3 to 0.9 in steps of 0.1 The learning rule, the learning mode and the number of hidden layer elements remained constant.

93% of the testing examples were classified correctly by the ANN trained with a momentum of 0.6. The performance of the network using a momentum of 0.5 was however slightly higher than that trained with a momentum of 0.6; 646 responses were correctly classified (93.1%) as opposed to 643 (92.7%). It was therefore established that for this particular problem the optimal network was one trained using the quickprop learning algorithm with a momentum of 0.5 and 12 elements in its hidden layer. The

training data were presented to it 200 times during the learning process, using incremental learning. The sensitivity and specificity values achieved by this network were 96% and 90% respectively therefore classifying a response as 'delayed' when its P1 latency was less than 43ms was the principal source of error. Of those incorrectly classified 74% had a P1 latency (as defined by the expert) within 1ms of the boundary between normal and abnormal while 90% were within 2ms of this threshold. It has thus been demonstrated that the ability of the system to classify mfERG responses as 'delayed' or 'not delayed' has been improved by training a network to analyse the responses when presented with the outputs from ANN 1 and spline1. This can be seen in table 7.10:

Technique applied	% agreement with expert
ANN 1	86
spline 1	90
ANN 3 (combining above 2 techniques)	**93**

Table 7.10 Summary of performance achieved by each individual technique and combined approach when distinguishing between 'delayed' and 'not delayed'. Results are shown when using an individual ANN, the spline fitting technique, and an ANN combining these two techniques.

The network which yielded the highest accuracy will be termed ANN 3 in all subsequent sections.

7.3 'Decreased' or 'not decreased'

In chapter 6 (section 6.8.3) it was shown that the spline fitting technique was capable of correctly classifying a waveform as 'decreased' in amplitude or 'not decreased' for 93% of the 1000 responses tested; the baseline drift was removed from each of these responses prior to analysis with the spline. The spline had a resolution of ten and utilised a fifth order polynomial. Of those waveforms incorrectly classified, 80% had a P1 amplitude (as stated by the expert) within 5nV of the boundary between normal and abnormal. Clear mistakes were therefore only made for 1% of the responses tested. Consequently it was concluded that this technique had the potential to be used successfully as part of the multilayered system. This method will be referred to as spline2 for the remainder of the chapter.

7.4 Testing the multilayered system

It has been shown that the performance of the system has been improved by combining the different techniques investigated thus far in this thesis; when tested on 1000 waveforms the system classified 95% of waveforms correctly as 'response' or 'no response'. Of those said to have a response 93% were categorised accurately, both in terms of their latency and their amplitude. The individual approaches and the multilayered system were each tested on 1000 individual mfERG waveforms. However in clinical practice the system would be utilised to analyse trace arrays. It was therefore important to study the performance of the final, combined system when presented with trace arrays of varying recording qualities and with different retinal conditions.

7.4.1 Methods

20 trace arrays, representing a wide range of retinal function were selected. These were chosen from the set of recordings used to test the system's ability to grade recording quality in chapter 4. The group of 20 mfERG tests comprised seven 'excellent', six 'moderate', six 'noisy' and one 'unreportable' recording, as classified by the three experts. The system concurred with the experts for each of the 'excellent' and 'moderate' recordings and for four of the 'noisy' tests. The remaining two 'noisy' recordings were categorised as 'moderate' by the system; one of the three experts did however class each of these as 'noisy'. Finally, a mfERG test which the experts thought should not be reported was included, as the system classed it as 'noisy'. If using this multilayered system the test would have been analysed, therefore it was of interest to assess the ability of the system to report a very noisy recording.

One expert analysed each of the 1220 waveforms within the 20 trace arrays. These were initially classified as 'response' or 'no response'; for those said to represent an area of functioning retina both the P1 latency and amplitude were measured. The location of the response in the trace array was also noted as the normal range for the P1 amplitude decreases with eccentricity in the trace array. Responses were therefore defined as normal or abnormal both in terms of their amplitude and latency; the normative data presented in chapter 3 (section 3.4.2) defined these limits. The 1220 waveforms were subsequently presented to the multilayered system and analysed. Its

assessment of the waveforms was compared with that of the expert in each case. An overview of the multilayered system is shown in figure 7.2:
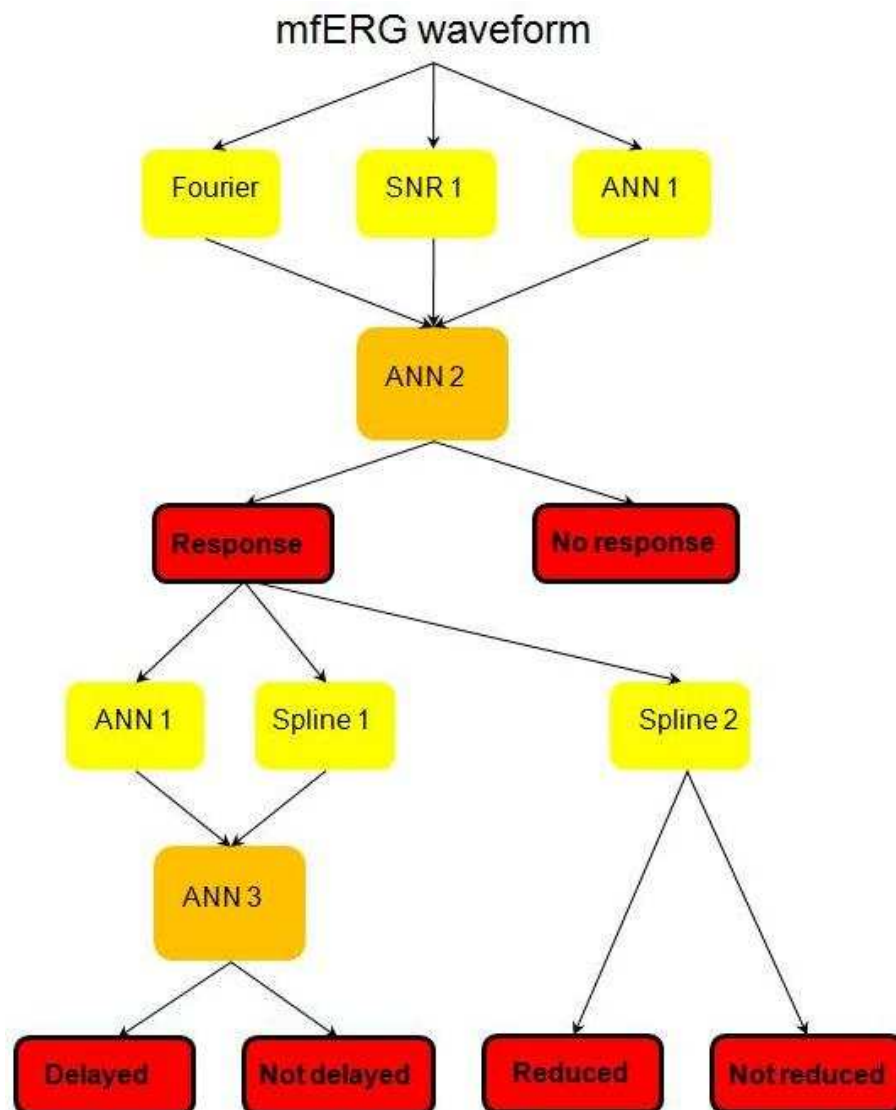
Figure 7.2 An overview of the multilayered system.

Each waveform was initially analysed using the Fourier profile, SNR 1 and ANN 1. The three classifications provided by these techniques were presented to ANN 2 which stated whether or not a waveform represented an area of functioning retina. Those defined as 'response' by ANN 2 were categorised as 'delayed' or 'not delayed' by ANN 1 and spline1. The output of each of these was then presented to ANN 3 which, based on these values, stated if a response was delayed or within normal timing limits. The P1 amplitude of a response was subsequently defined as decreased or normal using spline2, accounting for the location of the response in the trace array. Results have been grouped in terms of their quality grading to establish if the performance of the multilayered system is affected by the quality of a mfERG recording.

An initial objective for the system was to state the latency of P1 accurately. It was however shown in chapter 6 that when using the spline fitting technique the agreement between the spline and the expert was very low; 17%. The system could not therefore be relied upon to report an exact latency for P1. The discrepancy between the spline and the expert was however within 2ms for 78% of the responses analysed. It was therefore decided to investigate the possibility of further sub-classifying the responses in terms of their latency. The waveforms classified as delayed by ANN 3 were grouped into different time periods based on the P1 latency stated by spline1. The agreement between the expert and the spline was investigated when the delayed responses were classed into either two or three different time periods; these were:

1) 43-49ms;   >49ms   when two classifications;

2) 43-46ms;   47-50ms;   >50ms,   when three classifications.

Four trace arrays have then been utilised to show the performance of the multilayered system.

7.4.2 Results

The following table details the ability of the multilayered system to classify the mfERG waveforms, first as 'response' or 'no response' and then as normal or abnormal in terms of their amplitude and latency. When comparing the amplitude and latency classifications of the expert and the system only waveforms classified as 'response' by each were utilised.

	Recording quality			
	Excellent	Moderate	Noisy	Unreportable
% agreement response/no response	97	95	90	90
% agreement decreased/not decreased	96	96	89	82
% agreement delayed/not delayed	94	94	89	86

Table 7.11 The agreement between the multilayered system and the expert when classifying mfERG tests of different recording qualities, ranging from excellent to unreportable. The ability of the system to classify waveforms as response or no response, and normal or abnormal based on their P1 latency and amplitude is presented for the four different categories of recording quality.

The system performed well when the recording quality was 'excellent', with an accuracy of 97% when classifying the waveforms as 'response' or 'no response'; this fell to 90% when the system was presented with a mfERG test which the experts thought should not be reported. It was observed that 69% of those waveforms mislabelled as 'response' were from 'noisy' recordings. When analysing the trace arrays the expert stated that 8% of the waveforms were very difficult to categorise as 'response' or 'no response'. Of those waveforms incorrectly identified as having no significant physiological function 68% had been labelled as challenging to analyse by the expert. The expert had only experienced difficulties categorising 6% of responses misclassified as 'response'.

When stating if the amplitude and latency values were within normal limits, the highest level of agreement between the system and the expert was attained when the recording quality was 'excellent'. Accuracies of 96% and 94% were achieved for the amplitude and latency classifications respectively. These were decreased to 82% and 86% when the system was presented with the mfERG test thought to be too noisy to report by the experts (i.e. that classified as unreportable).

When the system was utilised to classify the delayed responses into one of three time periods an accuracy of 80% was obtained; this was increased to 89% when the number of categories was decreased to two. Despite the increased information provided when using three classifications an error rate of 20% was deemed to be unacceptable. Ideally an accuracy of greater than 89% would have been achieved when splitting the delayed waveforms into one of two categories however it was decided that the extra detail provided by having two latency categories for delayed responses justified its introduction to the multilayered system if appropriate. Again the classification accuracy was related to the quality of the recording; the agreement with the expert when allocating the delayed responses to one of two categories was 91%, 89%, 86% and 86% when the integrity of the mfERG test was 'excellent', 'moderate', 'noisy' and 'unreportable', respectively.

It is evident that the proficiency of the multilayered system is affected by the quality of a recording - for those recordings classified as 'excellent', superior results were attained. Four examples have been selected to compare the assessment of an expert and the multilayered system when presented with a trace array. The first was classified as 'excellent', the second as 'moderate', the third as 'noisy' and finally the fourth was said to be 'unreportable' by the experts, but was classified as 'noisy' by the system. In

each instance one expert analysed the waveforms in the trace array; these classifications are shown initially, followed by the output of the system. Finally the discrepancies between the expert and the system have been highlighted. To simplify the results, response latencies have been classified as 'delayed' or 'not delayed' in the first instance. If it was considered important to sub-classify the delayed responses this was also included.

Example 1

Waveforms classed by the expert as 'no response' are shown in purple while responses which were normal are green. Waveforms have been shaded yellow when decreased in amplitude, orange when delayed and pink when decreased and delayed:

Figure 7.3 The expert's analysis of an excellent recording. Purple depicts no response; pink is decreased and delayed; orange is delayed; yellow is decreased; green is normal. Diffuse delays were evident while there were no significant responses superiorly.

It can be seen that no significant responses were recovered in the superior region. Of those waveforms classified as 'response' diffuse delays were noted, with those shown in pink also decreased in amplitude. When this trace array was presented to the multilayered system the following results were obtained (left); the classification discrepancies between the expert and the system can be seen on the right, with each difference highlighted:
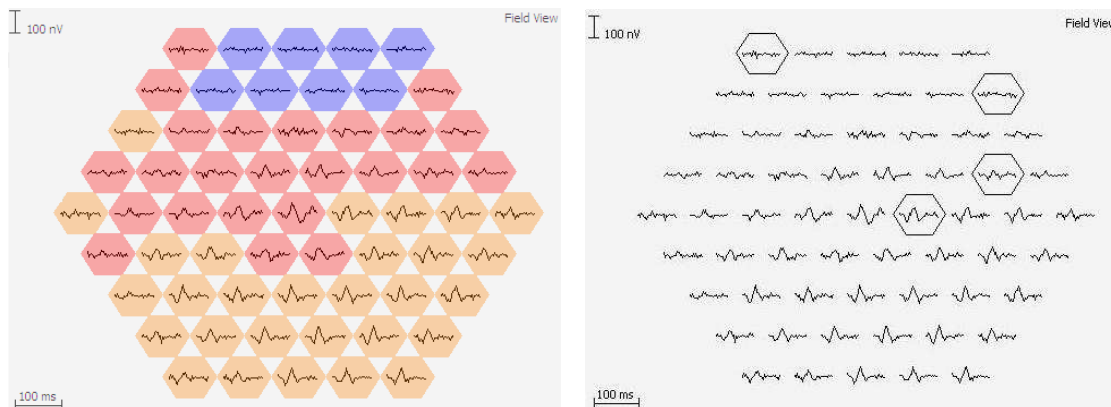
Figure 7.4 The system's analysis of the excellent recording shown in figure 7.3 (left), and the discrepancies between the expert and the system (right). Purple is no response, pink is decreased and delayed, while orange is delayed. The trace array on the right highlights the four classification discrepancies between the expert and the system.

In this instance four of the waveforms in the trace array were misclassified by the system, equating to an overall accuracy of 93%. Two waveforms were classed as decreased and delayed by the system when they were said to contain no significant physiological response by the expert. Of those correctly identified as 'response' ANN 3 classified each waveform as 'delayed'; this was in agreement with the expert. The amplitude was incorrectly categorised for two waveforms; it was noted that the P1 amplitude was within 2nV of the boundary between normal and abnormal for each of these. It was seen in figure 7.3 that there were diffuse delays in this recording however no information regarding the severity of these delays was provided.

The following trace array demonstrates the latency of the responses in figure 7.3 when classified into three categories: normal (<43ms); moderate delays (43-49ms); and severe delays (>49ms). These have been depicted in green, orange and pink respectively. Waveforms classed as 'no response' are shown in purple:
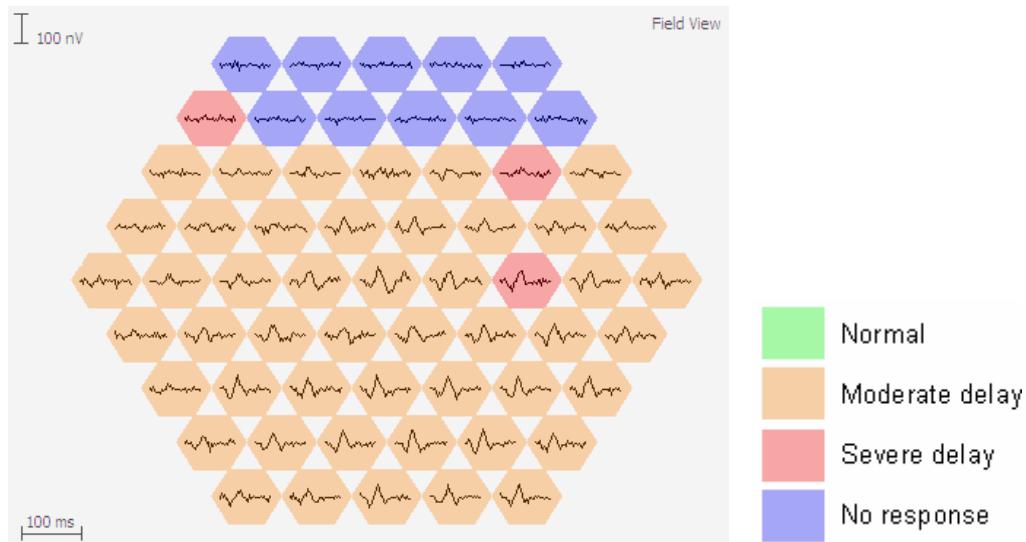
Figure 7.5 The expert's analysis of the excellent recording in figure 7.3 when sub-classifying the response latencies as normal, a moderate delay or a severe delay. Purple depicts no significant response, pink is a severe delay (P1>49ms), orange is a moderate delay (P1=43-49ms) while green is normal latency (P1<43ms). Moderate delays were seen for the majority of the responses.

It is evident that the majority of responses are moderately delayed in this case. When the system categorised the responses based on their P1 latency the following results were obtained:
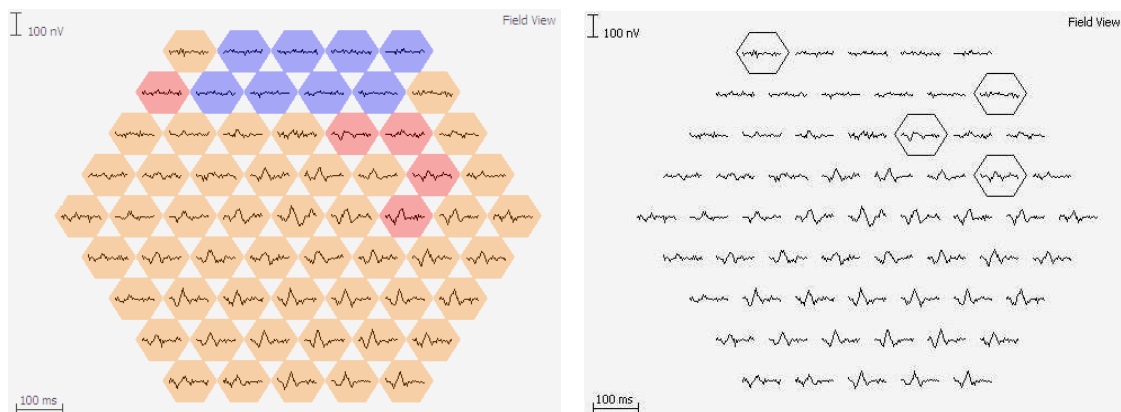
Figure 7.6 The system's analysis when sub-classifying the response latencies of the excellent recording shown in figure 7.3 (left). The discrepancies between the expert and the system are also displayed (right). Purple defines no response, pink is a severe delay (P1>49ms) and orange is a moderate delay (P1=43-49ms). The trace array on the right highlights the four classification discrepancies between the expert and the system.

Again the two responses mistakenly identified as 'response' can be seen. Of those correctly categorised by ANN 2 as representing an area of functioning retina two were said to be severely delayed when the expert classified them as being moderately delayed. The P1 latency of each of these was 49ms which is within 1ms of the threshold between a severe and moderate delay using this grading system. It is evident that the multilayered system performed well when presented with this patient trace array and that the waveforms on which mistakes were made had amplitude and

latency values very close to the boundary between normal and abnormal or a moderate and a severe abnormality.

Example 2

The following example demonstrates the ability of the system to analyse a recording graded as being of a moderate quality. Figure 7.7 depicts the classifications as defined by the expert:
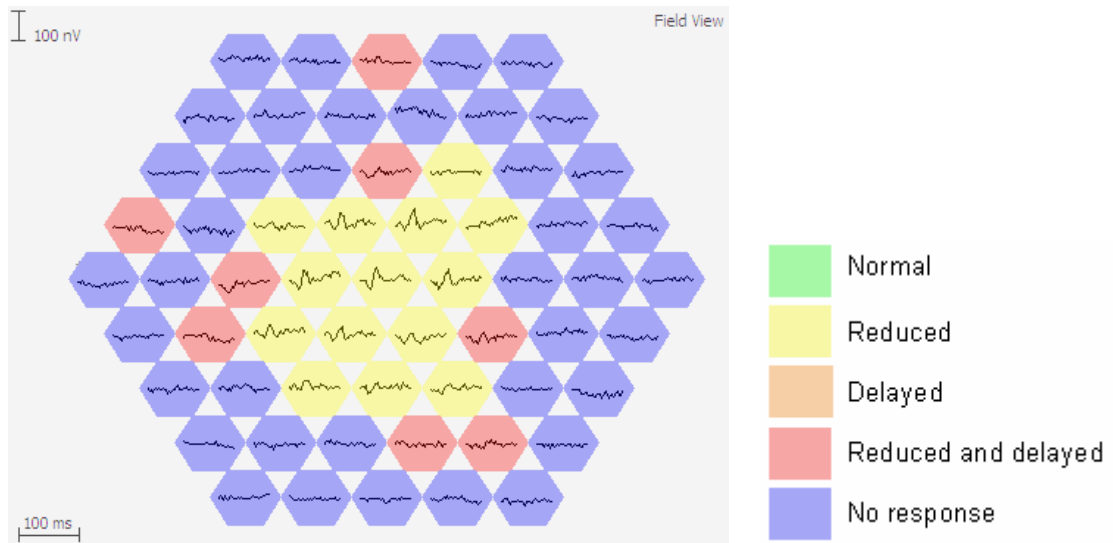
Figure 7.7 The expert's analysis of a moderate recording. Purple represents no response, pink is decreased and delayed, orange is delayed, yellow is decreased, and green is normal. Peripheral waveforms contained no significant physiological response while central responses were decreased in amplitude.

It can be seen that no significant responses were obtained from the peripheral regions while those waveforms located centrally were decreased in amplitude. A number of decreased and delayed responses were also observed. When analysed by the multilayered system the following results were obtained (left); again the discrepancies are shown (right):
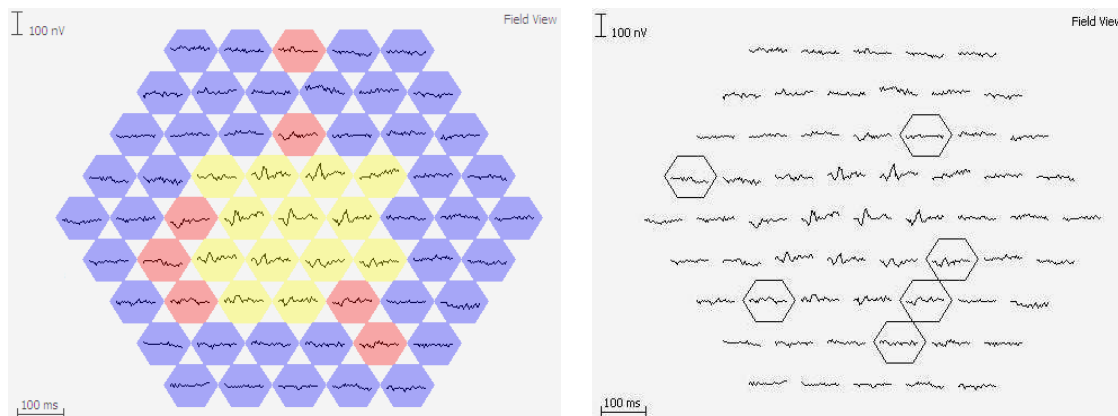
Figure 7.8 The system's analysis of the moderate recording shown in figure 7.7 (left), and the discrepancies between the expert and the system (right). Purple is no response, pink is decreased and delayed, while yellow is decreased. The trace array on the right highlights the six classification discrepancies between the expert and the system.

In this instance 55 of the 61 waveforms were correctly analysed by the system, corresponding to an agreement of 90%. Three waveforms were incorrectly identified by ANN 2 as 'no response' while one was mistakenly said to be 'response'; the expert had stated that two of these were very difficult to categorise as either 'response' or 'no response'. Spline 2 classified each of the responses as being decreased in amplitude; the expert agreed with this in each case. Two mistakes were however made by ANN 3 when distinguishing responses which were delayed from those within normal timing limits; one was incorrectly said to be delayed while the other was mislabelled as 'not delayed'. The former had a P1 latency (as stated by the expert) of 42ms while the latter had a latency of 43ms; this value was within 1ms of the boundary between normal and abnormal in each instance. In this case the central responses were within normal timing limits while no significant response were obtained from the peripheral areas, therefore further classification of the delayed responses was not essential.

Example 3

Example 3 illustrates the ability of the multilayered system to interpret a mfERG test categorised as a 'noisy' recording. The expert's analysis of the noisy mfERG recording can be seen in figure 7.9:
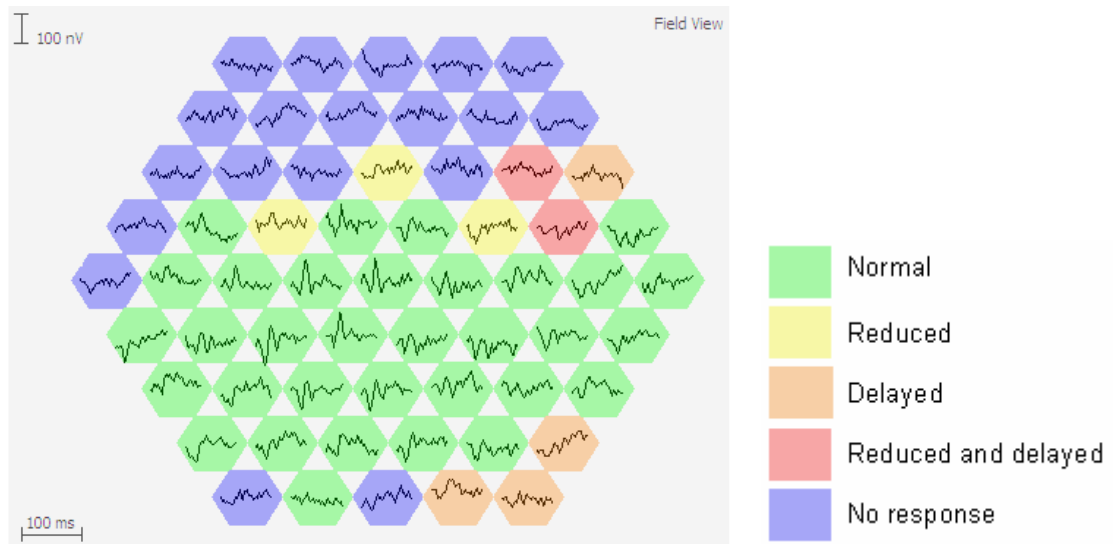
Figure 7.9 The expert's analysis of a noisy recording. Purple shows no response, pink is decreased and delayed, orange is delayed, yellow is decreased and green is normal. In general the inferior responses were normal while no significant function was observed superiorly.

No significant responses were recovered from the superior areas of the outer/mid retina (field view) while the majority of responses in the lower half of the trace array were within normal limits. When presented to the multilayered system the following results were obtained:
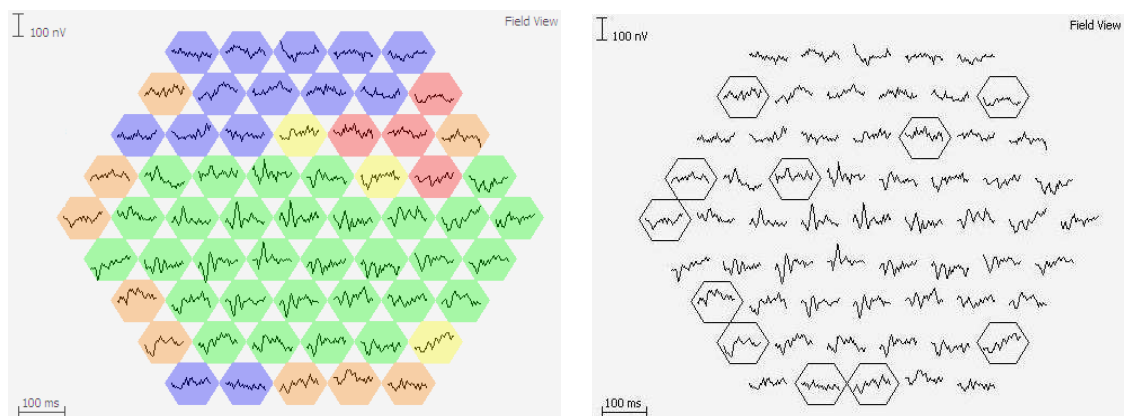
Figure 7.10 The system's analysis of the noisy recording shown in figure 7.9 (left), and the discrepancies between the expert and the system (right). The following colour coding system was utilised: purple is no response; pink is decreased and delayed; orange is delayed; yellow is decreased; green is normal. The trace array on the right shows the eleven classification differences between the expert and the system.

A greater number of discrepancies were noted in this trace array; eleven, as opposed to four and six for examples 1 and 2 respectively. One waveform was incorrectly said to have no significant response while five waveforms were mistakenly categorised as 'response' by ANN 2. The expert and the system disagreed when categorising three of the responses based on their P1 latency; each of these had a latency value within 2ms of the threshold between normal and abnormal and were difficult waveforms to analyse. Spline2 made two mistakes when categorising responses based on their P1

amplitude; in each instance the amplitude as stated by the expert was more than 5nV from the boundary between normal and abnormal.

Example 4

Finally, a recording said to be 'unreportable' by the experts in chapter 4, but classed as 'noisy' by the system is presented. An expert's assessment of the trace array can be seen in figure 7.11:
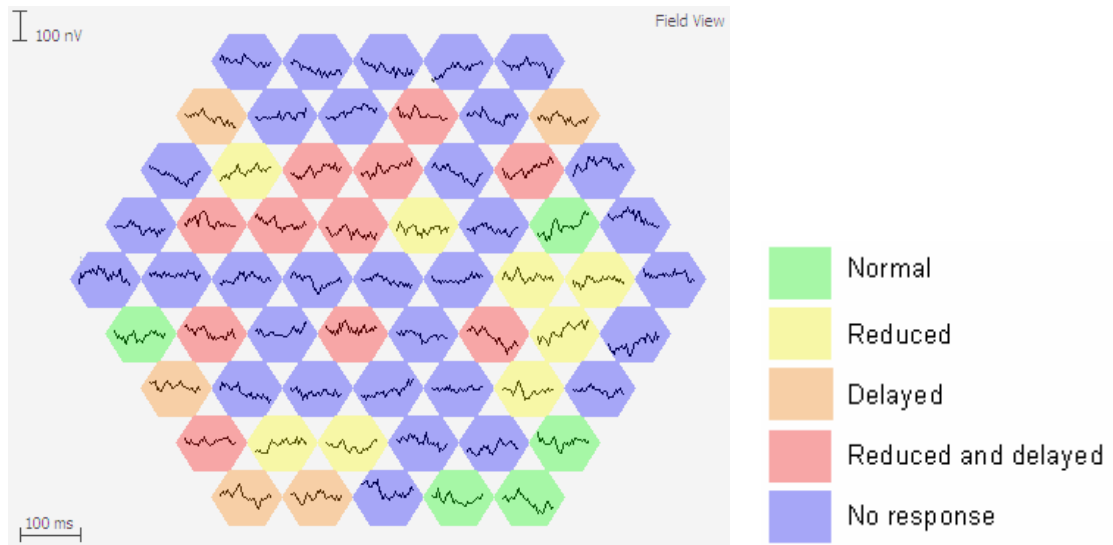
Figure 7.11 The expert's analysis of the recording considered to be unreportable. Purple depicts no response, pink is decreased and delayed, orange is delayed, yellow is decreased, while green is normal.

Figure 7.12 depicts the multilayered system's interpretation of the test (left) and the discrepancies between it and the expert (right):
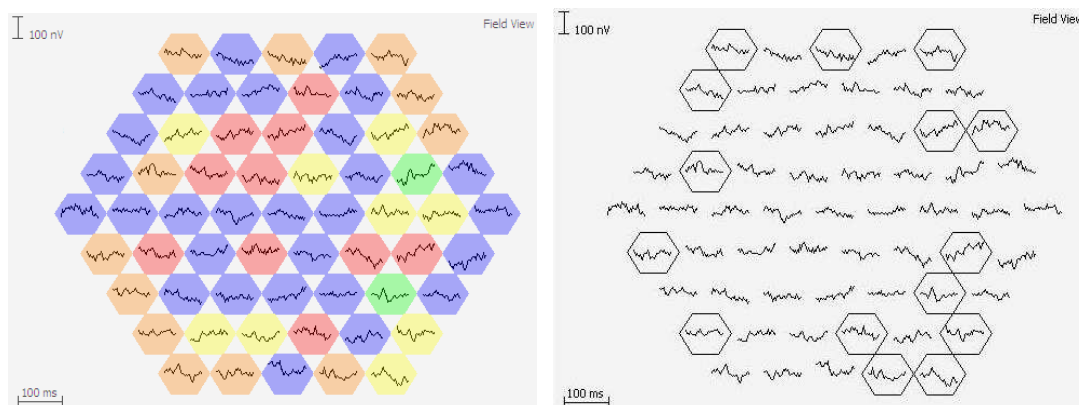
Figure 7.12 The system's analysis of the unreportable recording shown in figure 7.11 (left), and the discrepancies between the expert and the system (right). The following colour coding system has been employed: purple is no response; pink is decreased and delayed; orange is delayed; yellow is decreased in amplitude; and green is normal. The trace array on the right demonstrates the fifteen classification discrepancies between the expert and the system. This is the poorest performance of the four examples.

15 of the waveforms were incorrectly analysed by the system in this example. This corresponds to a classification accuracy of 75% which is relatively poor. The expert and the system disagreed for six of the waveforms when differentiating between 'response' and 'no response'; five of these were mistakenly identified as 'response' while one was said to have no significant function. The latter was said to be difficult to analyse by the expert. When distinguishing an abnormal from a normal amplitude five mistakes were made while four discrepancies were seen when classifying the latency of the responses as 'delayed' or 'not delayed'. Of those incorrectly classified by ANN 3, two had a P1 latency within 1ms of the boundary between normal and abnormal while the remainder were within 3ms.

It is evident that the multilayered system performed best when presented with high quality recordings and that its efficacy fell when the integrity of the recording was decreased.

7.5 Discussion

The principal aim of this study was to develop an objective method for analysing mfERG data. A number of approaches have been explored in the previous three chapters to realise this; each showed potential however a higher accuracy was ideally required than that achieved by each of the individual methodologies. It was therefore decided to combine the techniques using an ANN with a view to yielding a higher performance from the system. This approach was taken as at least one of the methods agreed with the expert for 96% of the examples classified as 'response' or 'no response' while one technique was correct for 94% of the responses being categorised as 'delayed' or 'not delayed'.

ANN 2, (the network trained using the outputs from the original network ANN 1, the frequency profile and SNR 1) accurately classified 95% of the waveforms presented to it as 'response' or 'no response'. This was superior to that achieved by each technique in isolation. Given that for 4% of the 1000 waveforms none of the individual techniques agreed with the expert, a performance of 95% was commendably successful. The sensitivity and specificity values were 89% and 97% respectively. Stating that an area of retina with no significant function contained a physiological response was therefore the most common source of error.

It was noted that when 20 trace arrays were subsequently examined 69% of those waveforms misclassified as 'response' were obtained from recordings classified as 'noisy'. In chapter 6 it was found that of the 1000 waveforms classified using the SNR method (i.e. one of the inputs into ANN2), 17% of those mistakenly identified as an area of functioning retina were contaminated by 50Hz noise; in general, when the uncorrelated data of the recordings were studied in the frequency domain, the magnitude of the peak at 50Hz was greater in recordings graded as 'noisy' than in those said to be 'excellent'. This may therefore explain why 69% of those waveforms mistakenly said to be a physiological response were recovered from poor quality recordings. The percentage of these waveforms said to be difficult to analyse by the expert was very small. In contrast, the expert experienced difficulties classifying 68% of the waveforms mistakenly identified as 'no response', implying that this distinction was of a more subjective nature.

The multilayered system and the expert concurred for 93% of the testing examples when categorising them as within normal timing limits or delayed, representing an increase in performance when compared to that of the two individual techniques. Again this was a relatively good performance as both spline1 and ANN 1 were incorrect for 6% of the responses presented to them. The sensitivity of ANN 3 was 96% while the specificity was 90% meaning that classifying a response as delayed when it had a P1 latency of less than 43ms was the network's main mistake. Both ANN 1 and spline1 had a lower specificity than sensitivity therefore it was unsurprising that a network training using their combined outputs had a lower specificity. It was important to identify if those responses misclassified had a P1 latency close to the boundary between normal and abnormal or if large errors had been made by ANN 3. It was established that of the responses incorrectly analysed, 74% had a P1 latency within 1ms of the boundary while 90% were within 2ms of this limit. It is therefore evident that the majority of mistakes made by ANN 3 were not gross errors, thus instilling more confidence in the system.

An original aim was to develop a system capable of automatically stating the exact P1 latency, thus allowing responses from sequential visits to be compared, and the severity of delays in different regions of the trace array to be examined. The spline fitting technique was unable to achieve this, however by sub-classifying the delayed responses into one of two groups the system could provide the user with slightly more detailed information regarding the extent of the delay.

It was important to assess the ability of the multilayered system to analyse a full trace array as opposed to individual waveforms and to establish the impact of recording quality on the system's capacity to interpret a mfERG test. 20 trace arrays, representing a wide range of patient compliance, were therefore selected. The multilayered system subsequently classified each of the waveforms into one of five groups: 1) normal; 2) decreased in amplitude; 3) delayed; 4) decreased and delayed; or 5) no significant response. It was shown that the system's performance was significantly better when presented with recordings classified as 'excellent' or 'moderate' than when given tests said to be 'noisy' or 'unreportable'. Four of the trace arrays were discussed in greater detail, each taken from one of the four recording quality groups. The system classified 93% of the waveforms correctly for the example said to be an excellent recording, while this agreement with the expert was 90%, 82% and 75% when the system analysed the trace arrays said to be of a 'moderate', 'noisy' and 'unreportable' recording quality, respectively. These findings imply that the user can have more confidence in the interpretation provided by the system when analysing recordings obtained from compliant patients.

Prior to analysing the responses, the multilayered system provides an assessment of the recording quality. A grading of the confidence that the user should have in the system's analysis could therefore be developed based on the integrity of the recording. Furthermore, the system can provide a grading for the recording quality during the test, thus warning the operator of problems with patient compliance. This could enable the operator to rectify the problem, if possible, thus improving the final recording quality and consequently increasing the system's ability to classify the final cross correlated waveforms.

It should be noted that each of the mfERG waveforms used to test the multilayered system was only analysed by one human expert, due to the laborious nature of the task. Given that locating the exact position of the P1 component can be a very difficult task, the classifications provided by the expert may differ to those if the same data had been presented to an alternative expert. It is not therefore possible to state with complete certainty that the multilayered system was incorrect (or correct) when it disagreed (or agreed) with the expert. It would therefore be of interest to ask an additional two human experts to analyse the data utilised to test the system. The majority opinion could therefore be established, and compared with that of the system in each instance. Additionally, the agreement between the three experts when interpreting mfERG data could be established by calculating the Kappa value. This

would provide a more in depth assessment of inter-observer differences when analysing the mfERG. Finally it would be of interest to calculate the Kappa agreement between two experts and the system; this would establish if the current rate of classification differences between the one expert and the system is similar to that between two humans, or if it a classification problem inherent to the design of the system.

It is evident that the multilayered system developed has the potential to provide an objective and automated assessment of the mfERG test, particularly when the integrity of the recording has been classified as either 'excellent' or 'moderate'. This would be particularly useful when large volumes of data require processing, both quickly and consistently, for example in multi centre trials. Given that the human and the system disagreed when analysing a number of the waveforms this multilayered system would not replace the human, but could provide an initial analysis for an expert to review.

7.6 Conclusions

The performance of the system was improved considerably by combining each of the techniques investigated in chapters 4, 5 and 6 (analysis of data in the frequency domain, the use of neural networks and the SNR value, respectively). When distinguishing between 'response' and 'no response' the expert and the system agreed for 95% of the 1000 mfERG waveforms. Of those said to represent an area of functioning retina both the expert and the system categorised them as normal or abnormal in terms of their P1 amplitude and latency. They concurred for 93% of responses for each of these classifications, with the majority of mistakes made when analysing waveforms with a P1 amplitude or latency close to the boundary between normal and abnormal. Finally it was demonstrated that a superior performance was realised by this multilayered system when presented with recordings classified as 'excellent' or 'moderate' by the experts than when analysing those said to be 'noisy' or 'unreportable'. This multilayered system has the potential to be employed for the analysis of the mfERG.

8 Conclusions and further work

In recent years the mfERG has become more widely used as an objective method for monitoring the function of the outer/mid retina, however it is limited by the subjective nature of its interpretation process. This technique has a potentially important role to play in multicentre clinical trials but even experts in the field often disagree when interpreting the data. Furthermore the experience of operators varies considerably; analysis of the responses is relatively simple when recordings are obtained under optimal conditions however difficulties can arise when patient cooperation is reduced. A technique for improving the objectivity and consistency of the analysis process is therefore required. A number of approaches, discussed in chapter 3, have been proposed to achieve this; however each has particular limitations associated with it. Consequently the aim of this thesis was to develop an automated and objective method for interpreting a mfERG recording.

Ideally the system should provide a consistent and objective method for grading the integrity of the recording, both during and after the test. The former would warn the operator if the recording was of an insufficient standard, allowing them an opportunity to address the problem while the patient was still available. The latter would determine whether or not a test should be analysed. Of those mfERG recordings said to be of a suitable quality to report, the system should state if the cross correlated waveforms contain a physiological response or indicate if they represent an area with no significant retinal function. Finally it should report if a response is within normal P1 amplitude and latency values. It would ideally allow P1 to be located accurately to enable responses from sequential visits to be compared, and in the case of delays, provide knowledge of the severity of this delay.

A technique was presented in chapter 4 for grading the quality of a recording. This involved studying the raw, uncorrelated data in the frequency domain. The efficacy of the approach was assessed by presenting 50 mfERG tests to three experts and the system; an agreement of 94% was achieved between the four analyses when making the distinction between which tests should or should not be reported, which was concluded to be acceptable. The agreement between the system and the experts was however poorer when the quality was categorised into one of four groups (excellent, moderate, noisy or unreportable); it fell to 62%. When classifying the integrity of one segment of a recording the experts and the system concurred for 84% of the testing set

when categorising the data into one of two groups: acceptable; or unreportable. It was of interest to note that when classifying the recording integrity, both for the full recording and for one data segment, the agreement between the three experts was similar to that between two experts and the system. The inconsistencies seen between the three human experts highlight the problems associated with using humans as the gold standard as there were many instances of disagreement between the three people.

Three techniques were subsequently investigated to distinguish between 'response' and 'no response': analysis of the Fourier domain profile; artificial neural networks (ANN); and the signal to noise ratio (SNR). The performance of each approach was insufficient when used in isolation however when combined an agreement of 95% between the system and the expert was achieved when tested on 1000 waveforms.

The P1 latency of the mfERG response is the principal measure used to determine if an area of functioning retina is normal or compromised; a delay indicates an area of disease. Two methods were therefore studied to state if the P1 latency was delayed or within normal timing limits: ANNs; and spline fitting. Again a higher performance was required than that yielded by either technique. However when their outputs were combined, an agreement with the expert of 93% was achieved, with the majority of disagreements occurring close to the boundary between normal and abnormal. Finally spline fitting was used to determine if the P1 amplitude of a response was within normal limits. The normal range for the P1 amplitude has been shown to decrease with eccentricity in the trace array therefore a correction was made for this. Again the expert and the spline agreed with one another for 93% of the responses when classifying them as decreased or within normal amplitude limits. The majority of those misclassified had an amplitude value within 5nV of the threshold between normal and decreased.

When testing the system's ability to analyse the 1000 individual waveforms its classification was compared with that of one human expert in each instance. Using a human expert as a gold standard does however have a number of problems associated with it, one of which is the differences in opinion offered by experts when analysing the data. A relatively poor consensus was for example seen between experts in chapter 4 when three people were asked to grade the recording quality of 50 mfERG recordings. It is therefore likely that if an alternative human expert, for example one working in a different visual electrophysiology department, had been asked to analyse the 1000 mfERG waveforms, a certain portion of them would have been graded

differently. Consequently it is difficult to truly assess the ability of the system to interpret the mfERG data as the gold standard itself is subjective and prone to inconsistencies. Furthermore, a human expert may provide a different assessment of the same waveform if presented it on two separate occasions; the system would however provide the same interpretation on each presentation of the data. It is therefore possible that the analysis provided by the system is superior to that of the expert as it is more consistent, always using the same mode of interpretation.

To provide a fairer assessment of the system's performance it would be useful to present the testing data set to additional human experts. Ideally these people would be selected from a variety of centres in the UK. The experts would first have to agree on the main parameters of interest and the basic method of interpretation; once this was decided they would analyse each waveform. The majority opinion could then be compared with that of the system, possibly enabling a fairer assessment of the system's performance. This would reveal if current discrepancies seen between the expert and the system were genuine mistakes, or simply a reflection of the current expert's interpretation. It would also be of interest to ask each additional expert to reanalyse the data on a separate occasion to quantify the repeatability of their analysis.

It is evident from the results presented in chapter 7 that this multilayered system has the potential to provide an automated and objective assessment of the mfERG test. This would provide an initial mode of analysis, which would then be reviewed by a clinician who would also take the patient's clinical history and all other test results into account. Although the system showed good potential it would be prudent to refine each of the individual techniques with a view to increasing the overall performance; a number of methods were suggested in previous chapters to improve each approach. It would also be beneficial to incorporate additional techniques into the system to yield more accurate classifications. One such modification would be to compare the amplitude and latency classification of a response with that of its neighbours; it is for example very unlikely that an individual response would be delayed if all surrounding responses were normal, therefore a possible mistake could be highlighted to the operator if this classification was different.

Chapter 5 utilised ANNs to classify the mfERG data into one of three categories (no significant function, delayed response or response within normal latency limits). The P1 amplitude of a response is also important when analysing the data. It would therefore be useful to develop a neural network dedicated to classifying a response as

decreased in amplitude or within normal limits. It was shown in chapter 3, when a group of 20 healthy individuals were tested to establish a normative range for the mfERG test, that response amplitudes decrease with eccentricity (section 3.4.2). Five different networks would therefore be required to state whether or not a response is decreased in amplitude: one for the central response; and one for each of the four concentric rings.

It would also be of interest to investigate the possible role of principal component analysis in the interpretation of the mfERG. This technique has been successfully applied to the analysis of physiological signals similar to the mfERG; its advantage is that it can reduce the noise present on a response while maintaining the shape of the underlying data. Its aim is to decrease the number of variables in a data set while retaining most of the information in the original data set. Similar to the process used in wavelet analysis the data is decomposed into a number of principal components, with the first components accounting for most of the variance in the data. All components higher than a certain number, for example the third principal component, are discarded as their contribution to the data set is minimal. The data is then reconstructed using the remaining principal components. Zhang *et al.* applied this method of analysis to the mfVECP (140;141), demonstrating that the noise present in the reconstructed mfVECP signal was decreased while its shape was similar to that of the original waveform. This property of principal component analysis would potentially be very useful when analysing the mfERG as the presence of noise causes difficulties when interpreting the waveforms. To ensure that the response was not significantly distorted the latency of the P1 component would have to be compared before and after the application of this technique. If superior results were obtained to those reported using digital and wavelet filtering (sections 6.4 and 6.5) principal component analysis could usurp these approaches. Consequently the ability of the spline fitting technique to locate P1 may be improved if the P1 latency is unaffected and if greater noise reduction is seen. This could potentially fulfil the original objective which was to compare test results from sequential visits, a function not thus far realised as the spline fitting technique is not sufficiently accurate at locating P1.

It was shown that the performance of the multilayered system increased when presented with recordings said to be 'excellent', both by the system and the experts, but fell when analysing tests classified as 'noisy'. It would therefore be appropriate to develop an index detailing the confidence a user should have in the system's interpretation; this would increase in instances when the recording was of a high

standard. To achieve this objective a large number of recordings would have to be studied. 150 tests (50 excellent; 50 moderate; and 50 noisy) could for example be utilised; the classifications provided by the expert and the system would be compared in each case. An overall performance would be obtained for each category of recording quality, both in terms of the system's ability to categorise a waveform as 'response' or 'no response' and to classify responses as normal or abnormal based upon the P1 amplitude and latency. These values could be used to define the confidence an operator should have in the system's analysis for each type of classification; this could be depicted as a percentage, with 100% being completely confident that the system is correct in all instances.

To enable this multilayered system to be used it would have to be incorporated into the mfERG system; this would ensure that the analysis tools used by different groups, for example in multi-centre trials, were identical. Prior to integrating this analysis package into a mfERG system the experimental set up employed would have to be accounted for as the amplitude and latency values used to define normal and abnormal in this thesis are reflective of the particular mfERG set up utilised to acquire the data. If parameters such as the type of electrode, stimulus or amplifier settings were different to those described in the mfERG protocol (section 3.4.1) the system would have to be modified accordingly to account for the differences in the final responses.

The principal aim of this thesis was to improve the objectivity and consistency of the analysis process for the mfERG. An additional factor limiting the expansion of this test is the difference in experimental set up used by different departments. A number of these factors were discussed in chapter 2, including the choice of stimulator, electrode and filter bandwidths for the amplifier. Each of these was shown to affect the final responses slightly differently. Consequently it is impossible to have a standard normal range for the technique. One of the principal differences in testing protocols is the type of stimulus utilised. Many groups use a CRT device to display the stimulus but this is a redundant technology, with problems of equipment replacement. A standard stimulus is therefore required to enable the expansion of this objective clinical technique. To achieve standardisation of the mfERG test a new stimulus is currently being developed and built in Glasgow. It is primarily being developed as a stand-alone device, enabling the different manufacturers to incorporate it into their own systems. As well as improving the consistency of the testing protocol this stimulus will enable temporal aspects of the physiological responses to be investigated as it can

be run at different stimulating frequencies, ranging up to 1kHz. This may lead to a tailoring of the test to specific retinal conditions. The stimulus utilises LEDs and allows the user to state both the intensity and the duration of the pulse of light for every 1ms throughout the test.

Finally, chapter 4 studied the Fourier profile of the uncorrelated data with a view to grading the integrity of a mfERG recording. In doing so the frequencies at which the retina is stimulated during testing were revealed, in addition to the frequencies to which it responded. It was shown that when there were diffuse delays of the cross correlated responses, the retina was unable to respond to the higher stimulation frequencies. Furthermore, the more severe the delays, the greater the loss of the upper frequencies in the Fourier profile. It was also observed that for those mfERG responses with no N2 component, the peak at the stimulus frequency in the Fourier profile was absent, while for those with no significant function, no stimulus-associated peaks were present, implying that the retina could not respond to any of the principal stimulation frequencies. It was therefore found that by viewing the uncorrelated data in the frequency domain, both the temporal and adaptation properties of the retina can be visualised in a very simple manner. This is a considerable advancement as the current method of analysis (measuring the amplitude and latency of P1 and N1 (71)), provides no information on the retina's ability to respond to the different stimulation frequencies. It would therefore be of interest to pursue further investigation into analysis of the Fourier profile as it may reveal important information embedded within the mfERG data, which remains elusive using current analysis techniques, regarding the function of the outer/mid retina. The staging and rate of progression of degenerative retinal conditions such as retinitis pigmentosa could, for example, be studied. The possible quantification of new treatment strategies using this mode of analysis could also be investigated.

In conclusion, the work presented in this thesis has demonstrated the successful development of an objective and automated method for analysing the mfERG data by combining a number of techniques. These included artificial neural networks, analysis of the data in the frequency domain, calculation of the signal to noise ratio, digital filtering, wavelet analysis and the use of curve fitting. This multilayered system can potentially (subject to widespread testing against a more standardised definition of a gold standard drawn from a larger panel of experts) provide a consistent and objective mode of analysis for the mfERG, thus removing some of the problems associated with human experts interpreting the data.

Appendices

Appendix 1

<u>Transforming the uncorrelated data from the time to the frequency domain</u>

This program removes sections of the uncorrelated data obtained from overlapping parts in the m-sequence, and then transforms the data to the Fourier domain. Prior to this transformation the uncorrelated data must be exported from the mfERG program and saved as a raw file (in this case it has been called 'recording1.raw'); this option is available when using the custom built mfERG system.

The raw file is opened and the data are read into Matlab; each overlapping part is named 'ignore' while all other parts are named 'data'. The 'data' parts are then spliced together to form 'y', the uncorrelated data set, which is transformed from the time to the frequency domain. As a reminder the following image demonstrates the overlap of the sequence, with the green areas representing the repetition; the aim is to splice the orange segments together.

1:16	Segment 1 (1:2048)	2032: 2048	2049: 2064	Segment 2 (2049:4096)	4080: 4096	4097: 4112	Segment 3 (4097:6144)	6128: 6144

fid = fopen ('recording1.raw','rb')   *opens raw file called 'recording 1'*

ignore1=fread(fid,2,'uint16');   *read in first two data samples from 'recording 1' file; called 'ignore1' as not required; 'uint16' is the format in which the data are read in (i.e. an unsigned 16 bit integer).*

ignore2=fread(fid,256,'uint16');   *read in next 256 data samples (i.e. first 16 points, each sampled 16 times. As these are repeated in segment, they are ignored.*

data1   = fread(fid,32768, 'uint16');   *read in next 32768 data points (i.e. steps 1 to 2048 in sequence, each sampled 16 times); named 'data1' as these are the data acquired from segment 1, therefore they are utilised.*

ignore3=fread(fid,512,'uint16');   *read in next 512 data points (2032: 2048, each sampled 16 times). An overlap with segment 1, therefore it is ignored.*

% *This process is repeated until reaching the end of the uncorrelated data set*

data2 = fread(fid,32768, 'uint16');
ignore4=fread(fid,512,'uint16');

```
data3 = fread(fid,32768, 'uint16');
ignore5=fread(fid,512,'uint16');
data4 = fread(fid,32768, 'uint16');
ignore6=fread(fid,512,'uint16');
data5 = fread(fid,32768, 'uint16');
ignore7=fread(fid,512,'uint16');
data6 = fread(fid,32768, 'uint16');
ignore8=fread(fid,512,'uint16');
data7 = fread(fid,32768, 'uint16');
ignore9 =fread(fid,512,'uint16');
data8 = fread(fid,32768, 'uint16');
ignore10=fread(fid,512,'uint16');
data9 = fread(fid,32768, 'uint16');
ignore11=fread(fid,512,'uint16');
data10 = fread(fid,32768, 'uint16');
ignore12=fread(fid,512,'uint16');
data11 = fread(fid,32768, 'uint16');
ignore13=fread(fid,512,'uint16');
data12 = fread(fid,32768, 'uint16');
ignore14=fread(fid,512,'uint16');
data13 = fread(fid,32768, 'uint16');
ignore15=fread(fid,512,'uint16');
data14 = fread(fid,32768, 'uint16');
ignore16=fread(fid,512,'uint16');
data15 = fread(fid,32768, 'uint16');
ignore17=fread(fid,512,'uint16');
data16 = fread(fid,32768, 'uint16');
ignore18=fread(fid,256,'uint16');
```

x=[data1;data2;data3;data4;data5;data6;data7;data8;data9;data10;data11;data12;data 13;data14;data15;data16];   *each of the data segments combined to form 'x'; this is the complete uncorrelated data set with overlapping segments removed and comprises 524288 values.*

fclose(fid)   *closes the file 'recording1'.*

% The uncorrelated data, with overlapping segments removed, is subsequently transformed from the time to the frequency domain.

Fs =1200;   *sampling frequency = 1200Hz;*
t=0:1/Fs:524287*1/Fs;   *analysis interval in steps of 1/sampling freq;*
n= 2^(nextpow2(length(x)));   *use next highest power of 2 greater than or equal to length(x) to calculate the fast Fourier transform (FFT); this is essential when using the FFT (section 4.1.3);*

Y = fft(x,n);   *Y is the transformed data in the frequency domain; fft is a built in Matlab function for the fast Fourier transform; x is padded with zeros so that length(Y) is equal to n (i.e. a power of 2);*

NumUniquePts = ceil((n+1)/2);   *calculates the number of unique points (i.e. accounting for symmetry around the Nyquist frequency (section 4.1.2));*

Y = Y(1:NumUniquePts);   *discards data at frequencies greater than Nyquist frequency;*

Y2=Y.*conj(Y)/n;   *to acquire the power spectral density (measure of the energy at each frequency) the complex conjugate is utilised;*

f = (0:NumUniquePts-1)*1200/n;   *this provides an evenly spaced frequency vector with 'NumUniquePts' data points;*

stem (f,Y2);   *plots a stem plot of the frequency against the power spectrum for the uncorrelated data.*

<u>Transforming one segment of the uncorrelated data from the time to the frequency domain</u>

The above program was utilised to transform one segment of data to the frequency domain, with several minor modifications:

1) Instead of splicing all 16 'data' parts together to form x, x comprised only one 'data' part, for example:

X = [data1];

2) The analysis interval was also altered to:

t=0:1/Fs:32768*1/Fs;

Appendix 2

Transforming the correlated data from the time to the frequency domain

The cross correlated data for each eye (i.e. the value of each of the 256 data points for every recovered response), are stored within an ascii file. Additional information such as the patient name is also stored within this file. The cross correlated data were imported into a new Excel spreadsheet and saved as a text file, in this case 'recording2'. This comprised a 256 X 61 matrix (i.e. 61 responses, each of length 256 data points). This was then loaded into Matlab using the following program, either as an individual wave, or as the entire text file; data were then transformed from the time to the frequency domain.

```
in_x= load (recording2.txt');
a=[1]
```
loading data from text file 'recording2';
in this case only the 1st response is selected to transform, but can do all 61   (a=[1:61]);

```
x=(in_x1(:,a));
Fs = 1200;
t=0:1/Fs:255*1/Fs;
```
the data to be transformed to frequency domain;
sampling frequency of 1200Hz;
analysis interval of 256 data points in steps of 1/Fsampling;

```
n= 2^(nextpow2(length(x)));
```
uses next highest power of 2 greater than or equal to length(x) to calculate FFT; this is essential when using FFT (section 4.1.3).

```
Y = fft(x,n);
```
Y is the transformed data in the frequency domain; fft is a built in Matlab function for the fast Fourier transform; x is padded with zeros so that length(Y) is equal to n (i.e. a power of 2);

```
NumUniquePts = ceil((n+1)/2);
```
calculates the number of unique points around the Nyquist frequency;

```
Y = Y(1:NumUniquePts);
```
discards data at frequencies greater than the Nyquist frequency;

```
Y2=Y.*conj(Y)/n;
```
to acquire the power spectral density (measure of the energy at each frequency) the complex conjugate is utilised;

```
f = (0:NumUniquePts-1)*1200/n;
```
this provides an evenly spaced frequency vector with 'NumUniquePts' data points;

```
stem(f,Y2);
```
plot a stem plot of the frequency against the power spectrum for the uncorrelated data.

Appendix 3

Creating synthetic data set: shifting and multiplying by a scaling factor

The original waveform is shifted to the left and the right in increments of 2 data points (1.67ms) until the P1 latency of the secondary waveforms ranges from 35-67ms. They are also multiplied by a scaling factor, generating waveforms with P1 amplitudes ranging from approximately 10nV to 150nV.

To shift a waveform to the right, for example by 4 data points, data point 1 of the original waveform becomes data point 5 on the new wave. Values must therefore be generated for data points 1 to 4. This is done by generating a matrix of random numbers (in this case a 1 by 4 matrix), and adding these to the values of the original data points 1 to 4.

The expert analysed each of the original waveforms, stating the P1 latency and amplitude if the P1 component was present. These values determined the scaling factors chosen and the number of shifts to the right and left. For this example the P1 latency and amplitude are 47ms and 94nV respectively. The following program was written to generate secondary waveforms from this waveform. The values for 'a', 'b' and 'd' were determined by the amplitude and latency of the P1 component.

```
fid = fopen ('wave1_shift_scale.xls','w');     opens Excel file to write data to;

        if (fid==-1)                           fid is a file identifier; if fopen cannot open the
                                               Excel file 'wave1_shift_scale.xls' it
error ('cannot open file for writing');        returns the value -1 and a message stating
end                                            that it cannot open the file is shown     ;
```

% shift to right first. Original response has a P1 latency of 47ms therefore it is shifted by up to 24 data points (corresponds to approximately 20ms).

```
for a = [0: 2: 24]                    for shifting right (0 to 24 in increment of 2);
for b = [0.1 : 0.16: 1.54]            for multiplying by 10 different scaling factors;
c = (2*rand(1,a)-1);                  matrix with dimensions 1 X a, filled with
                                      random numbers ranging from -1 to 1.

for n = a + 1 : 120                   for data points 'a' to 120 each point on wave
    y2(n) = b * (wave1 (n-a));        shifted by 'a'; also multiplied by 'b';
    fprintf (fid, '%f\t', y2(n))      writing each new data point to the Excel
end                                   sheet; tab after each one, i.e. filled into a row

for n = 1 : a                         for data points 1 to 'a' random number added to
                                      the original value;
```

```
   y2(n) = b * (wave1 (n) + c (n)');        also multiplied by 'b';
   fprintf (fid, '%f\t', y2(n))             writing each new data point to the Excel
end                                          sheet; tab after each one, i.e. filled into a row


fprintf (fid, '%f\n', y2(n))                new line; then returns to start of loop again;
end
end
```

%130 secondary waveforms are thus created (13 different shifts, with 10 scaling factors). Each is delayed.

% shift to left – in this instance it is shifted by up to 14 data points (corresponding to approximately 12ms).

```
for d = [2: 2: 14]                          for shifting left (2 to 14 in increment of 2);
for b = [0.1: 0.16: 1.54]                   for multiplying by 10 different scaling factors;
e = = (2*rand(1,d)-1);                       matrix with dimensions 1 X d, filled with
                                             random numbers ranging from -1 to 1;


for n = 1 : 120 – d                          for data points 1 to (120-d), each point is
   y2(n) = b * (wave1(n+d));                 shifted by 'd' points; also multiplied by 'b.'
   fprintf (fid, '%f\t', y2(n))             writing each new data point to Excel file-tab
end                                          after each new value thus new waveform written
                                             to a row;


for n = 120-d+1 : 120                        for data points (120)-d to 120, random number, e
                                             , added to original value;

   y2(n) = b * (wave1 (n-1) + e (d)');       also scaled by 'b;.
   fprintf (fid, '%f\t', y2(n))             writing each new data point to Excel file-tab
end                                          after each new value thus new waveform written
                                             to a row;


fprintf (fid, '%f\n', y2(n))                new line therefore next waveform written to next
                                             line during next round of this 'for' loop;

end
end
```

% An additional 70 secondary waveforms are thus created (7 different shifts, with 10 scaling factors). The first 20 are delayed while the last 50 are within normal P1 latency limits.

```
fclose(fid);                                closes Excel file secondary waveforms have been
                                             written to.
```

Assigning target values (i.e. a classification) to each of the secondary waveforms

The target value assigned to each secondary wave is determined by the process performed to generate it. In this instance the first 150 secondary waveforms are

delayed while numbers 151 to 200 are within normal timing limits. A target value of 0 is therefore assigned to waveforms 1 to 150, while the last 50 responses are given a target value of 0.9.

```
fid = fopen ('TARGETS_shift_scale_wave1.xls','w');      Opens a new Excel file to write
                                                        target values to;

        if (fid==-1)                                    fid is a file identifier; if fopen cannot open the
                                                        Excel file 'wave1_shift_scale.xls' it
error ('cannot open file for writing');                 returns the value -1 and a message stating
end                                                     that it cannot open the file is shown

for k = [ 1 : 150]                                      k will change, depending on original P1 latency;
    target = 0                                          assigns a target value '0' to first 150 waves;
    fprintf (fid, '%f\t', target)                       writing target value to Excel file;
    fprintf (fid, '\n')                                 new line- then returns to start of loop;
end

for k = [ 151 : 200]                                    k will change, depending on original P1 latency;
    target = 0.9                                        assigns a target value '0.9' to last 50 waves;
    fprintf (fid, '%f\t', target)                       writing target value to Excel file;
    fprintf (fid, '\n')                                 new line - then returns to start of loop;
end

fclose (fid);                                           closes Excel file which target values have been
                                                        written to.
```

Appendix 4

<u>Creating synthetic data set: stretching and multiplying by a scaling factor</u>

To generate a stretch the first 30 data points on the waveform are not shifted. The latter 90 data points are however shifted to the right, by varying amounts. The shift on data point 120 is greatest, while it is zero at point 30; it varies linearly between data points 31 and 120. Each shifted waveform is also multiplied by a scaling factor – this is determined by the P1 amplitude of the original waveform as stated by the expert. The classification of the secondary waveforms is determined by the P1 latency of the original waveform, and the amount by which it is stretched.

The aim was to generate waveforms with an amplitude range varying from 10nV to 150nV. In this instance the original waveform, termed 'wave1', had a P1 latency of 47ms and an amplitude of 94nV.

```
fid = fopen ('stretch_wave1.xls','w');
if (fid==-1)


error ('cannot open file for writing');
end
```
opens the Excel file to write data to;
fid is a file identifier- if fopen cannot open the Excel file it returns the value -1, in addition to a message stating that it cannot open the file;

```
for m=5:5:25
```
stretch factor - this ranges from 5 to 25 data points, in increments of 5;

```
for b = 0.1: 0.16: 1.54
```
scaling factor (dependent on the P1 amplitude of original signal);

%first 30 data points are only multiplied by a scaling factor – they are not shifted.

```
for n=1:30
   y2(n)=b*wave1(n);

   fprintf (fid,'%f\t',y2(n));
end
```
selecting the first 30 data points of wave 1;
first 30 data points of all secondary waveforms are multiplied by a scaling factor, but not shifted;
writing to the Excel file 'stretch_wave1.xls';

%the stretch is incorporated at this point – the shift of the final point is varied from 5 to 25 in increments of 5. The shift for the 90 previous data points varies linearly from 0 to that at point 120 (e.g. shift of 25).

```
for n=31:120
   y(n)=del2(round(n-(m*((n-30)/90))));
```
selecting data points 31 to 120 of wave 1;
stretching the signal – the shift is varied linearly along the waveform;

```
    y1(n)=(y(n)+y(n-1))/2;              smoothing the signal by averaging neighbouring
                                        points;
    y2(n)=b*y1(n);                      multiplying by a scaling factor;
    fprintf (fid, '%f\t', y2(n));       writing the data to the Excel file (in a row)
end

fprintf(fid, '%f\n',y2(n));            new line, then returns to the start of the loop;
end
end

fclose(fid);                           closes the Excel file.
```

% The target value assigned to each secondary wave is determined by the process performed to generate it. In this instance all secondary waveforms are delayed. A target value of 0 is therefore assigned to each response.

```
fid = fopen ('TARGETS_stretch_scale_wave1.xls','w');    Opens a new Excel file to write
                                                        the target values to;

if (fid==-1)                           fid is a file identifier- if fopen cannot open
                                       the Excel file it returns the value -1, in
error ('cannot open file for writing');   addition to a message stating that it cannot
end                                    open the file;

for k = [ 1 : 50]                      k will change, depending on the original P1
                                       latency;
    target = 0                         assigning a target value '0' to all 50 waves;
    fprintf (fid, '%f\t', target)      writing target value to the Excel file;
    fprintf (fid, '\n')                new line, then returns to the start of loop to write
                                       to file;
end

fclose (fid);                          closes the Excel file which the target values have
                                       been written to.
```

Appendix 5

Adding 50Hz noise, random noise and drift to the waveforms

Waveforms are randomly selected by this program; 50Hz and random noise are then added to the waveforms. The amplitude of the noise is varied from 10% to 20% of the P1 amplitude of the response (step of 10%), while the phase of the 50Hz is changed from 0^0 to 180^0. 8 waveforms are generated for each waveform selected. Drift is then added to these, creating an additional 12 waveforms for each of the noisy responses.

In this case responses generated from wave 1 using the program detailed in appendix 3 are selected (i.e. shifted and scaled versions of wave 1). 20 responses are selected from a possible 200.

shiftedandscaledwaves = load ('wave1_shift_scale.xls'); *loads file from appendix 3;*

fid=fopen ('noisy_wave1_shift_scale.xls','w'); *creates a new file to write data to;*

if (fid==-1)   *fid is a file identifier- if fopen cannot open*
                                                                         the Excel file it returns the value -1, in
error ('cannot open file for writing');   *addition to a message stating that it cannot*
end   *open the file;*

P1amp=94;   *this is the P1 amplitude of wave 1;*
b=[0.1:0.16:1.54];   *the scaling factors used in appendix 3;*
for k=randsample (200,20)'   *selecting 20 of the 200 scaled and shifted*
                                                                         responses to add the noise to - the randsample
                                                                         command ensures there is no repetition of the
                                                                         random numbers generated;

% 50Hz and random noise are both added to the selected waveforms. The magnitude of this noise is determined by the P1 amplitude of the response to which it is being added. When shifting and scaling wave 1, the wave was shifted and then 10 different scaling factors were applied; it was then shifted again, and scaled 10 times. Consequently waveforms with the same amplitude are to be found in every 10 rows in the file 'wave1_shift_scale'.

% The examples selected from 'wave1_shift_scale' are determined by 'k' (the random number), thus it is different each time the program is run. A switch and case statement is therefore used to direct the program to the appropriate part, ensuring that the correct amplitude of noise is chosen each time the program is run.

switch (k)

*% If the example selected is from rows 1, 11....191 of 'wave1_shift_scale' this loop will be executed. The amplitude is b(1) * P1 amp in this case, where b (1) = 0.1:*

        case {1,11,21,31,41,51,61,71,81,91,101,111,121,131,141,151,161,171,181,191}

```
for a = P1amp*b(1)*0.1 : P1amp*b(1)*0.1 : P1amp*b(1)*0.2   10% - 20% of P1amp
for h = P1amp*b(1)*0.1 : P1amp*b(1)*0.1 : P1amp*b(1)*0.2   10% - 20% of P1amp
        for g=0:pi:pi                                 0⁰ and 180⁰
           y=(h*sin(2*pi*50*t -g));              generating 50Hz signal (t is in seconds)
           outy = a * (2*rand(1,120)-1);         generating random signal
           outpy(1:120)=y(1:120) + outy';        adding 50Hz and random signal together
           outputy =outpy + shiftedandscaledwaves (k,(1:120)); adding noise to wave
           fprintf(fid,'%f\t',outputy)           writing data to Excel file (wave in a row)
           fprintf(fid,'\n')                      new line, then starts the loop again
        end
        end
        end
```

% If the example selected is from rows 2, 12....192 of 'wave1_shift_scale' the following loop will be executed. The only difference between it and the previous loop is that the amplitude of the noise is determined by the second scaling factor used, b(2), instead of b(1) (i.e. 0.26 rather than 0.1):

```
        case {2,12,22,32,42,52,62,72,82,92,102,112,122,132,142,152,162,172,182,192}

        for a = P1amp*b(2)*0.1 : P1amp*b(2)*0.1 : P1amp*b(2)*0.2
        for h = P1amp*b(2)*0.1 : P1amp*b(2)*0.1 : P1amp*b(2)*0.2
        for g=0:pi:pi

           y=(h*sin(2*pi*50*t -g));
           outy = a * (2*rand(1,120)-1);
           outpy(1:120)=y(1:120) + outy';
           outputy =outpy + shiftedandscaledwaves (k,(1:120));
                    fprintf(fid,'%f\t',outputy)
           fprintf(fid,'\n')
        end
        end
        end
```

% This is continued, until that starting with 10, as shown below (i.e. 7 additional ones in between these). In this case the P1 amplitude of the responses to which noise is added is b(10)*P1amp, where b(10) is 1.54:

```
case {10,20,30,40,50,60,70,80,90,100,110,120,130,140,150,160,170,180,190,200}

        for a = P1amp*b(10)*0.1 : P1amp*b(10)*0.1 : P1amp*b(10)*0.2
        for h = P1amp*b(10)*0.1 : P1amp*b(10)*0.1 : P1amp*b(10)*0.2
        for g=0:pi:pi

           y=(h*sin(2*pi*50*t -g));
           outy = a * (2*rand(1,120)-1);
           outpy(1:120)=y(1:120) + outy';
           outputy =outpy + shiftedandscaledwaves (k,(1:120));
                    fprintf(fid,'%f\t',outputy)
           fprintf(fid,'\n')
```

```
            end
            end
            end


end
end


fclose(fid);
```

% 8 waves are thus created for each of the 20 waveforms selected (2 amplitudes for random noise, 2 for 50Hz and 2 phases for 50Hz) therefore 160 waveforms are generated. Drift is then added to each of these 160 responses.

noisyshiftedandscaledwaves=load('noisy_wave1_shift_scale.xls'); *loading file created above when adding 50Hz and random noise;*

fid=fopen('Drift_noisy_wave1_shift_scale.xls','w'); *creates new file to write data to;*
if (fid==-1)   *fid is a file identifier- if fopen cannot open the Excel file it returns the value -1, in*
error ('cannot open file for writing');   *addition to a message stating that it cannot*
end   *open the file;*

% A line of the form y=mx + c is created – this is then added to the waveform to emulate drift.

for g=1:160   *wanting to add drift to all 160 responses;*

for m=-0.75:0.5:0.75   *the gradient of the line is varied from-0.75 to 0.75, in steps of 0.5;*

for c=-20:20:20   *the y-intercept of the line is varied from -20 to 20, in steps of 20;*

 for i=1:120   *generating a line with 120 data points;*
        ydrift(i)=(m*0.83*(i)) + c;   *y=mx+c;*
        outputydrift(1:120)=ydrift(1:120) + noisyshiftedandscaledwaves (g,(1:120));
                                                 adding line of form y=mx+c to waveform
        fprintf(fid,'%f\t',outputydrift(i));   *writing to the Excel file (waveform fills one row);*
end
        fprintf(fid,'%f\n',outputydrift(i));   *new line - then returns to the start of loop;*
end
end
end
fclose(fid);   *closes the Excel file.*
```

*%An additional 12 waveforms are therefore created for each of the responses to which 50Hz and random noise were added, therefore 96 responses are created in total for each of the 20 waveforms selected from the file 'shift_scale_wave1'.*

*Finally a file containing the target value for each of these waveforms was created.*

Assigning target values to each of the noisy waveforms

The P1 latency is unaffected by adding noise and drift therefore the target value of each new response is the same as that of the wave selected from the file 'wave1_shift_scale' to create it.  It was shown in appendix 3 that for wave 1, the first 150 waveforms created were delayed (therefore classified as 0) while numbers 151 to 200 were within normal latency limits (therefore classified as 0.9).  'k', the array of random numbers utilised to choose the 20 responses to which noise were added, is referred to as this determines the target value.

```
fid = fopen ('TARGETS_drift_noisy_wave1_shift_scale.xls','w');   creates new file;

if (fid==-1)                              fid is a file identifier- if fopen cannot open
                                          the Excel file it returns the value -1, in
error ('cannot open file for writing');   addition to a message stating that it cannot
end                                       open the file;

for j=1:20                                for each of the 20 responses selected;
if k(j) > 150                             if the row number of the wave selected >150;
target = 0.9*ones(1,96)'                  the target value for the following 96 responses is
                                          0.9 (i.e. not delayed);

fprintf(fid, '%f\n',target)               write this to the target Excel file
else                                      if the row number of the wave is 150 or less;
target = zeros(1,96)'                     the target value for the following 96 responses is 0
                                          (i.e. delayed);

fprint(fid, '%f\n',target)                write this to the target Excel file
end
end
fclose(fid);                              closes Excel file.
```

Appendix 6

Assigning synthetic data to training and testing set at random

The secondary waveforms created using the methods described in appendices 3-5 were collated into one Excel spreadsheet, named 'syntheticwaves'. Similarly the target values were collected into one Excel file, named 'targetanswer'. The former comprised 120500 rows and 120 columns (120500 waveforms, each of 120 data points), while the latter consisted of 120500 rows. Waveforms and their corresponding target answers (i.e. classification) were selected at random using the following short program, to form the training set (40000 examples) and the testing set (1000 examples):

```
waves = load ('syntheticwaves.xls');
```
loads synthetic waves from Excel file into Matlab – this matrix has been named 'waves';

```
targets = load ('targetanswers.xls');
```
loads answers for waves from Excel into Matlab – this matrix has been named 'targets';

% The aim was to select 41000 examples and their corresponding answers at random from the 120500 waveforms. It was essential to select a random number generator which avoided repetition of numbers, to ensure that the same examples did not appear multiple times in the training set, and that no examples used to train the network featured in the testing set. The Matlab command 'randsample' was selected as this fulfils this requirement.

```
Y=randsample (120500, 41000);
```
a 1 by 41000 matrix (Y) is generated – this encompasses random integers with values ranging from 1 to 120500, with no repetition (i.e. this is used to dictate which waveforms are selected from the complete data set);

```
Y_training = Y (1:40000);
Y_testing = Y (40001:41000);
```
selects the first 40000 of these random numbers;
selects the last 1000 of these random numbers;

```
trainingwaves = waves (Y_training, (1:120));
    trainingtargets = targets (Y_training, (1:120));
```
40000 waves selected for the training set;
40000 answers selected for training set; first 40000 values of Y determined the waves (and their targets) chosen;

```
testwaves = waves (Y_testing, (1:120));
testtargets = targets (Y_testing,(1:120));
```
1000 waves selected for the testing set;
1000 answers selected for the testing set; last 1000 values of Y determined the waves (and their targets) chosen;

% These four data sets were subsequently written to four separate Excel files, which could then be used to train and test the neural networks.

```
fid = fopen ( 'trainingset.xls', 'w');
```
Opens a new Excel file to write training examples to;

```
if (fid==-1)
```
fid is a file identifier- if fopen cannot open

```
error ('cannot open file for writing');
end
```
the Excel file it returns the value -1, in addition to a message stating that it cannot open the file;

```
fprintf (fid, '%f\n', trainingwaves );
```
writes the training waves data to the Excel file;

```
fclose (fid);
```
closes the Excel file to which the training examples have been written.

```
fid = fopen ( 'traininganswers.xls', 'w');
```
Opens a new Excel file to write the training example answers to;

```
if (fid==-1)
```
fid is a file identifier- if fopen cannot open the Excel file it returns the value -1, in

```
error ('cannot open file for writing');
end
```
addition to a message stating that it cannot open the file;

```
fprintf (fid, '%f\n', trainingtargets );
```
writes the training answers to the Excel file;
```
fclose (fid);
```
closes the Excel file to which the training example answers have been written.

```
fid = fopen ( 'testingset.xls', 'w');
```
Opens a new Excel file to write the testing examples to;

```
if (fid==-1)
```
fid is a file identifier- if fopen cannot open the Excel file it returns the value -1, in

```
error ('cannot open file for writing');
end
```
addition to a message stating that it cannot open the file;

```
fprintf (fid, '%f\n', testwaves );
```
writing the testing data to the Excel file;

```
fclose (fid);
```
closes the Excel file to which the testing examples have been written.

```
fid = fopen ( 'testinganswers.xls', 'w');
```
Opens a new Excel file to write testing example answers to;

```
if (fid==-1)
```
fid is a file identifier- if fopen cannot open the Excel file it returns the value -1, in

```
error ('cannot open file for writing');
end
```
addition to a message stating that it cannot open the file;

```
fprintf (fid, '%f\n', testtargets );
fclose (fid);
```
writing the testing answers to the Excel file;
closes the Excel file to which the testing example answers have been written.

Reference List

(1)   Hubel D. Eye, Brain and Vision. Scientific American Library, 1988.

(2)   Lewis A, Zhaoping L. Are cone sensitivities determined by natural color statistics? Journal of Vision 2006; 6(3):285-302.

(3)   Curcio CA, Sloan KR, Kalina RE, Hendrickson AE. Human Photoreceptor Topography. Journal of Comparative Neurology 1990; 292(4):497-523.

(4)   Osterberg G. Topography of the layer of rods and cones in the human retina. Acta Ophthalmologica 1935; 6:1-103.

(5)   Curcio CA, Allen KA. Topography of Ganglion-Cells in Human Retina. Journal of Comparative Neurology 1990; 300(1):5-25.

(6)   Clifford Rose F. The Eye in General Medicine. 1983.

(7)   Celesia G. Visual evoked potentials. In: Celesia G, editor. Disorders of Visual Processing. The Netherlands: Elsevier, 2005: 117-130.

(8)   Carr RE, Siegel IM. Electrodiagnostic Testing of the Visual System: A Clinical Guide. F.A.Davis Company, USA, 1990.

(9)   Odom JV, Bach M, Brigell M, Holder GE, McCulloch DL, Tormene AP et al. ISCEV standard for clinical visual evoked potentials (2009 update). Documenta Ophthalmologica 2010; 120(1):111-119.

(10) Brown M, Marmor M, VAEGAN, Zrenner E, Brigell M, Bach M. ISCEV Standard for Clinical Electro-Oculography (EOG) 2006. Documenta Ophthalmologica 2006; 113(3):205-212.

(11) Frishman LJ. Origins of the electroretinogram. In: Heckenlively JR, Arden GB, editors. Principles and Practice of Clinical Electrophysiology of Vision. Cambridge, Massachusetts: MIT Press, 2006: 139-183.

(12) Marmor MF, Fulton AB, Holder GE, Miyake Y, Brigell M, Bach M. ISCEV standard for full-field clinical electroretinography. Documenta Ophthalmologica 2008; 118:69-77.

(13) Bush RA, Sieving PA. A Proximal Retinal Component in the Primate Photopic Erg A-Wave. Investigative Ophthalmology & Visual Science 1994; 35(2):635-645.

(14) Sieving PA, Murayama K, Naarendorp F. Push-Pull Model of the Primate Photopic Electroretinogram - A Role for Hyperpolarizing Neurons in Shaping the B-Wave. Visual Neuroscience 1994; 11(3):519-532.

(15) Gouras P, Niemeyer G. Electroretinography. In: Celesia G, editor. Disorders of Visual Processing. The Netherlands: Elsevier, 2005: 87-97.

(16) Holder GE. Pattern electroretinography (PERG) and an integrated approach to visual pathway diagnosis. Progress in Retinal and Eye Research 2001; 20(4):531-561.

(17)   Holder GE, Brigell MG, Hawlina M, Meigen T, VAEGAN, Bach M. ISCEV standard for clinical pattern electroretinography - 2007 update. Documenta Ophthalmologica 2007; 114(3):111-116.

(18)   Sutter EE, Tran D. The Field Topography of Erg Components in Man .1. the Photopic Luminance Response. Vision Research 1992; 32(3):433-446.

(19)   Hood DC, Holopigian K, Greenstein V, Seiple W, Li J, Sutter EE et al. Assessment of local retinal function in patients with retinitis pigmentosa using the multi-focal ERG technique. Vision Research 1998; 38(1):163-179.

(20)   Seeliger M. Multifocal electroretinography in retinitis pigmentosa (vol 125, pg 214, 1998). American Journal of Ophthalmology 1998; 125(5):743.

(21)   Dolan FM, Parks S, Hammer H, Keating D. The wide field multifocal electroretinogram reveals retinal dysfunction in early retinitis pigmentosa. British Journal of Ophthalmology 2002; 86(4):480-481.

(22)   Palmowski AM, Sutter EE, Bearse MA, Fung W. Mapping of retinal function in diabetic retinopathy using the multifocal electroretinogram. Investigative Ophthalmology & Visual Science 1997; 38(12):2586-2596.

(23)   Fortune B, Schneck ME, Adams AJ. Multifocal electroretinogram delays reveal local retinal dysfunction in early diabetic retinopathy. Investigative Ophthalmology & Visual Science 1999; 40(11):2638-2651.

(24)   Han Y, Bearse MA, Schneck ME, Barez S, Jacobsen CH, Adams AJ. Multifocal electroretinogram delays predict sites of subsequent diabetic retinopathy. Investigative Ophthalmology & Visual Science 2004; 45(3):948-954.

(25)   Dolan FM, Parks S, Keating D, Dutton GN, Evans AL. Multifocal electroretinographic features of central retinal vein occlusion. Investigative Ophthalmology & Visual Science 2003; 44(11):4954-4959.

(26)   Ohshima A, Hasegawa S, Takada R, Takagi M, Abe H. Multifocal electroretinograms in patients with branch retinal artery occlusion. Japanese Journal of Ophthalmology 2001; 45(5):516-522.

(27)   Kretschmann U, Seeliger MW, Ruether K, Usui T, Apfelstedt-Sylla E, Zrenner E. Multifocal electroretinography in patients with Stargardt's macular dystrophy. British Journal of Ophthalmology 1998; 82(3):267-275.

(28)   McDonagh J, Stephen LJ, Dolan FM, Parks S, Dutton GN, Kelly K et al. Peripheral retinal dysfunction in patients taking vigabatrin. Neurology 2003; 61(12):1690-1694.

(29)   Kellner U, Kraus H, Foerster MH. Multifocal ERG in chloroquine retinopathy: regional variance of retinal dysfunction. Graefes Archive for Clinical and Experimental Ophthalmology 2000; 238(1):94-97.

(30)   Hood DC, Odel JG, Zhang X. Tracking the recovery of local optic nerve function after optic neuritis: A multifocal VEP study. Investigative Ophthalmology & Visual Science 2000; 41(12):4032-4038.

(31) Klistorner A, Fraser C, Garrick R, Graham S, Arvind H. Correlation between full-field and multifocal VEPs in optic neuritis. Documenta Ophthalmologica 2008; 116(1):19-27.

(32) Hood DC, Greenstein VC. Multifocal VEP and ganglion cell damage: applications and limitations for the study of glaucoma. Progress in Retinal and Eye Research 2003; 22(2):201-251.

(33) Graham SL, Klistorner AI, Goldberg I. Clinical application of objective perimetry using multifocal visual evoked potentials in glaucoma practice. Archives of Ophthalmology 2005; 123(6):729-739.

(34) Miele DL, Odel JG, Behrens MM, Zhang X, Hood DC. Functional bitemporal quadrantopia and the multifocal visual evoked potential. Journal of Neuro-Ophthalmology 2000; 20(3):159-162.

(35) Massicotte EC, Semela L, Hedges TR. Multifocal visual evoked potential in nonorganic visual field loss. Archives of Ophthalmology 2005; 123(3):364-367.

(36) Keating D, Parks S. Multifocal techniuqes. In: Heckenlively JR, Arden GB, editors. Principles and Practice of Clinical Electrophysiology of Vision. Cambridge, Massachusetts: MIT Press, 2006: 319-340.

(37) Keating D, Parks S, Evans A, Smith DC. The multifocal ERG: Unmasked by selective cross-correlation. Investigative Ophthalmology & Visual Science 2002; 43:U409.

(38) Sutter EE. The interpretation of multifocal binary kernels. Documenta Ophthalmologica 2000; 100:49-75.

(39) Hood DC, Seiple W, Holopigian K, Greenstein V. A comparison of the components of the multifocal and full-field ERGs. Visual Neuroscience 1997; 14(3):533-544.

(40) Hood DC, Frishman LJ, Saszik S, Viswanathan S. Retinal origins of the primate multifocal ERG: Implications for the human response. Investigative Ophthalmology & Visual Science 2002; 43(5):1673-1685.

(41) Hare W, Ton H. Effects of APB, PDA, and TTX on ERG responses recorded using both multifocal and conventional methods in monkey. Documenta Ophthalmologica 2002; 105:189-222.

(42) Horiguchi M, Suzuki S, Kondo M, Tanikawa A, Miyake Y. Effect of glutamate analogues and inhibitory neurotransmitters on the electroretinograms elicited by random sequence stimuli in rabbits. Investigative Ophthalmology & Visual Science 1998; 39(11):2171-2176.

(43) Hood DC, Greenstein V, Frishman L, Holopigian K, Viswanathan S, Seiple W et al. Identifying inner retinal contributions to the human multifocal ERG. Vision Research 1999; 39(13):2285-2291.

(44) Sutter EE, Bearse MA. The optic nerve head component of the human ERG. Vision Research 1999; 39(3):419-436.

(45) Parks S, Keating D. The multifocal electroretinogram. In: Celesia G, editor. Disorders of Visual Processing. The Netherlands: Elsevier, 2005: 99-115.

(46) Rudolph G, Kalpadakis P, Bechmann M, Haritoglou C, Kampik A. Scanning laser ophthalmoscope-evoked multifocal ERG (SLO-mfERG) in patients with macular holes and normal individuals. Eye 2003; 17(7):801-808.

(47) Rohrschneider K, Bueltmann S. Correlation between findings in multifocal ERG using the SLO versus a CRT monitor for stimulation. Investigative Ophthalmology & Visual Science 2002; 43:U1048.

(48) Seeliger M, Narfstrom K, Reinhard J, Zrenner E, Sutter EE. Continuous monitoring of the stimulated area in multifocal ERG. Documenta Ophthalmologica 2000; 100:167-184.

(49) Kondo M, Miyake Y, Horiguchi M, Suzuki S, Tanikawa A. Recording multifocal electroretinograms with fundus monitoring. Investigative Ophthalmology & Visual Science 1997; 38(5):1049-1052.

(50) Bearse MA, Sutter EE. Imaging localized retinal dysfunction with the multifocal electroretinogram. Journal of the Optical Society of America A-Optics Image Science and Vision 1996; 13(3):634-&.

(51) Huang SZ, Wu DZ, Jiang FT, Luo GW, Liang JJ, Wen F et al. The multifocal electroretinogram in X-linked juvenile retinoschisis. Documenta Ophthalmologica 2003; 106(3):251-255.

(52) Hennessy MP, VAEGAN. Amplitude Scaling Relationships of Burian-Allen, Gold Foil and Dawson, Trick and Litzkow Electrodes. Documenta Ophthalmologica 1995; 89(3):235-248.

(53) Coupland SG. Electrodes for visual testing. In: Heckenlively JR, Arden GB, editors. Principles and Practice of Clinical Electrophysiology of Vision. Cambridge, Massachusetts: MIT Press, 2006: 245-254.

(54) Arden GB, Carter RM, Hogg C, Siegel IM, Margolis S. Gold Foil Electrode - Extending the Horizons for Clinical Electroretinography. Investigative Ophthalmology & Visual Science 1979; 18(4):421-426.

(55) Bultmann S, Rohrschneider K. Reproducibility of multifocal ERG using the scanning laser ophthalmoscope. Graefes Archive for Clinical and Experimental Ophthalmology 2002; 240(10):841-845.

(56) Parks S, Keating D, Williamson TH, Evans AL, Elliott AT, Jay JL. Functional imaging of the retina using the multifocal electroretinograph: A control study. British Journal of Ophthalmology 1996; 80(9):831-834.

(57) Dawson WW, Trick GL, Litzkow CA. Improved Electrode for Electroretinography. Investigative Ophthalmology & Visual Science 1979; 18(9):988-991.

(58) Kretschmann U, Bock M, Gockeln R, Zrenner E. Clinical applications of multifocal electroretinography. Documenta Ophthalmologica 2000; 100:99-113.

(59) Feigl B, Haas A, El Shabrawi Y. Multifocal ERG in multiple evanescent white dot syndrome. Graefes Archive for Clinical and Experimental Ophthalmology 2002; 240(8):615-621.

(60) Mohidin N, Yap MKH, Jacobs RJ. The repeatability and variability of the multifocal electroretinogram for four different electrodes. Ophthalmic and Physiological Optics 1997; 17(6):530-535.

(61) Harding GFA, Wild JM, Robertson KA, Lawden MC, Betts TA, Barber C et al. Electro-oculography, electroretinography, visual evoked potentials, and multifocal electroretinography in patients with vigabatrin-attributed visual field constriction. Epilepsia 2000; 41(11):1420-1431.

(62) Si YJ, Kishi S, Aoyagi K. Assessment of macular function by multifocal electroretinogram before and after macular hole surgery. British Journal of Ophthalmology 1999; 83(4):420-424.

(63) Seeliger MW, Kretschmann UH, Apfelstedt-Sylla E, Zrenner E. Implicit time topography of multifocal electroretinograms. Investigative Ophthalmology & Visual Science 1998; 39(5):718-723.

(64) Bock M, Andrassi M, Belitsky L, Lorenz B. A comparison of two multifocal ERG systems. Documenta Ophthalmologica 1999; 97(2):157-178.

(65) Gerth C, Garcia SM, Ma L, Keltner JL, Werner JS. Multifocal electroretinogram: age-related changes for different luminance levels. Graefes Archive for Clinical and Experimental Ophthalmology 2002; 240(3):202-208.

(66) Shimada Y, Li Y, Bearse MA, Sutter EE, Fung W. Assessment of early retinal changes in diabetes using a new multifocal ERG protocol. British Journal of Ophthalmology 2001; 85(4):414-419.

(67) Heinemann-Vernaleken B, Palmowski AM, Allgayer R, Ruprecht KW. Comparison of different high resolution multifocal electroretinogram recordings in patients with age-related maculopathy. Graefes Archive for Clinical and Experimental Ophthalmology 2001; 239(8):556-561.

(68) Keating D, Parks S, Evans A, Williamson TH, Elliot AT, Jay JL. The effect of filter bandwidth on the multifocal electroretinogram . Documenta Ophthalmologica 1997; 92(4):291-300.

(69) Seeliger MW, Weber BHF, Besch D, Zrenner E, Schrewe H, Mayser H. mfERG waveform characteristics in the RS1h mouse model featuring a 'negative' ERG. Documenta Ophthalmologica 2003; 107(1):37-44.

(70) Gonzalez P, Parks S, Dolan F, Keating D. The effects of pupil size on the multifocal electroretinogram. Documenta Ophthalmologica 2004; 109(1):67-72.

(71) Hood DC, Bach M, Brigell M, Keating D, Kondo M, Lyons JS et al. ISCEV guidelines for clinical multifocal electroretinography (2007 edition). Documenta Ophthalmologica 2008; 116(1):1-11.

(72) Jackson GR, Ortega JD, Girkin C, Rosenstiel CE, Owsley C. Aging-related changes in the multifocal electroretinogram. Journal of the Optical Society of America A-Optics Image Science and Vision 2002; 19(1):185-189.

(73) Mohidin N, Yap MKH, Jacobs RJ. Influence of age on the multifocal electroretinography. Ophthalmic and Physiological Optics 1999; 19(6):481-488.

(74) Fortune B, Johnson CA. Decline of photopic multifocal electroretinogram responses with age is due primarily to preretinal optical factors. Journal of the Optical Society of America A-Optics Image Science and Vision 2002; 19(1):173-184.

(75) Chisholm J, Keating D, Parks S, Evans A. The impact of fixation monitoring on the multifocal electroretinogram. Documenta Ophthalmologica 2001; 102:131-139.

(76) Zhang X, Hood DC, Chen CS, Hong JE. A signal-to-noise analysis of multifocal VEP responses:an objective definition for poor records. Documenta Ophthalmologica 2000; 104:287-302.

(77) Han Y, Bearse MA, Schneck ME, Barez S, Jacobsen C, Adams AJ. Towards optimal filtering of "standard" multifocal electroretinogram (mfERG) recordings: findings in normal and diabetic subjects. British Journal of Ophthalmology 2004; 88(4):543-550.

(78) Bronson-Castain KW, Bearse MA, Han Y, Schneck ME, Barez S, Adams AJ. Association between multifocal ERG implicit time delays and adaptation in patients with diabetes. Investigative Ophthalmology & Visual Science 2007; 48(11):5250-5256.

(79) Greenstein VC, Holopigian K, Seiple W, Carr RE, Hood DC. Atypical multifocal ERG responses in patients with diseases affecting the photoreceptors. Vision Research 2004; 44(25):2867-2874.

(80) Chisholm JA. The multifocal visual evoked cortical potential in visual field mapping: a methodological study. 2008.

(81) Keating D, Ainslie G, Chisholm J, Evans A, Smith D, Parks S. Signal to noise ratio in multifocal ERG records. Investigative Ophthalmology & Visual Science 45, U404. 2004.
Ref Type: Abstract

(82) Keating D, Parks S, Evans A. Technical aspects of multifocal ERG recording. Documenta Ophthalmologica 2000; 100:77-98.

(83) Miyake Y, Horiguchi M, Tomita N, Kondo M, Tanikawa A, Takahashi H et al. Occult macular dystrophy. American Journal of Ophthalmology 1996; 122(5):644-653.

(84) Hood DC, Li J. A technique for measuring individual multifocal ERG records. In: Yager D, editor. Non invasive assessment of the visual system. 1997: 33-41.

(85) Greenstein VC, Holopigian K, Hood DC, Seiple W, Carr RE. The nature and extent of retinal dysfunction associated with diabetic macular edema. Investigative Ophthalmology & Visual Science 2000; 41(11):3643-3654.

(86) Han Y, Adams AJ, Bearse MA, Schneck ME. Multifocal electroretinogram and short-wavelength automated perimetry measures in diabetic eyes with little or no retinopathy. Archives of Ophthalmology 2004; 122(12):1809-1815.

(87) Schneck ME, Bearse MA, Han Y, Barez S, Jacobsen C, Adams AJ. Comparison of mfERG waveform components and implicit time measurement techniques for detecting functional change in early diabetic eye disease. Documenta Ophthalmologica 2004; 108(3):223-230.

(88) Kurtenbach A, Langrova H, Zrenner E. Multifocal oscillatory potentials in type 1 diabetes without retinopathy. Investigative Ophthalmology & Visual Science 2000; 41(10):3234-3241.

(89) Li D, Horiguchi M, Kishi S. Tomographic and multifocal electroretinographic features of idiopathic epimacular membranes. Archives of Ophthalmology 2004; 122:1462-1467.

(90) Yoshii M, Yanashima K, Suzuki S, Okisaka S. Artifact Removal procedure distorts multifocal electroretinogram. Japanese Journal of Ophthalmology 2000; 44(4):419-423.

(91) Bock M, Gerth C, Lorenz B. Impact of notch filter use on waveforms of first- and second-order-kernal responses from multifocal ERGs. Documenta Ophthalmologica 2000; 101:195-210.

(92) Bagolini B, Porciatti V, Falsini B. Binocular interaction and steady-state visual evoked potentials, 1A study in normal subject and in subjects with defective binocular vision. Graefes Archive for Clinical and Experimental Ophthalmology 1988; 226:401-406.

(93) Johansson B, Jokobsson P. Fourier analysis of steady-state visual evoked potentials in subjects with normal and defective stereo vision. Documenta Ophthalmologica 2000; 101:233-246.

(94) Bagolini B, Falsini B, Cermola S, Porciatti V. Binocular interactions and steady-state VEPs. A study in normal and defective binocular vision (Part II). Graefes Archive for Clinical and Experimental Ophthalmology 1994; 232:737-744.

(95) Johansson B, Jakobsson P. Fourier-analysed steady-state VEPs in pre-school children with and without normal binocularity. Documenta Ophthalmologica 2006; 112:13-22.

(96) Asi H, Perlman I. Relationships Between the Electroretinogram A-Wave, B-Wave and Oscillatory Potentials and Their Application to Clinical-Diagnosis. Documenta Ophthalmologica 1992; 79(2):125-139.

(97) Meigen T, Bach M. On the statistical significance of electrophysiological steady-state responses. Documenta Ophthalmologica 1999; 98(3):207-232.

(98) Klistorner AI, Graham SL. Electroencephalogram-based scaling of multifocal visual evoked potentials: Effect on intersubject amplitude variability. Investigative Ophthalmology & Visual Science 2001; 42(9):2145-2152.

(99) Jones AS, Taktak AGF, Helliwell TR, Fenton JE, Birchall MA, Husband DJ et al. An artificial neural network improves prediction of observed survival in

patients with laryngeal squamous carcinoma. European Archives of Oto-Rhino-Laryngology 2006; 263(6):541-547.

(100)   Lundin M, Lundin J, Burke HB, Toikkanen S, Pylkkanen L, Joensuu H. Artificial neural networks applied to survival prediction in breast cancer. Oncology 1999; 57(4):281-286.

(101)   Chen YC, Yang WW, Chiu HW. Artificial neural network prediction for cancer survival time by gene expression data. Bioinformatics and Biomedical Engineering, 3rd International Conference 2009;1-4.

(102)   Niwas SI, Kumari RSS, Sadasivam V. Artificial neural network based automatic cardiac abnormalities classification. Proceedings of the Sixth International Conference on Computational Intelligence and Multimedia Applications 2005.

(103)   Bortolan G, Willems JL. Diagnostic Ecg Classification Based on Neural Networks. Journal of Electrocardiology 1993; 26:75-79.

(104)   Guven A, Kara S. Diagnosis of the macular diseases from pattern electroretinography signals using artificial neural networks. Expert Systems with Applications 2006; 30(2):361-366.

(105)   Kara S, Guven A. Neural network-based diagnosing for optic nerve disease from visual-evoked potential. Journal of Medical Systems 2007; 31(5):391-396.

(106)   Guven A, Kara S. Classification of electro-oculogram signals using artificial neural network. Expert Systems with Applications 2006; 31(1):199-205.

(107)   Lipoth LL, Hafez HM, Goubran RA. Electroretinographical (ERG) based classification of eye diseases. Annual International Conference of the IEEE Engineering in Medicine and Biology Society 1991; 13(3):1417-1418.

(108)   Fisher AC, Hagan RP, Brown MC. Automatic positioning of cursors in the transient pattern electroretinogram (PERG) with very poor SNR using an Expert System. Documenta Ophthalmologica 2007; 115(2):61-68.

(109)   Chouakri SA, Bereksi-Reguig F, Ahmaidi S, Fokapu O. Wavelet denoising of the electrocardiogram signal based on the corrupted noise estimation. Computers in Cardiology 2005; 32:1021-1024.

(110)   Quian Quiroga R, Garcia H. Single-trial event-related potentials with wavelet denoising. Clinical Neurophysiology 2003; 114(2):376-390.

(111)   Fisher AC, Hagan RP, Mackay A, Brown M. Removal of the frame break-through artefact in PRVEP recordings using a system of wavelet decomposition. Annual Meeting of the British Society for Clinical Electrophysiology and Vision (BriSCEV) , 102. 2004.
Ref Type: Abstract

(112)   Hazarika N, Chen JZ, Tsoi AC, Sergejew A. Classification of EEG signals using the wavelet transform. Signal Processing 1997; 59(1):61-72.

(113) Englehart K, Hudgins B, Parker PA, Stevenson M. Classification of the myoelectric activity using time-frequency based representations. Medical Engineering and Physics 1999; 21(6-7):431-438.

(114) Rogala T, Brykalski A. Wavelet feature space in computer-aided electroretinogram evaluation. Pattern Analysis and Applications 2005; 8(3):238-246.

(115) Penkala K. Analysis of bioelectrical signals of the human retina (PERG) and visual cortex (PVEP) evoked by pattern stimuli. Bulletin of the Polish Academy of Sciences, Technical Sciences 2005; 53(3):223-229.

(116) Jimenez JM, Velasco RB, Vazquez LB, Ascariz JM, Villa Polo P. Multifocal electroretinography. Glaucoma diagnosis by means of the wavelet transform. Electrical and Computer Engineering, CCEC, Canadian Conference 2008;867-870.

(117) Chan HL, Brown B. Investigation of retinitis pigmentosa using the multifocal electroretinogram. Ophthalmic and Physiological Optics 1998; 18(4):335-350.

(118) Holder G E, Brigell M G, Hawlina M, Meigen T, Vaegan, Bach M. Iscev standard for clinical pattern electroretinography - 2007 update. Documenta Ophthalmologica 2007; 114(3):111-116.

(119) Bach M, Meigen T. Do's and don'ts in Fourier analysis of steady-state potentials. Documenta Ophthalmologica 1999; 99(1):69-82.

(120) Johansson B, Jakobsson P. Fourier analysis of steady-state visual evoked potentials in subjects with normal and defective stereo vision. Documenta Ophthalmologica 2000; 101:233-246.

(121) Kreyszig E. Advanced Engineering Mathematics. 9th ed. John Wiley & Sons, 2006.

(122) Marks RJ. Handbook of Fourier Analysis and its Applications. 1st ed. Oxford University Press, 2009.

(123) Stoler MH, Schiffman M. Interobserver reproducibility of cervical cytologic and histologic interpretations - Realistic estimates from the ASCUS-LSIL triage study. Jama-Journal of the American Medical Association 2001; 285(11):1500-1505.

(124) Skaane P, Engedal K, Skjennald A. Interobserver variation in the interpretation of breast imaging - Comparison of mammography, ultrasonography, and both combined in the interpretation of palpable noncalcified breast masses. Acta Radiologica 1997; 38(4):497-502.

(125) Jensen P, Krogsgaard MR, Christiansen J, Braendstrup O, Johansen A, Olsen J. Observer Variability in the Assessment of Type and Dysplasia of Colorectal Adenomas, Analyzed Using Kappa-Statistics. Diseases of the Colon & Rectum 1995; 38(2):195-198.

(126) Viera AJ, Garret JM. Understanding interobserver agreement: The Kappa statistic. Family Medicine 2005; 37(5):360-363.

(127)   Landis JR, Koch GG. The measurement of observer agreement for
        categorical data. Biometrics 1977; 33(1):159-174.

(128)   Jain AK, Mao JC, Mohiuddin KM. Artificial neural networks: A tutorial.
        Computer 1996; 29(3):31-44.

(129)   Basheer IA. Selection of methodology for neural network modeling of
        constitutive hystereses behavior of soils. Computer-Aided Civil and
        Infrastructure Engineering 2000; 15(6):440-458.

(130)   Callan R. The Essence of Neural Networks. Great Britain: Prentice Hall
        Europe, 1999.

(131)   Hagan MT, Demuth HB, Beale M. Neural Network Design. Boston: PWS
        Publishing Company, 1996.

(132)   Wang W, Pieter H, Gelder V, Vrijling J. Some Issues About the
        Generalisation of Neural Networks for Time Series Production. Conference
        Proceedings from Artificial Neural Networks: Formal Models and their
        Applications 2005; 3697:559-564.

(133)   Mehrotra K, Mohan C, Ranka S. Elements of Artificial Neural Networks.
        2nd ed. Massachusetts Institute of Technology, 2000.

(134)   Masters T. Practical Neural Network Recipes in C++. 3rd ed. Academic
        Press, USA, 1993.

(135)   Prechelt L. Neural Networks: Tricks of the Trade. Berlin: Springer, 1998: 55-
        69.

(136)   Guven A, Kara S. Classification of electro-oculogram signals using artificial
        neural network. Expert Systems with Applications 2006; 31(1):199-205.

(137)   Kara S, Guven A. Neural network-based diagnosing for optic nerve disease
        from visual-evoked potential. Journal of Medical Systems 2007; 31(5):391-
        396.

(138)   Chisholm J. The multifocal visual evoked cortical potential in visual field
        mapping: a methodological study. 2008.

(139)   Matlab 2007. 2007.

(140)   Zhang M, Hood DC. A principal component analysis of multifocal pattern
        reversal VEP. Journal of Vision 2004; 4(1):32-43.

(141)   Zhang M, Hood DC. Increasing the sensitivity of the multifocal visual
        evoked potential (mfVEP) technique: incorporating information from higher
        order kernels using a principal component analysis method. Documenta
        Ophthalmologica 2004; 108(3):211-222.