2016

Optimised Mammogram Displays for Improved Breast Cancer Detection

Riley, Graeme Alexander

http://hdl.handle.net/10026.1/5150

RESEARCH DEGREES WITH PLYMOUTH UNIVERSITY

OPTIMISED MAMMOGRAM DISPLAYS FOR IMPROVED BREAST CANCER
DETECTION

by

GRAEME ALEXANDER RILEY

A thesis submitted to Plymouth University in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

Health and Human Sciences Doctoral Training Centre

March 2016

Acknowledgements

I would like to express my gratitude to my supervisor, Dr. William Simpson, firstly for offering me the opportunity to study for a PhD in the first place but, subsequently for his support and guidance in all aspects of my research practice. Without his expertise and knowledge this thesis could never have even been imagined let alone reached the threshold of publication. I would also like to thank my second supervisor, Dr Giorgio Ganis for his support and advice along the way. A very special thanks goes to Dr Jim Steele from the Primrose Breast Care Unit at Derriford Hospital in Plymouth, not just for the material support of anonymised images, but also for his own expertise, advise and support in building my knowledge of the highly specialised discipline of mammography.

There are many participants to thank, particularly as psychophysical research demands a high level of concentration and many hours spent searching for the metaphorical needle in a haystack, but two that have provided outstanding support are my wife, Janice Riley and my friend, Sharon Power. Without their hours of participating my data would be far less rich.

Apart from her contribution as a participant, I would also like to acknowledge the unwavering support that I have received from my wife throughout my PhD, as a sounding board for my, often incomprehensible, thoughts, for her understanding during difficult times, her proof reading as well as her love and encouragement without which I could not have finished this thesis.

Author's Signed Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Sub-Committee. Work submitted for this research degree at the Plymouth University has not formed part of any other degree either at Plymouth University or at another establishment.

This study was financed with the aid of a studentship from Plymouth University.

Relevant scientific seminars and conferences were regularly attended at which work was often presented and several papers were prepared for publication.

Presentation and Conferences Attended:

- Only Connect Psychology Conference, Plymouth University, 14 Jun 2013

- Plymouth University Post Graduate Society Conference, Plymouth University, 18 Jun 2013

- Medical Image Perception Society Conference, Washington DC, 13-16 Aug 2013

- Plymouth University Psychology Conference, Plymouth University, 12-13 Jun 2014

External Contacts:

- Dr Jim Steel, Director of Breast Screening, Consultant Breast Radiologist, Primrose Breast-care Unit, Derriford Hospital, Plymouth

Word count of main body of thesis: 58,325

Signed...

Date ....

Abstract

Graeme Alexander Riley

OPTIMISED MAMMOGRAM DISPLAYS FOR IMPROVED BREAST CANCER
DETECTION

In current mammography practice, radiologists typically view mammograms in a symmetric, side-by-side, configuration in the belief that abnormalities will be made salient because they break the perceived symmetry. The literature on the use of symmetry as an aid to signal detection is limited and this thesis has taken a psychophysical approach to investigate the radiologist's task of detecting a small mass (a blob) in paired mammogram backgrounds. Initial experiments used Gaussian white noise and synthetic mammogram backgrounds to test observer performance for the radiologist's task using symmetric (side-by-side) displays and animated (the two images of a pair alternated sequentially in the same location) displays. The use of animated displays was then tested using real mammogram backgrounds in the subsequent experiments. The results showed that side-by-side presentation of paired images does not provide any benefit for the detection of a blob, whereas, alternated presentation enabled the observer to use the correlation present between the paired images to improve detection performance. The effect of alternation was not evident when applied to the task of detecting a small mass in real mammogram pairs and subsequent investigation suggested that the loss of effect resulted from the lack of scale invariance of real images. This meant that, regardless of the level of global correlation between two images, the localised correlation, at a region size reflecting the visual angle subtended by the fovea, was much lower. Thus, decorrelation by the visual system was ineffective and performance for the detection of a blob in the paired images was also ineffective. This thesis suggests that, whilst animated displays can be a powerful tool for the identification of differences between paired images, the underpinning mechanism of decorrelation makes them unsuited for mammograms where scale invariance means that correlation at local levels is a fraction of the global correlation level.

Contents

List of Figures

21

22

List of Tables

List of Abbreviations

2AFC	Two alternative forced choice
BIRADS	The Breast Image Reporting and Data Systems
CAD	Computer aided detection
CC	Cranio-caudal
CI	Confidence interval
d'	Detectability index
GWN	Gaussian white noise
LCL	Lower confidence limit
MAFC	Multiple alternative forced choice
MLO	Medio-lateral Oblique
PDF	Probability density function
PLN	Power law noise
RMS	Root mean square
ROC	Receiver operating characteristic
SD	Standard deviation
SDT	Signal detection theory
SE	Standard error
SKE	Signal known exactly
SKS	Signal known statistically
UCL	Upper confidence limit

Quick Reference Guide

BIRADS	The Breast Image Reporting and Data Systems (BIRADS) is a scheme for categorising the findings of mammogram sceenings.
Contrast	This thesis uses Weber contrast which can be defined, mathematically, as the difference between the luminance of the signal and the luminance of the background divided by the luminance of the background (Whittle, 1994), as shown in the following Equation:

$$C_W = \frac{L_s - L_b}{L_b}$$

Where:

C_W = Weber Contrast

L_s = Luminance of the signal

L_b = Luminance of the background

Contrast threshold	Contrast threshold is defined as "the minimum amount of contrast necessary for a stimulus to be just detectable." (Kingdom & Prins, 2010, p. 10)
Criterion	The criterion refers to the bias of an observer towards making one decison over another.
d' (pronounced d prime)	The detectability index, d', provides a measure of the separation between the means of signal and noise distributions compared against the standard deviation of the signal plus noise distributions.
Fluctuating asymmetry	Fluctuating asymmetry is a measure of biological asymmetry and refers to small random deviations from perfect symmetry in bilaterally paired structures such as may be present between the left and right sides of the human body (Tomkins & Kotiaho, 2002).
Forced choice	Forced choice refers to tasks where the observer is required to make a choice from a predetermined set of choices. This thesis uses a forced choice procedure with two alternative choices (2AFC).

Gaussian white noise (GWN)	GWN is a noise where the pixel values have a normally distributed probability density function and are independent and identically distributed (uncorrelated) (Lu & Dosher, 2014, p. 34).
Parenchyma	The parenchyma refers to the functional tissue of an organ which, in the case of the breast, includes both glandular and fibrous tissues. This is distinct from the connective and supporting tissue (stroma).
Pedestal	A baseline stimulus to which the signal is added (Kingdom & Prins, 2010, p. 265)
Power law noise (PLN)	PLN is a noise with a power spectral density (energy or power per Hertz) that is inversely proportional to its frequency. PLN with a power spectrum of the from $1/f^3$, where f refers to the noise frequency, was used in this thesis because of its similarity to real mammograms. See section 1.6.4 on page 77.
Signal known exactly	The observer has complete a priori information about the signal including its size, shape and location (Abbey & Eckstein, 2000, p. 630).
Signal known statistically	The observer has a priori information about the signal statistics including its size and shape but signal location is uncertain (Burgess, 2010, p. 30).

Chapter 1

The Detection of Visual Signals

This thesis is about the detection of visual signals in noise backgrounds and how this relates to the applied discipline of mammography. Detecting a lesion in a mammogram in psychophysical terms is the detection of a spatial signal in a noise background and there is a long line of psychophysical research related to this. The application of mammography will be considered in chapter 2, however, this chapter will concentrate on the theoretical and psychophysical aspects of the task.

Beginning with the sources of observer variability, the nature and origins of the "noise" that takes the decision making task from being a categorical decision, with a step change from yes to no, to one with a grey area of ambiguity between yes and no will be discussed. With the causes of this ambiguity established, how it can be quantified will be considered in the section on signal detection theory, enabling an empirical basis for decision making to be developed. With this theoretical baseline established, the concept of the ideal observer will be introduced, along with the related measure of observer efficiency.

Following on from the discussion of the decision making process, the subsequent sections will move from the simple description of a signal to consider the more complex qualities of a real signal, such as its spatial and temporal characteristics. The historical research into the response of the human visual system to spatial and temporal frequency will be discussed, introducing the contrast sensitivity function, through Fourier analysis, the temporal contrast sensitivity function and onto the concept of sustained and transient channels. Similarly, a more in depth treatment of the types of noise that will feature in this thesis will be given, along with the display modes that will be employed. The chapter will close by looking at the theoretical concepts so far discussed from the perspective of a human observer and will review of the literature relating to the ability of the human observer to implement those theoretical concepts.

This chapter will review and discuss literature relating to each of the aforementioned areas to

establish the historical and theoretical background for this thesis. The broad scope of research that impacts upon this thesis, as well as the extensive literature produced over the past 80 years, means that this review will, necessarily, be selective, attempting to focus on key moments and breakthrough research, to build the theoretical foundations upon which our own research is based.

1.1 The Sources of Observer Variability

Possibly the first consideration when discussing the detection of visual signals should be the question of what leads to the variability in the decision making process of the observer about the presence, or absence, of a signal when it is close to the observer's threshold for seeing? Variability results from noise which is "usually" considered to be a random process where each individual outcome is variable but, given a large enough number of samples, can be described by a probability distribution. The word "usually" has been inserted because, in the context of medical imaging, non random features, such as anatomical structure, are often classed as noise. This is not limited to medical images and natural images, such as landscapes, people or buildings, also feature structured noise (power law noise) and there is a large literature related to the statistics of natural images (Hyvärinen et al., 2009; Simoncelli & Olshausen, 2001; Field, 1999), however, for the purposes of this section, we will consider noise as a random process. Non random features in mammograms, such as anatomical structure, will be discussed in section 1.6.5 and in chapter 2. Random noise poses important problems for the visual system, and recognising the sources of random noise and quantifying them has occupied researchers over many decades but is an important aspect when considering the detection of a signal. This subsection will outline those sources and some of the important research milestones leading to their understanding.

1.1.1 The effect of the quantum variation of light

Perhaps the first "port of call" in determining the sources of observer variability is to determine the absolute physical threshold of seeing. What is the lowest intensity of light that the human visual system can see? This was the aim of the pioneering work on the absolute threshold of vision carried out by Hecht et al. (1942). The researchers carefully controlled the parameters of the experiment to optimise the conditions for detecting a disc of light which was located 20 degrees from the eye's fixation point so that the light would be incident upon the area of the retina with the highest concentration of rods (Hecht et al., 1942). The spatial and temporal frequencies of the light, as well as its wavelength, were specified to minimise thresholds for detection. The participants were also dark adapted to minimise their sensitivity to light. The disc of light was flashed and participants

were asked to respond "yes" or "no" to say whether they had seen the flash and the procedure was repeated at a number of different flash intensities with a number of participants. Hecht et al. (1942) found that, under optimal conditions, the human visual system can detect a flash of light 60% of the time when it contains about 90 photons. Taking into account the losses experienced in the eye, they estimated that only about 10% of the photons arriving at the eye are actually absorbed by the visual pigment of the rods and, as these photons were spread over about 500 rods (since estimated to be nearer to 350 rods (Cornsweet, 1970, p. 25)), they were able to statistically conclude that only a single photon is necessary to activate a rod. Hecht (1945) clarified this estimate suggesting that, whilst one photon may be sufficient to excite a rod, five photons would be necessary for seeing because of the possibility for spontaneous excitation occurring within a rod. Less than 5 photons would not be sufficient to enable discrimination between these spontaneous excitations and genuine photon excitation, whereas it would be highly unlikely for 5 or more spontaneous excitations to occur together, thus establishing a fixed threshold, below which seeing would not occur (Hecht, 1945). Whilst recognising that neural noise (in the form of spontaneous excitation of rods) did occur, Hecht et al. (1942) did not credit the visual system as able to respond to this, concluding that the variations in human response resulted from the quantum fluctuations in light once above a fixed internal threshold; not from any variability in the participant themselves and any false positive responses below the fixed internal threshold were the result of guesswork rather than any sensory activity. This idea of a fixed internal threshold became known as the high threshold theory (Green & Swets, 1966, p. 127) and, whilst this part of their conclusion would later be dismissed, the concept of noise being inherent in any visual signal was an extremely important finding.

1.1.2 A variable criterion for seeing

The assumption from the high threshold theory that false positive responses below the fixed internal threshold were simply guesses led to the practice of adjusting the psychometric functions to account for this. After all, the theory supposed, how could the visual system respond to neural activity below the fixed threshold?

Therefore, a measured psychometric function, as shown by the dotted line in Figure 1.1, would be corrected to the function shown by the solid line in Figure 1.1 using the formula shown at equation 1.1.

$$p_{corr} = \frac{p_{act} - Q_{fp}}{1 - Q_{fp}} \qquad (1.1)$$

where:
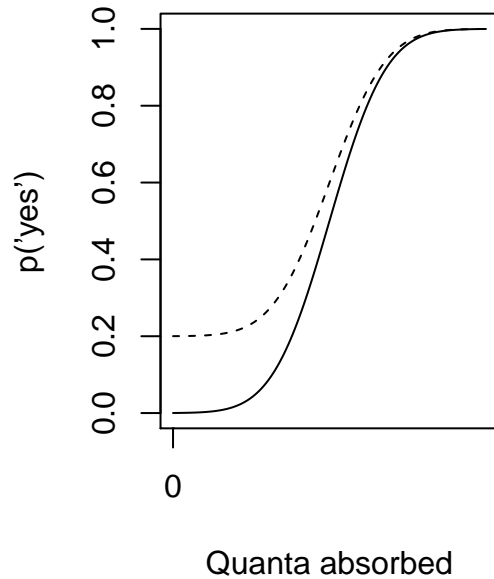
Figure 1.1: Example psychometric function for the probability of saying "yes" signal present against the number of quanta absorbed at the retina. The "as measured" function is shown by the dotted line and the corrected function is shown by the solid line.

p_{corr} = corrected probability of saying yes

p_{act} = measured probability of saying yes

Q_{fp} = false positive response at quanta=0

The concept of the high threshold theory and the notion that responses below a fixed, neurophysiologically determined threshold were the result of guesswork, and thus requiring the psychometric function to be corrected, was challenged by Tanner & Swets (1954). Tanner & Swets (1954) contended that, rather than the result of guesswork, these responses were the result of the visual system responding to neural activity, which must be generated within the visual system itself rather than resulting from the stimulus. Tanner & Swets (1954) proposed that, rather than a fixed, neurophysiologically determined threshold, the observer had a variable threshold, known as a criterion. A criterion refers to the observer's implicit rule for converting the internal response elicited by the stimulus into an external response or decision; for example, internal responses above the criterion will elicit a "yes" response and internal responses below the criterion will elicit a "no" response. This study was largely responsible for the introduction of the theory of signal detectability, now more commonly referred to as signal detection theory (SDT), into the domain of psychophysics (Cohn, 1993, p. 4), though it largely remained in the audition sphere until the 1970s.
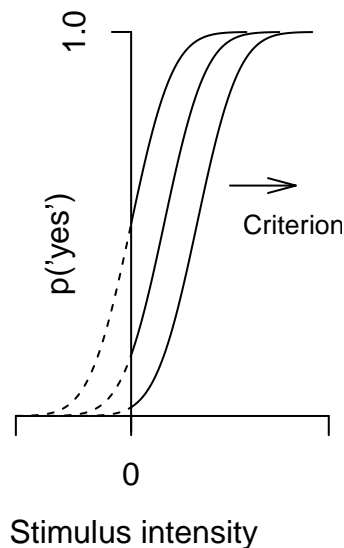
Figure 1.2: The effect of shifting the observers criterion on the position of the psychometric function.

SDT will be treated in greater detail in section 1.2, however, at this juncture, the experiment of Tanner & Swets (1954) will be introduced because of its relevance in countering the concept of high threshold theory, as suggested by Hecht et al. (1942). Tanner & Swets (1954) argued that, rather than observer responses to stimuli below the high threshold limit being the result of guesswork, the response was the result of both the small number of quanta received at the retina from the presented stimulus and a noise generated within the observer that is independent of the signal. Tanner & Swets (1954) presented flashed signals on a blank background to participants in both forced choice and yes/no experiments with a range of signal intensity levels. The participants had to identify the interval in which the signal occurred and their criterion was shifted by informing them of the prior probability of signal presentation as well as with varying financial inducements for each possible decision outcome (correct detection, correct rejection, false alarm or miss).

The results produced similar shape functions to those produced by high threshold theory but, as illustrated in Figure 1.2, by showing how changing the observer's willingness to say signal present, or not, shifted the threshold level for the same signal, they were able to refute the notion of a fixed threshold as suggested by the high threshold theory (Green & Swets, 1966, pp. 127-136). Integral with the concept of SDT was the notion of an internal noise source, independent of the signal, that drove the observer's responses in the absence of a signal, rather than their responses being

guesswork, but the origin of this noise was not elucidated by Tanner & Swets (1954) and, whilst Hecht (1945) recognised the existence of internal noise, it would fall to Barlow (1957) to provide a more detailed investigation into its effects.

The results from Tanner & Swets (1954) suggest that, given an appropriate criterion (and a well trained observer) it should be possible to demonstrate the minimum light intensity that the human visual system is able to detect and, many years later, Sakitt (1972) used SDT to do just this, arguing that only a single photon (or quantum) was necessary for the experience of seeing. Sakitt (1972) used an experimental set up similar to Hecht et al. (1942), with dark adapted participants and disc signals on blank backgrounds flashed to the temporal retina, but instead of a yes/no protocol, he used a rating scale ranging from 0 (we did not see anything) to 6 (we saw a very bright light) and three signal strengths, blank (no signal), weak (average 55 photons at the cornea) and strong (average 66 photons at the cornea). Plotting the average rating score for each signal strength against the number of quanta incident on the cornea, a linear relationship was observed and for subject BS (the study author) the following linear model was derived:

$$\bar{i} = 0.0274 Qc + 0.36 \tag{1.2}$$

where

$\bar{i} =$ the average rating score

$Qc =$ average number of quanta at cornea per flash

The form of the linear model for the average rating score closely matched the linear model for the average number of rod signals:

$$a = f\left(Qc + Xc\right) \tag{1.3}$$

where

$a =$ the average number of rod signals

$f =$ the fraction of incident quanta that produce rod signals

$Qc =$ average number of quanta at cornea per flash

$Xc =$ dark light (now more commonly referred to as internal noise)

From the close similarity of equations 1.2 and 1.3, Sakitt (1972) drew the conclusion that the average rating score was equal to the number of rod signals. As shown in equation 1.3, the number of rod signal results from the average number of effective quantum absorptions plus noise events and using the Poisson distribution of the average rating scores and the known values for f and Qc, Sakitt (1972) was able to calculate the number of quanta incident on the retina as shown in

equations 1.4 to 1.6.

Assuming the average rating score is equal to the number of rod signals then:

$$a = 0.0274 \left(Qc + \frac{0.36}{0.0274} \right) \tag{1.4}$$

$$a = 0.0274 \left(Qc + 13.1 \right) \tag{1.5}$$

From the cumulative probability distribution of rating scores, Sakitt (1972) estimated the average number of rod signals for a criterion of 1 as $a = 0.7$. Thus:

$$Q_c = \frac{0.7}{0.0274} - 13.1 = 12.4 \tag{1.6}$$

Sakitt (1972) actually calculated $Q_c = 12.6$, which was consistent with the estimated display luminance, and he therefore argued that a trained observer (such as himself), could actually count every quantum absorption such that the absolute threshold for seeing would be a single photon.

The results from Sakitt (1972) show that there is no absolute physiological threshold for seeing, rather the threshold for seeing is variable, as dictated by the observer's criterion, and, with the right conditions, training and a low enough criterion, can be as low as a single quantum incident on the retina.

1.1.3 The effect of background luminance on the threshold of seeing

The research cited in sections 1.1.1 and 1.1.2 was concerned with the absolute threshold of the human eye to flashes of light in either totally dark conditions (Hecht et al., 1942) or with a uniform background luminance, constant throughout the experiment (Tanner & Swets, 1954), such that the only variability considered was the quantum fluctuation of the light emitted by the flash. Whilst this early research, as well as that of Sakitt (1972), established absolute thresholds for seeing, the detection of a spatial signal, as will be used in the experiments in this thesis, is a more complex task. With a spatial signal, the observer is no longer simply counting photons, but is now required to match the spatial signal with a template of that signal, in an operation known as cross-correlation. Cross-correlation will be one of the underpinning concepts for all of the experiments in this thesis and will be discussed fully in section 1.3, however, before introducing the concept more formally, the formative research providing the foundations for cross-correlation will be introduced, beginning with the research conducted for the US Navy during World War II by Blackwell in 1946.

This research introduced another source of variability by introducing variation in background

illumination and different sizes of target signal and Blackwell (1946) used spatial signals, rather than light flashes. These were displayed on blank backgrounds, with luminance ranging from zero to 1000 footlamberts (3426 cd/m^2), and the experiments determined the contrast threshold of the human eye (rather than measuring the simple quantum strength of a light flash as used by Hecht et al. (1942); Hecht (1945)).

Before proceeding further, it would be appropriate to clarify and define what contrast means and what is meant by contrast threshold and contrast sensitivity. The contrast of the signal is the relative difference in luminance from its background and for this thesis the Weber contrast will be used. Weber contrast can be defined, mathematically, as the difference between the luminance of the signal and the luminance of the background divided by the luminance of the background (Whittle, 1994), as shown in Equation 1.7

$$C_W = \frac{L_s - L_b}{L_b} \qquad (1.7)$$

Where:

C_W = Weber Contrast

L_s = Luminance of the signal

L_b = Luminance of the background

Contrast threshold is defined as "the minimum amount of contrast necessary for a stimulus to be just detectable." (Kingdom & Prins, 2010, p. 10) and, therefore, the contrast detection threshold would be the minimum value of Weber contrast required to just detect the stimulus. Contrast sensitivity is the reciprocal of contrast threshold and represents the ability of the observer to detect differences between the stimulus and its background.

The circular stimuli used by Blackwell (1946) ranged in diameter from 3.6 to 121 minutes of arc with varying levels of contrast and were projected in one of eight locations on the periphery of a blank circular screen of varying brightness. The participants were instructed to identify the position they thought that the stimulus occupied and threshold contrasts were calculated from the 50% correct point on the observers' psychometric function. Blackwell (1946) was thus able to show how the contrast threshold for the detection of a spatial signal varied with background luminance and disc size.

The results shown in figure 1.3 show that at high background luminance and with large stimuli the relationship between contrast and background luminance obeyed the Weber-Fechner law. The Weber-Fechner law, when applied to contrast threshold, states that if the background luminance is doubled, then to maintain the signal just visible, the luminance difference between the signal and background must also double. Thus, the contrast threshold remains constant as shown in equation
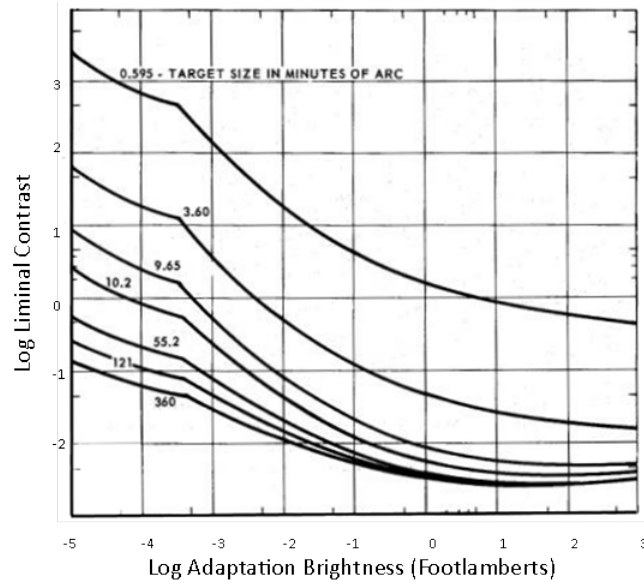
Figure 1.3: A reproduction of Blackwell (1946) Figure 10 showing the variation of contrast levels with room luminance with seven disc sizes.

1.8.

$$\frac{I - I_b}{I_b} = \frac{\Delta I}{I_b} = K_w = Contrast\,Threshold \tag{1.8}$$

where

ΔI = luminance difference

I = signal luminance

I_b = average luminance

K_w = the Weber Fraction

The Weber-Fechner law would, thus, predict a flat function of contrast sensitivity versus background luminance and this is what we see at high luminance levels with large stimuli. At lower background luminance, with smaller stimuli, the relationship between contrast and background luminance no longer obeys the Weber-Fechner law; the contrast needed to see a signal in low luminance increases approximately with the square root of the decreasing background luminance. Whilst not a conclusion drawn by Blackwell (1946), Barlow (1957) later suggested that the deviation from the Weber-Fechner law seen in low luminance levels may result from the increasing influence of what he referred to as "dark light", a concept now more commonly referred to as internal noise.

1.1.4 The effect of internal noise

Barlow (1956) proposed that "dark light" resulted from noise occurring in the optic pathway, in-

39

dependent of any external signal. This was demonstrated by Barlow (1957) in a psychophysical experiment requiring observers to detect spatial disc signals on blank backgrounds which established the threshold values for several disc sizes at varying levels of background luminance. The results were compared against the theoretical prediction that at higher background luminance, the level of background luminance would determine the observer's threshold, whilst at low background luminance, the threshold would level off at a lower threshold determined by the dark light of the optic pathway. This can be seen in Figure 1.4, showing the function of the log increment threshold (the log of the difference between the intensity of the signal and the intensity of the background) plotted against the log of the background intensity.

There are two sources of noise contributing to the total noise in the system, the noise resulting from the intensity of the background and the unknown internal noise. At high background luminance levels the internal noise is insignificant and the function is determined by the background intensity and, in accordance with the Weber-Fechner law, the difference between the increment threshold must be increased to detect the signal as external noise increases. At very low levels of background luminance, internal noise increasingly becomes important and eventually dominates causing a levelling off of the increment threshold required for detection of the signal. As shown in Figure 1.4, the straight sections of the two parts of the curve intersect at the point at which the noise resulting from the background luminance is equal to the internal noise and the dotted line from the intersection onto the x axis indicates its value. The results were consistent with the prediction made by Barlow (1957) and enabled the dark light to be estimated as being equivalent to approximately 1000 quanta entering the eye.

The dark light proposed by Barlow (1957) also fitted well with SDT as proposed by Tanner & Swets (1954), providing the source of noise independent of the signal that drove the observer's response in the absence of a signal, leading to false alarms. Barlow (1957) also recognised the quantal fluctuation of the background luminance as a source of noise independent of the signal and, hence, was recognising the effect of noise internal to the visual system (the dark light) and noise external (the quantal fluctuation of the background luminance).

Nagaraja (1964) used the same principle for estimating internal noise as Barlow (1957) but took an engineering methodology, used to measure the noise in amplifiers, of adding external noise to a system (known as equivalent input noise measurement (Pelli, 1990, p. 4)) and adapted this to measure noise in human visual perception tasks (Lu & Dosher, 2008, 2014, p. 272). Using disc signals on backgrounds with variable noise levels, Nagaraja (1964) plotted contrast levels for the detection of the disc signal as a function of background noise level for several disc sizes and background luminance. It was observed that adding external noise to the system didn't change
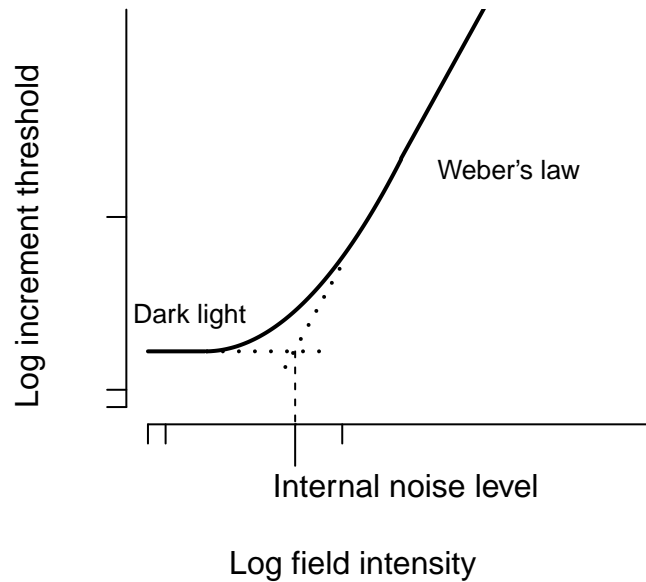
Figure 1.4: Example measurement of equivalent input noise.

the shape of the psychometric function, but simply shifted it to higher contrast thresholds and, thus, Nagaraja (1964) suggested that the contrast threshold is the result of the internal noise plus the additional external noise. Nagaraja (1964) observed that at higher levels of external noise, the internal noise was negligible and could be ignored and using the methodology illustrated in Figure 1.4, was able to estimate the level of internal noise. Decreasing the disc size simply shifted the curve of signal to noise ratio higher without changing the measured level of internal noise.

1.1.5 The internal and external components of internal noise

The approach of adding external noise to a system, as used by Nagaraja (1964), was extended and widely used by numerous researchers over the following decades (for example: Burgess (1985); Burgess & Ghandeharian (1984a,b); Burgess et al. (1981); Legge et al. (1987); Pelli (1981, 1990)). Pelli (1990, p. 12) argued that the internal noise arose from two sources; absorbed photon noise and contrast invariant (or fixed) neural noise which together he referred to as "equivalent noise". Re-analysing data from several earlier studies, Pelli (1990, p. 11) showed that the equivalent noise level reduced as luminance levels increased and went on to show that this change resulted from the variation of absorbed photon noise whilst the contrast invariant neural noise remained constant. As the luminance levels increase so absorbed photon noise decreases and at high luminance levels the

equivalent noise is dominated by neural noise. At low luminance levels, the opposite is true and, as luminance decreases, the equivalent input noise becomes increasingly dominated by absorbed photon noise, though this is at extremely low light levels where rod vision predominates.

Whilst Pelli (1990) showed that the quantum fluctuations of the background luminance make an inversely proportional contribution to the level of internal noise, decreasing as background luminance increases, (Burgess & Colborne, 1988) suggested that added background noise also influenced the level of internal noise. Burgess & Colborne (1988) proposed that neural noise was made up of a fixed component and a second component that was dependent upon the strength of the external noise. Their experiment used two techniques to estimate the ratio of internal noise to external noise; firstly a method called response consistency and secondly using same and different noise fields. The method of response consistency compares the responses of observers to two identical detection experiments, conducted several weeks apart; if the observers had no internal noise their performance should be consistent across experiments and the ratio would approach one, whereas with high levels of internal noise, the results would be inconsistent and the probability of agreement would approach chance (ratio = 0.5). Using the probability ratios determined from the two separate experiments and the known data for the external noise, Burgess & Colborne (1988) calculated an internal noise to external noise ratio of 0.75 at high external noise levels with an increasing ratio as external noise levels reduce, indicating that the internal noise reduced as external noise fell.

A similar result was achieved using the same/different noise field method. The theoretical basis for this was explained using the formulas for the detection of a signal in a two alternative forced choice (2AFC) detection task, as shown in equations 1.11 to 1.20, as detailed by Burgess & Colborne (1988). Burgess & Colborne (1988) used the detectability index $\left(d^{'}\right)$ and before discussing this treatment, it would be worthwhile to briefly introduce the concept of $d^{'}$.

The detectability index $\left(d^{'}\right)$ represents the difference between the mean values of two internal response distributions (for example, signal and noise distributions or two signal distributions) divided by the standard deviation of the two distributions (Green & Swets, 1966, p. 60). The detectability index represents the discriminability between two distributions and, using the detection of a signal in noise, the detectability index can be defined mathematically as:

$$d^{'} = \frac{\mu_{s+n} - \mu_n}{\sigma} \tag{1.9}$$

where

μ_{s+n} = Internal noise distribution of signal plus noise

μ_n = Internal noise distribution of noise only

$\sigma =$ Standard deviation of the distributions (assuming equal variance distributions)

The detectability index $\left(d'\right)$ can also be measured from experimental data using the formula:

$$d' = z\left(H\right) - z\left(F\right) \tag{1.10}$$

(Macmillan & Creelman, 2005, pp. 8-10)

where

$z =$ Inverse of the normal distribution function

$H =$ Hit rate

$F =$ False alarm rate

Referring back to Burgess & Colborne (1988), the detectability index for a signal in different noise fields is as follows in equation 1.11.

$$\left(d'_2\right)^2 = \frac{E^2}{\sigma_0^2 + \sigma_c^2} \tag{1.11}$$

where

$d'_2 =$ detectability index for detection of a signal in different noise fields

$E^2 =$ signal energy

$\sigma_0^2 =$ variance external noise

$\sigma_c^2 =$ variance internal noise

Note: Signal energy $\left(E^2\right)$ is defined as the integral over space and time of the squared signal function. Mathematically this is represented by:

$$E^2 = \int \int \int S^2\left(x, y, t\right) dx dy dt \tag{1.12}$$

where S is defined as the signal function (or contrast), x and y are the signal dimensions measured in degrees of arc and t is time in seconds (Legge et al., 1987)

Burgess & Colborne (1988) stated that for identical noise fields, the ideal observer would subtract the two noise fields, leaving a noiseless image and, therefore, the detectability index for a signal in same noise fields is as follows in equation 1.13.

$$\left(d'_1\right)^2 = \frac{E^2}{\sigma_c^2} \tag{1.13}$$

where

$d'_1 =$ detectability index for detection of a signal in same noise fields

Thus, manipulating equations 1.11 and 1.13:

$$\left(d_2^{'}\right)^2\left(\sigma_0^2+\sigma_c^2\right)=E^2 \qquad and \qquad \left(d_1^{'}\right)^2\sigma_c^2=E^2 \tag{1.14}$$

For the same signal energy:

$$\left(d_2^{'}\right)^2\left(\sigma_0^2+\sigma_c^2\right)=\left(d_1^{'}\right)^2\sigma_c^2 \tag{1.15}$$

and: re-arranging

$$\left(\sigma_0^2+\sigma_c^2\right)=\frac{\left(d_1^{'}\right)^2}{\left(d_2^{'}\right)^2}\sigma_c^2 \tag{1.16}$$

Take σ_c^2 outside the left hand side bracket

$$\sigma_c^2\left(\frac{\sigma_0^2}{\sigma_c^2}+1\right)=\frac{\left(d_1^{'}\right)^2}{\left(d_2^{'}\right)^2}\sigma_c^2 \tag{1.17}$$

Cancel σ_c^2

$$\left(\frac{\sigma_0^2}{\sigma_c^2}\right)=\frac{\left(d_1^{'}\right)^2}{\left(d_2^{'}\right)^2}-1 \tag{1.18}$$

$$\left(\frac{\sigma_0}{\sigma_c}\right)=\left[\frac{\left(d_1^{'}\right)^2}{\left(d_2^{'}\right)^2}-1\right]^{\frac{1}{2}} \tag{1.19}$$

and, finally, the ratio of internal to external noise equals:

$$\left(\frac{\sigma_c}{\sigma_0}\right)=\left[\frac{\left(d_1^{'}\right)^2}{\left(d_2^{'}\right)^2}-1\right]^{-\frac{1}{2}} \tag{1.20}$$

Using this method the internal-external noise ratios were estimated as 0.73 for one observer and 0.63 for the second observer, roughly consistent with the response consistency method. These results enabled Burgess & Colborne (1988) to conclude that internal noise had two components: the fixed, contrast invariant, internal noise as suggested by previous researchers (Barlow, 1957; Nagaraja, 1964; Pelli, 1981) and a variable component directly dependent upon the external noise.

1.1.6   What is internal noise?

Although internal noise can be conveniently categorised into a fixed and variable components, it is likely that both components result from a combination of neuronal noise and noise resulting

from the variability in the decision making process. Considerable work has been carried out into the causes of neuronal noise and Destexhe & Rudolph-Lilith (2012) list a number of sources including thermal noise, channel noise, shot noise and burst noise, amongst others, however a simpler designation is made by Baylor et al. (1980) who suggested that internal noise results from two sources; firstly continuous fluctuations of the neural signal, thought to arise in the transduction process, and, secondly, occasional discrete events most likely caused by thermal or spontaneous isomerisation of the photoreceptor rhodopsin molecules (although this second factor will not be important at the light levels used for the experiments in this thesis). One effect of neuronal noise is to create a transition region for the detection of a signal between not seeing and seeing the signal, where the observer cannot be certain about its presence, or absence, and this, therefore, creates a region where the observer can exercise a variable criterion for seeing. Thus, internal noise is a combination of neuronal noise and noise resulting from criterion variability.

1.1.7 Summary of the sources of observer variability

The research cited here (along with a great number of other contributions) has established that, whilst the absolute threshold for seeing may be as low as 100 photons of light incident on the eye and a single photon absorption at the retina, the actual threshold for the detection of a signal is inherently dependent upon the various sources of noise. That noise has a number of components resulting from factors external and internal to the visual system. For the simple case of a flash of light Hecht et al. (1942) and Hecht (1945) demonstrated that observer variability was caused by the effect of quantum fluctuations of the signal luminance. Barlow (1957) also recognised that the luminance of the background was also subject to quantum fluctuation and, as shown by Blackwell (1946) and Nagaraja (1964), the background luminance itself contributes to the external noise levels impacting upon the observer's ability to detect a stimulus.

Tanner & Swets (1954) found that observer variability also resulted from sources internal to the observer in the form of criterion variability, enabling the observer's response to be shifted independently of any factors associated with the external signal. Identification of the underlying cause of the observer's variability to response criterion changes fell to Barlow (1956, 1957) who coined the term "dark light" to describe, what would become known as, internal noise. Further research has shown that internal noise is made up of induced and fixed components, with the induced component variable and proportional to the strength of the external noise and the fixed component, invariant of external noise. All these separate components need to be considered when determining the contrast threshold (as defined on page 38) for detecting a signal.

1.2 Signal Detection Theory

When studying the decision making process for the detection of a signal, whether it is static or animated, under conditions of variability resulting from noise, a general model for the decision making process is useful. Signal detection theory has been found to be one of the most successful models and is widely used in psychophysical applications, as well as more general contexts (Macmillan & Creelman, 2005, pp. xiii; Wickens, 2001, pp. v-ix). As discussed in section 1.1.2, the widespread use of SDT in psychophysics followed on from research carried out by Tanner & Swets (1954) (Cohn, 1993, p. 4). Tanner & Swets (1954) implemented the mathematical / engineering basis developed by Peterson et al. (1954) and the use of SDT in psychophysics, although initially predominately in audition applications, was refined by Swets (1961) and Swets et al. (1961) and formalised in the classic text by Green & Swets (1966).

Although the experiments in this thesis will involve spatially extended patterns which can be represented by matrices, for simplicity, the concept of SDT will be introduced using a scalar stimulus, such as a flash of light. The application of SDT to spatially extended patterns will be introduced in section 1.3 once the basic principles have been established.

The presentation of a scalar stimulus to a receiver, such as a flash of light containing some number of photons, will result in an internal neural response, however, as a result of the quantum nature of light and the sources of variability discussed in section 1.1, the response to the stimulus will vary across a range of response strengths. If displayed as a probability density function showing the strength of each response against the frequency with which each value of strength occurs, it can be approximately described by a Gaussian distribution (Figure 1.5).
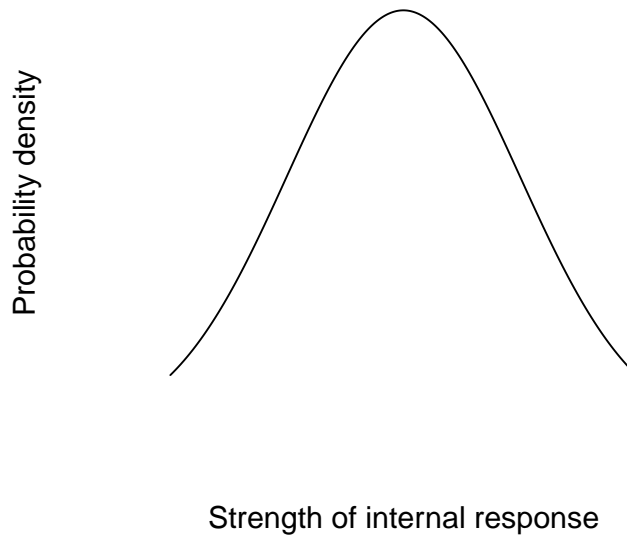
Figure 1.5: Probability density function for the neural internal response to a scalar stimulus, such as a flash of light containing some number of photons.
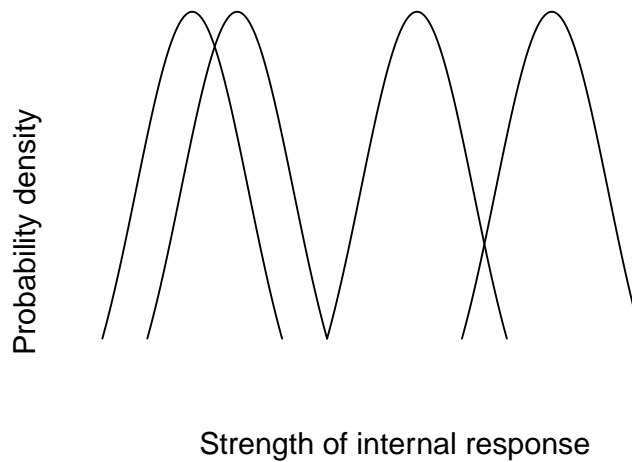
Figure 1.6: Probability density function for the neural internal response to a stimulus consisting of noise only and noise plus a signal. The left hand pair of stimuli represents the noise only stimulus and a stimulus containing a weak signal and the weakness of the signal results in a lot of overlap between the two stimuli. The right hand pair of stimuli represents the noise only stimulus and a stimulus containing a stronger signal which results in less overlap between the two stimuli.

If a signal is added to the original stimulus, the internal response will once again take a range of values, dependent upon the strength of the signal and, once again, reflecting the sources of variability discussed in section 1.1, and this can also be represented by a probability density function. The challenge for the receiver, or observer, is to decide from one internal response whether it was elicited by the noise alone or the noise plus signal. Clearly with a very strong signal this is easy, however, for a weak signal, the probability density functions of noise only and noise plus signal have considerable overlap making the decision of whether the signal is present much more difficult. A criterion is useful here and this enables the observer to make consistent decisions. A criterion refers to the observer's implicit rule for converting the internal response elicited by the stimulus into an external response or decision; for example, internal responses above the criterion will elicit a "yes" response and internal responses below the criterion will elicit a "no" response. Selection of the criterion will inevitably lead to some errors as well as correct decisions and in signal detection theory these can be categorised into one of four response types:

a.   Hit

b.   Miss

c.   False alarm

d.   Correct rejection

As shown in Figure 1.7 :
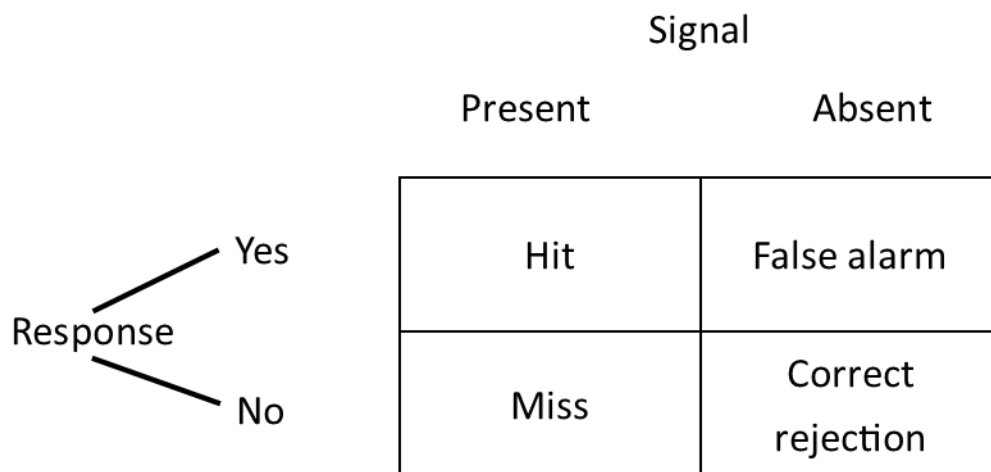
Figure 1.7: Response types to signal present and signal absent.

Taking the criterion level to be λ; if the received stimulus on a single trial elicits a value greater than λ we will say yes, signal present and if it elicits a value less than λ we will say no, signal absent. If the stimulus presented is signal plus noise and the trial elicits a value greater than λ,

we will say yes and this represents a hit. Because the distributions overlap, the trial may elicit a value above λ for a stimulus of noise only, we will again say yes, but this time it will be a false alarm. Note that, whilst stating that the two distributions overlap, they do not actually appear together, indeed, they do not appear to the observer at all and are just pictorial representations of the range of internal responses from a noise only and a signal plus noise stimulus. The observer only sees a single stimulus and they must make a decision regarding whether it is best represented by the noise alone distribution or by the signal plus noise distribution.

We can calculate the probability of these decisions being a hit or a false alarm by calculating the area under probability distribution function of the stimulus above λ.

The probability density function of a Gaussian distribution of a random variable x with zero mean and a variance of σ^2 $(N(\mu, \sigma^2))$ is given by:

$$\phi\left(x\right) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}} \tag{1.21}$$

where:

$\phi\left(x\right) =$ probability density function

$x =$ a random variable

$\sigma^2 =$ variance

The area under this probability density function is found by integrating equation 1.21, so that the probability of correctly saying yes, signal present, given the stimulus $x_{s+n} = signal + noise$ is shown in Figure 1.8 and, mathematically, by equation 1.22:

$$P\left(Yes|signal + noise\right) = \int_{\lambda}^{\infty}\phi\left(x_{s+n}\right)dx \tag{1.22}$$

Similarly, the probability of a false alarm given the stimulus $x_n = noise\,only$ is shown in Figure 1.9 and, mathematically, by equation 1.23.

$$P\left(Yes|noise\,only\right) = \int_{\lambda}^{\infty}\phi\left(x_n\right)dx \tag{1.23}$$

1.2.1   Likelihood ratio

Given the response from a single observation, it would be useful to be able to make a decision regarding whether the observation belonged to the noise only distribution or from the signal plus noise distribution and the likelihood of each can be calculated using the probability density function of the noise and signal plus noise and the value returned by the single observation (Wickens, 2001,
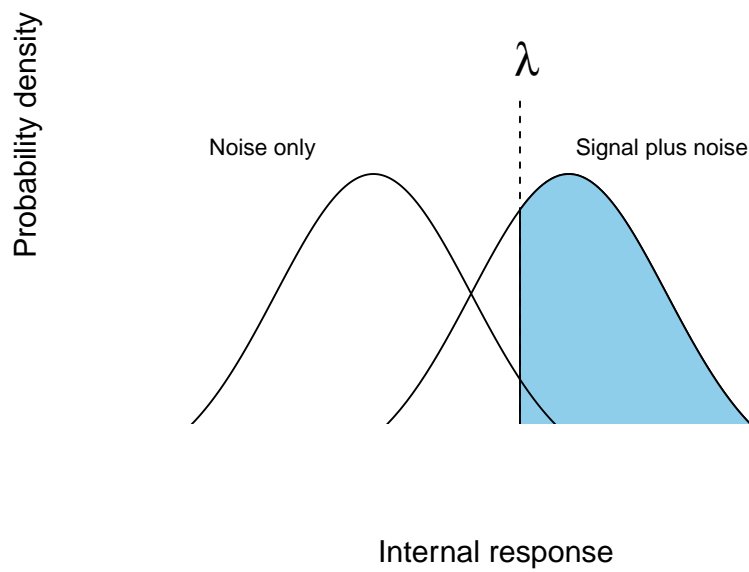
Figure 1.8: The probability density functions of noise only and signal plus noise. The shaded area shows the probability of correctly saying "yes" given a signal from the signal plus noise distribution that is equal to or greater than λ.
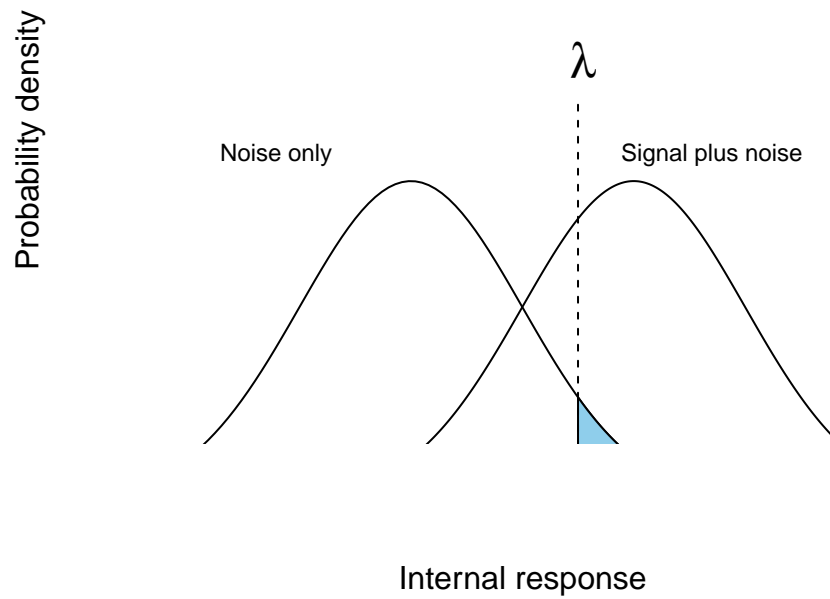
Figure 1.9: The probability density functions of noise only and signal plus noise. The shaded area shows the probability of incorrectly saying "yes" given a signal from the noise only distribution that is equal to or greater than λ.
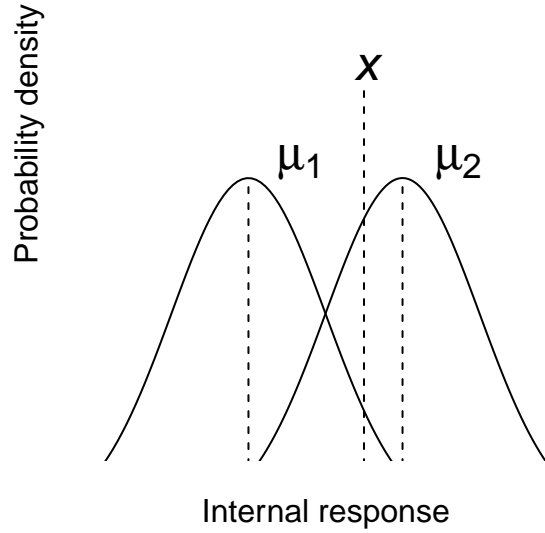
Figure 1.10: A noise only distribution with a mean of μ_1 and the signal plus noise distribution with a mean of μ_2, both with an equal variance of σ^2. The observer receives a stimulus resulting in an internal response of X.

pp. 154-157). Figure 1.10 shows a noise only distribution with a mean of μ_1 and the signal plus noise distribution with a mean of μ_2, both with an equal variance of σ^2.

The likelihood that the observation X belongs to the noise only distribution ($Hypothesis\ 1\ (H_1)$) or to the signal plus noise distribution ($Hypothesis\ 2\ (H_2)$) is given by equations 1.24 and 1.25:

$$Likelihood\ that\ Hypothesis\ 1\ is\ true\ L(H_1): \qquad L\left(X \mid H_1\right) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(X-\mu_1)^2}{2\sigma^2}} \qquad (1.24)$$

of

$$Likelihood\ that\ Hypothesis\ 2\ is\ true\ L(H_2): \qquad L\left(X \mid H_2\right) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(X-\mu_2)^2}{2\sigma^2}} \qquad (1.25)$$

Which can be written as equation 1.26:

$$ratio\ H_1 : H_2 = \frac{L\left(H_1\right)}{L\left(H_2\right)} \qquad (1.26)$$

To ensure manageable numbers and a symmetrical distribution the likelihood ratio is calculated using the logarithm of each likelihood as shown in equation 1.27:

$$\ln\left(ratio\,H_1:H_2\right) = \ln\left(\frac{L\left(H_1\right)}{L\left(H_2\right)}\right) = \ln L\left(H_1\right) - \ln L\left(H_2\right) \tag{1.27}$$

An unbiased observer will have a criterion of zero (the intersection point of the two distributions) and values greater than zero will indicate the likelihood that the signal X comes from the noise only distribution and values less than zero will indicate the likelihood that the signal X comes from the signal plus noise distribution. We can, of course, set the criteria to increase or decrease the likelihood of each decision.

1.2.2   The Bayesian observer

The likelihood ratio simply indicates the likelihood of the observation falling in one distribution or the other, however, it does not give the probability of the observation actually being correct. By taking account of the prior probabilities of each alternative the likelihood ratio can be converted into a probability model; this is the Bayesian observer. By including the prior probability of each outcome $(P\left(H\right))$ with the likelihood of its occurrence $(P\left(X\mid H\right))$ we obtain an *a posteriori* probability $(P\left(H\mid X\right))$ for that event using equation 1.28 (Wickens, 2001, pp. 157-160).

$$P\left(H\mid X\right) = \frac{P\left(X\mid H\right)P\left(H\right)}{P\left(X\right)} \tag{1.28}$$

$P\left(X\right)$ is the probability of observed stimulus occurring and, for the trials used in this thesis, is the same for all observations and, therefore doesn't contribute to the optimal decision rule. By following the same strategy as used with likelihoods of finding the ratio, the value of $P\left(X\right)$ is not required as can be seen from equation 1.30.

$$Bayes\,ratio = \frac{P\left(H_1\mid X\right)}{P\left(H_2\mid X\right)} = \frac{\frac{P(X|H_1)P(H_1)}{P(X)}}{\frac{P(X|H_2)P(H_2)}{P(X)}} \tag{1.29}$$

$$Bayes\,ratio = \frac{P\left(X\mid H_1\right)P\left(H_1\right)}{P\left(X\mid H_2\right)P\left(H_2\right)} \tag{1.30}$$

If hypothesis H_1 is that the signal is present, the Bayesian observer will say signal present if the Bayes ratio is equal to or greater than 1 and signal absent if the ratio is less than 1. We can see from equation 1.30 that if the prior probability of each hypothesis is the same as, for example, in the case of a 2AFC task, then the Bayes ratio will be the same as the likelihood ratio and, in this case, both would describe the optimum decision making strategy for the observer.

1.3   The Detection of Spatially Extended Patterns

Section 1.2 discusses the case of detecting a scalar signal, for example a packet of some number of photons as encountered in many of the early experiments (Barlow, 1957; Hecht et al., 1942; Hecht, 1945). A lesion embedded in a mammogram, and, indeed, a Gaussian blob superimposed on an image, is a more complex situation and involves a two dimensional spatial signal. Now, rather than a scalar quantity representing the strength of a signal, the spatial signal is represented by a matrix which can be flattened into a vector quantity. However, the theoretical approach of SDT, the likelihood observer and the Bayesian observer can still be applied by reducing the vector quantities to a single value through cross-correlation, as described in section 1.3.1. This single value can be compared to a criterion in the same way as section 1.2 or compared to an alternative stimulus, as in the case of a 2AFC task.

1.3.1   Cross-correlation

Cross-correlation is a technique that enables the similarity of two signals to be measured and was originally developed for use in electronics. The basic premise of cross-correlation is that the corresponding pixel values from the template of a signal and of the received stimulus (the image to be searched) are multiplied and summed to give a cross-correlation value of the stimulus and signal template. Taking a 2AFC, signal known exactly paradigm as an example, the cross-correlation would be carried out with the two alternative locations and the location that returns the largest cross-correlation value would be selected as the proposed location of the signal. For a multiple alternative forced choice (MAFC) paradigm, the cross-correlation would be carried out for all possible values of M. For a Yes/No task, we would say Yes, signal present, if the cross-correlation value exceeded the criterion set.

The proof of cross-correlation as the optimum strategy has been provided by Green & Swets (1966, pp. 162-164) and Kay (1998, pp. 95-96), among other texts, and, this will be summarised for a signal known exactly (SKE) task for the detection of a discrete signal in Gaussian white noise below. Taking the optimum decision making strategy of the likelihood ratio as the starting point, the proof will demonstrate that this equates to the cross-correlation of the signal template with the received stimulus $\sum_{n=0}^{N-1} r\,s$ where r is the received stimulus and s is the template of the signal.

Using the format of equations 1.24 and 1.25 as our starting point, and setting the mean of the noise only distribution to zero and assuming equal variances of σ^2, we can derive the likelihood values for the two hypotheses as follows:

$$H_1 : \ r = n \tag{1.31}$$

$$H_2 : \ r = s + n \tag{1.32}$$

where:

H_1 is hypothesis 1 (the received stimulus contains noise only)

H_2 is hypothesis 2 (the received stimulus contains the signal plus noise)

r is the received stimulus

n is Gaussian white noise (GWN) with a mean of zero and a variance of σ^2

s is a discrete signal with zero value outside the range $n = 0$ and $N - 1$

Note: The received stimulus will be summed between the range $n = 0$ and $N - 1$ throughout this proof, i.e. $\sum_{n=0}^{N-1}$, however, for clarity the summation sign will be shown without limits, i.e. $\sum$.

The likelihood that hypothesis H_1 is true (i.e. the received stimulus contains noise only) is given by equation 1.33.

$$Likelihood\,that\,Hypothesis\,1\,is\,true\ (L(H_1)) : \qquad L\left(r \mid H_1\right) = \frac{1}{\sqrt{2\pi}\sigma} exp^{-\frac{\sum (r)^2}{2\sigma^2}} \tag{1.33}$$

where:

$L\left(H_1\right)$ is the likelihood that the received stimulus contains noise only.

The likelihood that hypothesis H_2 is true (i.e. the received stimulus contains the signal plus noise) is given by equation 1.34.

$$Likelihood\,that\,Hypothesis\,2\,is\,true\ (L(H_2)) : \qquad L\left(r \mid H_2\right) = \frac{1}{\sqrt{2\pi}\sigma} exp^{-\frac{\sum (r-s)^2}{2\sigma^2}} \tag{1.34}$$

where:

$L\left(H_2\right)$ is the likelihood that the received stimulus contains the signal plus noise.

The likelihood ratio is:

$$L\left(r\right) = \frac{L\left(H_2\right)}{L\left(H_1\right)} = \frac{\frac{1}{\sqrt{2\pi}\sigma} exp^{-\frac{\sum (r-s)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi}\sigma} exp^{-\frac{\sum (r)^2}{2\sigma^2}}} \tag{1.35}$$

where:

$L(r)$ is the likelihood ratio

Therefore:

$$L(r) = exp\left[\left(-\frac{\sum(r-s)^2}{2\sigma^2}\right) - \left(-\frac{\sum(r)^2}{2\sigma^2}\right)\right] \tag{1.36}$$

Tidying up:

$$L(r) = exp\left[-\frac{1}{2\sigma^2}\left(\sum(r-s)^2 - \sum r^2\right)\right] \tag{1.37}$$

Taking the log of both sides

$$\ln L(r) = -\frac{1}{2\sigma^2}\left(\sum(r-s)^2 - \sum r^2\right) \tag{1.38}$$

Multiplying out:

$$\ln L(r) = -\frac{1}{2\sigma^2}\left(\sum\left(r^2 - 2r\,s + s^2 - r^2\right)\right) \tag{1.39}$$

Which simplifies to:

$$\ln L(r) = \sum\left(r\,s - \frac{s^2}{2\sigma^2}\right) \tag{1.40}$$

Because $\frac{s^2}{2\sigma^2}$ is known and, therefore a constant, its effect is simply to shift the value of $\ln L(r)$ down - it can, therefore be ignored and we are left with:

$$\ln L(r) = \sum r\,s \tag{1.41}$$

Equation 1.41 is the cross-correlation of the stimulus with the signal template and, therefore, provides the mathematical proof of cross-correlation leading to the optimum decision making strategy. Equation 1.41 can be written as a test statistic $T(r)$ that can be compared to a threshold λ, such that when the test statistic is greater than λ we conclude that hypothesis H_2 is true and the signal is present and when less than λ we conclude that hypothesis H_1 is true and the signal is not present .

$$T(r) = \sum r\,s > \lambda \qquad H_2\ is\ true \tag{1.42}$$

$$T(r) = \sum r\,s < \lambda \qquad H_1\ is\ true \tag{1.43}$$

The images presented to the observer can be represented by a column vector with the number of rows equal to the number of pixels and the ideal strategy of cross-correlation can be conveniently handled by linear algebra. Taking the received stimulus to be r, the column vectors for r will be either:

$$r = n \qquad (noise\ only)$$

or:

$$r = s + n \qquad (signal\ plus\ noise)$$

The received stimulus vector is first transposed to enable matrix multiplication and the likelihood ratio is determined by cross-correlating this with the signal vector as shown in equation

$$T\left(r\right) = \left(r\right)^{t} s \tag{1.44}$$

Where ^{t} indicates transpose and $T\left(r\right)$ is the test statistic.

For a yes/no task, if the test statistic exceeds the observer's criterion, the observer will say yes, signal present. For a 2AFC task, with stimuli presented left and right, the observer will cross-correlate with both stimuli and the observer will decide that the signal is present on the left if the cross-correlation value of the left stimulus is larger than that of the right, and right if the cross-correlation value of the right stimulus is larger than the left.

1.4   The Ideal Observer and Observer Efficiency

1.4.1   The ideal observer

The preceding sections have shown that the optimum decision making strategy for a signal known exactly in Gaussian white noise is to cross-correlate the stimulus with a template of the signal to be detected using a likelihood or Bayesian strategy to define the criterion and this defines the ideal observer for this task. Thus for the task pertinent to this thesis of detecting a signal in a noise background, the ideal observer will compute the test statistic $T\left(r\right)$, as shown in equation 1.44, compare this number to a criterion and if it is above the criterion will say "signal present", if below the criterion, "signal absent".

The ideal observer is a widely used technique with a long history in visual perception with early examples of its use exemplified by Peterson et al. (1954); Tanner & Birdsall (1958) and it

is described in detail by Green & Swets (1966, p. 151). Numerous examples of the use of the ideal observer theory occur both in general visual perception research (examples are: Burgess & Judy (2007); Conrey & Gold (2009); Simpson et al. (2003); Tapiovaara (1990)) and in medical image research (examples are: Abbey & Boone (2008); Abbey et al. (2006); Kupinski et al. (2003); Veldkamp et al. (2003)).

The Ideal observer is, thus, a theoretical device that achieves optimal performance for a designated task (Geisler, 2011; Green & Swets, 1966, p. 151). Biological systems such as the human visual system generally don't perform optimally, and therefore the ideal observer should not be thought of as a potentially realistic model of the actual performance of the human visual system. Rather, its value lies in providing a precise measure of the stimulus information available for performing the task and a computational framework that enables optimum performance of that task (Geisler, 2011). In its simplest form the performance of the ideal observer can be used as a benchmark against which to compare the human observer (Geisler, 2011; Green & Swets, 1966, p. 152; Kersten & Mamassian, 2009). In some cases, this can reveal what the human observer can do, but, just as importantly, it can also reveal what the human observer can't do. This information can aid in the design of algorithms for use in an interface that can assist the human observer towards optimum performance (He & Park, 2013).

1.4.2 Can human observers adopt the strategies of the ideal observer?

The preceding sections have shown that, for the detection of a signal in a noise background, the ideal observer will use a strategy of cross-correlating the stimulus with a template of the signal to be detected and make a decision about the presence or absence of the signal using a likelihood or Bayesian decision making strategy. The following subsections will review the literature investigating whether human observers can operate in the same way as ideal observers.

1.4.2.1 Can human observers cross-correlate?

As discussed in section 1.3, the optimum strategy for the detection of a SKE in Gaussian white noise is to cross-correlate the received image with a template of the signal (Green & Swets, 1966, p. 165) and use likelihood or Bayesian reasoning to select the option with the highest *a posteriori* probability and this is, therefore, the strategy that the ideal observer would use for this task. But can the human observer operate in the same way as an ideal observer?

Burgess & Ghandeharian (1984a) proposed that the Bayesian ideal observer would make use of *a priori* information and in the signal known exactly scenario this will include information about the size, shape and location of the signal. Thus, the Bayesian ideal observer will match a template

of the signal to the received stimulus and make a decision based on the most probable hypothesis; the strategy of cross-correlation or template matching. Burgess & Ghandeharian (1984a), in the first of a series of four papers, asked the question of whether human observers could also use a cross-correlation strategy and investigated this by comparing the performance of human observers against the predicted performance of the ideal cross-correlating observer and an alternative strategy, the ideal auto-correlating observer. Whereas the cross-correlating observer cross-correlates the received stimulus with a known template of the signal, the auto-correlating observer is simply an energy detector and, hence, cannot use all the known properties of the signal in the same way as a cross-correlating observer, as illustrated in Equations 1.45 and 1.46.

$$Cross\ correlator:\ say\ yes\ if \quad \sum r \cdot s > criterion \tag{1.45}$$

$$Energy\ detector:\ say\ yes\ if \quad \sum r^2 > criterion \tag{1.46}$$

Burgess & Ghandeharian (1984a) calculated the performance of the ideal cross-correlating observer and derived the performance of the auto-correlating observer, which we will refer to as an energy detector, using Monte Carlo simulations. They used a 2AFC protocol with static noise in one field and static noise plus the signal in the other field and compared the performance of human observers against the two ideal observers in two conditions; with phase information about the signal and without this information. An energy detector is unable to use properties of the signal, such as phase, and would predict the same performance in both conditions. This was not the case and, as shown in Figure 1.11, observers given phase information performed better than without phase information and, indeed, performed better than the ideal energy detector.

This supported the hypothesis of Burgess & Ghandeharian (1984a) that human observers could use a template with information, in this case phase, about the signal and supported the theory that humans can perform cross-correlation detection when given enough information about the signal that they can form a good template of it.

1.4.2.2   Can human observers make use of *a priori* and *a posteriori* information?

Whilst Burgess & Ghandeharian (1984a) demonstrated the ability of the human observer to cross-correlate, the format of the task didn't provide evidence that they could also use a Bayesian strategy in the decision making process, however, this was tested in Burgess & Ghandeharian (1984b) and Burgess (1985). The second paper of the series, Burgess & Ghandeharian (1984b) introduced uncertainty about the signal by carrying out the experiment with multiple possible
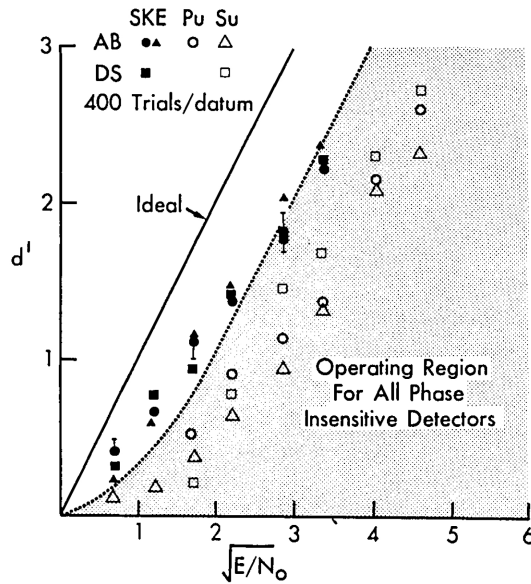
Figure 1.11: Figure reproduced from Burgess & Ghandeharian (1984a) showing that the performance of a human observer, for the detection of a two cycle sine wave, when given signal phase information (filled data points) is better than a human observer not given phase information (open data points). The performance of the human observer given phase information also exceeds that of the performance of a theoretical auto-correlating observer (energy detector) (dotted line) supporting the hypothesis that the human observers can cross-correlate.

locations ranging from 2 to 1800. The optimal, ideal observer, strategy was proposed as cross-correlating the signal locations with a known template of the signal and weighting the cross-correlation with the probability of the signal being in that location, a strategy referred to as the maximum *a posteriori* (MAP) decision strategy. The ideal observer performance was then further weighted by 50% to reflect the likely human observer efficiency. For the ideal observer, signal location uncertainty reduces performance as the number of locations increases. Once again, human performance was compared against the weighted ideal observer performance and the results for the human observers showed a good fit with that of the weighted ideal observer, showing similar decrements in performance as uncertainty increased. The observation that inefficiency remained at 50% regardless of the number of locations supported the hypothesis that human observers can also carry out a MAP strategy as employed by the Bayesian ideal observer.

These conclusions from the first two papers in the series (Burgess & Ghandeharian, 1984a,b) were further supported by the third paper in the series (Burgess, 1985), where signal uncertainty was introduced by increasing the the number of possible signal types to ten, with the signals selected from a Hadamard function set (Pratt, 1978). The study used an alternative forced choice paradigm with a signal selected from the set in a known location but conducted using three methods; detection with signal known, detection with signal unknown and an identification task with signal unknown.
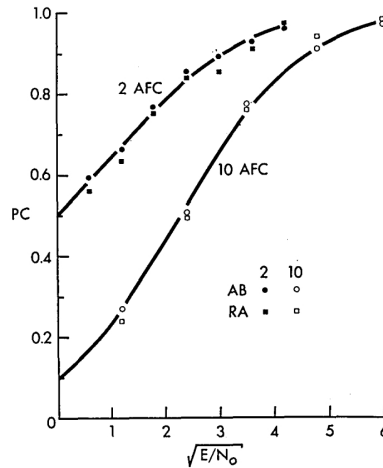
Figure 1.12: Example figure, reproduced from Burgess (1985), showing percent correct against the signal to noise ratio for the 2AFC and 10AFC identification tasks. The theoretical comparison for the ideal observer is shown by the solid line and the comparison for two human observers (AB and RA) are shown by the open and filled symbols.

Burgess (1985) predicted the ideal observer performance by calculation for the signal known exactly condition and using a Monte Carlo simulation for the signal unknown condition and compared this to the human observer performance. As would be expected, the effect of increased signal uncertainty was to reduce ideal observer performance. The same performance decrements were seen for the human observers, who were found to operate with, on average, 33% efficiency when compared to the ideal observer. Figure 1.12, reproduced from Burgess (1985), illustrates this, showing the close agreement between the ideal observer (weighted for human efficiency) and the two human observers for two alternative forced choice and the 10 alternative forced choice conditions.

As in the second paper (Burgess & Ghandeharian, 1984b), the human response varied linearly with that of the ideal observer showing that signal uncertainty has the same effect on the human observer as on the ideal observer and provides support for the hypothesis that human observers can utilise prior signal knowledge to cross-correlate with the received stimuli and use a Bayesian MAP strategy for decision making.

1.4.3   Observer efficiency

The concept of an ideal observer enables the performance of human observers to be compared and their efficiency to be calculated. Barlow (1978) re-introduced the concept of the ideal observer into the domain of the visual psychophysics to calculate the performance of the ideal observer and used this to determine and compare the efficiency of human observers when detecting various higher density dot patterns sited in lower density dot backgrounds. Barlow (1978) found an upper limit

61

of 50% efficiency for human observers carrying out this task.

Burgess et al. (1981) measured the efficiency of human observers when conducting a 2AFC discrimination task where the observers were presented with grating signals embedded in the centres of side-by-side square noise patches. The grating signals were presented as stationary gratings, periodic pulse burst signals or periodic sinusoidal signals and the observers were required to indicate which noise field contained the grating signal with the greatest amplitude. The efficiencies of the human observers were calculated from the ideal observer's performance which was determined using equation 1.47:

$$d'_{Ideal} = \frac{E}{N_0} \tag{1.47}$$

where

d'_{Ideal} = ideal observer performance

$E = \int_{-\infty}^{\infty} \int s^2(x, y)\, dxdy$ = signal energy

N_0 = noise spectral density

and human observer performance from:

$$d'_{Human} = 2erfi\,(2P - 1) \tag{1.48}$$

where

d'_{Human} = human observer performance

$erfi$ = inverse error function

P = proportion of correct responses

and efficiency from:

$$Efficiency = \left(\frac{d'_{Ideal}}{d'_{Human}}\right)^2 \times 100\% \tag{1.49}$$

Burgess et al. (1981) found efficiencies ranging from 54% for an aperiodic Gaussian signal to 83% for a 4.6 cycle/degree sine-wave grating.

Legge et al. (1987) proposed that the overall efficiency of the human observer could be partitioned into two components; one reflected by the observer's equivalent noise (reflecting the level of internal noise) and a second related to the observer's sampling efficiency. Legge et al. (1987) used the equivalent noise technique, described in section 1.1.4, and the established relationship between the threshold signal energy (E_t) and noise, described by equation 1.50:

$$E_t = k\,(N + N_{eq}) \tag{1.50}$$

where

E_t = signal energy at threshold

k = slope

N = noise spectral density. Noise spectral density refers to the noise power per unit of bandwidth and is sometimes known as the power spectral density of the noise. Legge et al. (1987) calculated this by multiplying the pixel area with the squared value of the root mean square (rms) contrast of the noise.

N_{eq} = equivalent noise

Note: The parameters k and N_{eq} can be estimated by plotting the performance data as shown in Figure 1.13 as described on the next page

The observer's sampling efficiency is related to the effectiveness by which the observer cross-correlates the signal template with the target signal with mismatch in template size and shape, along with incomplete spatial or temporal summation leading to reductions in sampling efficiency (Legge et al., 1987). The study included two separate experiments, one using a 2 cycle/degree sine wave as the target signal, the other using a 13.6 arcmin disc as the target signal, both embedded in a pedestal to facilitate a discrimination task between signal and pedestal, with added external noise that was static in the first experiment and dynamic in the second (Legge et al., 1987). In both cases the threshold signal energy for discrimination was plotted against the noise spectral density such that the separate contributions of equivalent noise and sampling efficiency could be estimated as shown in Figure 1.13, extracted from Legge et al. (1987).

For an ideal observer, with no internal noise and a sampling efficiency of 1, the signal detectability index d' will be given by:

$$d' = \sqrt{\frac{E_t}{N}} \tag{1.51}$$

And, therefore the signal energy E_t required by the ideal observer for detection will be:

$$E_t = d'^2 N \tag{1.52}$$

For a human observer with internal noise and a sampling efficiency of 1, the signal energy E_t required for detection will be:

$$E_t = d'^2 \left(N + N_{eq} \right) \tag{1.53}$$

Sampling efficiency is the reciprocal of the slope k, i.e. as the slope increases, sampling efficiency
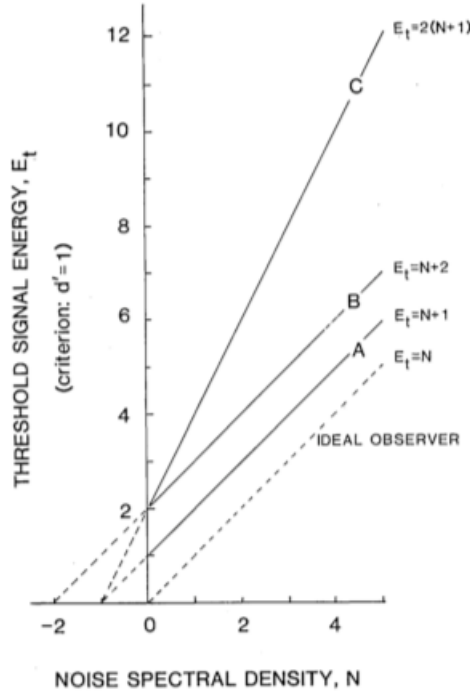
Figure 1.13: Illustration of Legge et al. (1987) methodology for partitioning contrast threshold for discrimination into the effect of sampling efficiency and the effect of equivalent noise.

reduces, and, therefore, sampling efficiency J can be defined as:

$$J = \frac{d'^2}{k} \tag{1.54}$$

and hence, equation 1.50 on page 62 can be rewritten as:

$$E_t = \frac{d'^2}{J}(N + N_{eq}) \tag{1.55}$$

With reference to Figure 1.13, we can determine the contrast threshold for discrimination with no added noise from the intercept on the Y axis and this shows that functions B and C have the same threshold energy, but, from threshold energy alone it would not be possible to identify the contribution of the individual sources of inefficiency. However, by adding external noise and plotting E_t against noise spectral density N, with the signal detectability index maintained at $d' = 1$, using equation 1.50, we are able to determine the contribution of sampling efficiency from the reciprocal of the slope of the function and the (negative) value of the equivalent noise from the intercept on the X axis.

From Figure 1.13, we can see that the ideal observer has a slope of 1 and no equivalent noise, therefore $E_t = N$. An observer with a sampling efficiency of 1 but with added equivalent noise of

64

1 is shown by function A, where $E_t = N + 1$. An observer with a sampling efficiency of 1 but with added equivalent noise of 2 is shown by function B, where $E_t = N + 2$. An observer with a sampling efficiency of 0.5 and an equivalent noise of 1 is shown by function C, where $E_t = 2\,(N + 1)$.

Using this methodology, Legge et al. (1987) were able to determine that the major contribution to the variation in contrast discrimination, and thus, to the overall efficiency of the human visual system for this task, resulted from the variation in equivalent noise, with sampling efficiency remaining relatively constant across the range of pedestal contrast.

Pelli (1990) also attempted to consolidate and clarify the components that together determined the overall quantum efficiency of vision, breaking down the process of vision from the presentation of the image to the performance of the assigned task, be it detection or discrimination, into three discrete stages. The first two stages being the production of the photon image at the retina and the production of an effective image within the cortex which, together, Pelli (1990) called transduction. The third stage Pelli (1990) referred to as calculation which represented the observer's use of the effective image in decision making and this is analogous to the sampling efficiency of Legge et al. (1987). Calculation efficiency is also known as central efficiency (Barlow, 1977) and detection efficiency (Kersten, 1987).

The first stage of transduction requires the conversion of photons from the stimulus into an image at the retina, and includes the random nature of the the luminance received at the cornea, the losses experienced within the eye plus the impact of noise added to the display and Pelli (1990) described the signal to noise ratio (SNR) at this stage as:

$$SNR_1 = \frac{E}{N + N_{photon}} \tag{1.56}$$

where

$\quad E$ = signal energy

$\quad N$ = added display noise

$\quad N_{photon}$ = corneal plus photon noise

The formation of the effective image in the cortex introduces the effect of neural noise and the contrast invariant quotient of this added to the corneal and photon noise can be estimated using the equivalent noise method described in section 1.1.4 such that the SNR at this stage is:

$$SNR_2 = \frac{E}{N + N_{eq}} \tag{1.57}$$

where

$\quad N_{eq}$ = equivalent noise = corneal plus photon noise and neural noise

Interestingly, Pelli (1990) has not included the noise resulting from the decision making process in equivalent noise which, by inference, previous researchers have done (Barlow, 1957; Nagaraja, 1964), but includes it in the final performance measure $\left(d'^2\right)$ for the task in a similar manner to Legge et al. (1987). Efficiency can be calculated for each stage and the overall quantum efficiency is equal to the product of the efficiencies at each stage. Hence:

$$Transduction\ efficiency\ (F_1) = \frac{\frac{E}{N+N_{eq}}}{\frac{E}{N+N_{photon}}} = \frac{N + N_{photon}}{N + N_{eq}} \qquad (1.58)$$

$$Calculation\ efficiency\ (F_2) = \frac{d'^2}{\frac{E}{N+N_{eq}}} \qquad (1.59)$$

and

$$Overall\ quantum\ efficiency\ (F) = F_1 F_2 = \frac{d'^2}{\frac{E}{N+N_{photon}}} \qquad (1.60)$$

Using these definitions of efficiency, Pelli (1990) demonstrated, using data from a number of previous studies that transduction efficiency is relatively constant at around 1% and, thus, variation in efficiency must result from variation in calculation efficiency. Whilst this appears to contradict the conclusion drawn by Legge et al. (1987), this may be explained by the methodology. Legge et al. (1987) showed that sampling efficiency within a task remained constant, however acknowledged that sampling efficiency between different tasks (static noise and dynamic noise) did change, in agreement with the conclusion drawn by Pelli (1990).

1.5 The Effect of Signal Characteristics on Signal Detectability

So far we have considered the detection of a signal in noise in isolation, without considering the characteristics of the signal such as its spatial or temporal frequency. In the clinical environment, the breast radiologist is trying to detect abnormalities that are represented by spatially extended patterns with a range of spatial frequencies and in this thesis we will use a Gaussian blob with a predetermined spatial frequency. This thesis will also add a temporal aspect to the task, introducing a signal that varies with time. These characteristics of spatial and temporal frequency will impact on the ability of the radiologist or observer to detect a signal in noise and this section will review the literature associated with these characteristics.

1.5.1 The effect of spatial frequency on signal detectability

The breast radiologist is faced with the task of detecting abnormalities of varying size and an observation from the studies cited previously is the variation of threshold with a change in the size of the signal (Barlow, 1957; Blackwell, 1946; Nagaraja, 1964). Size is, of course, related to spatial frequency; small signals correspond to higher spatial frequencies and larger signals to lower spatial frequencies (Hess, 2011). The response of the visual system to the variation in size can, therefore, be understood by studying its response to spatial frequency.

1.5.1.1 The units of spatial frequency in vision science and medical physics

This thesis straddles between the two distinct disciplines of visual science and medical physics and in doing so occasionally encounters different definitions or methodologies for the same concept. The definition of symmetry, to be discussed in section 2.5, is a case in point and reporting spatial frequency is another. In vision science visual angles are always used and spatial frequency is expressed in terms of the number of cycles (where, for a grating, one cycle consists of one light bar plus one dark bar) per degree subtended at the eye. Medical physics papers, however, will often report spatial frequency in cycles/mm (Bushberg, 2002, p. 269) or as a spatial dimension measured in mm or cm (for example: Bochud et al. (2004); Reiser et al. (2013)) or the number of pixels (for example: Judy et al. (1981)).

In vision science, visual angle is preferred as it enables a true measure of the image size on the retina and, thus, reflects the perceived size of a stimulus. For a stimulus measured in mm, cm or pixels the perceived size would depend upon the distance from the stimulus, however, given the stimulus size and viewing distance, the visual angle can be simply calculated.

1.5.1.2 The spatial contrast sensitivity function

Schade (1956) was the first researcher to measure contrast detection thresholds for the response of the human visual system to gratings of varying spatial frequency. This was achieved by using a television monitor to display vertical bands of varying width to represent spatial frequency at six different luminance levels and adjusting the contrast of the image until the bands were just visible, measuring the contrast at this point (Schade, 1956). Taking the reciprocal of the threshold contrast values at each spatial frequency Schade (1956) was able to describe the contrast sensitivity function for each luminance level and demonstrate the variation of sensitivity of the human visual system with spatial frequency. Using a similar approach Campbell and his colleagues (Campbell & Green, 1965; Campbell & Gubish, 1966) produced the more familiar, inverted U shaped, contrast sensitivity function, such as shown in Figure 1.14. The fall off in sensitivity at higher spatial
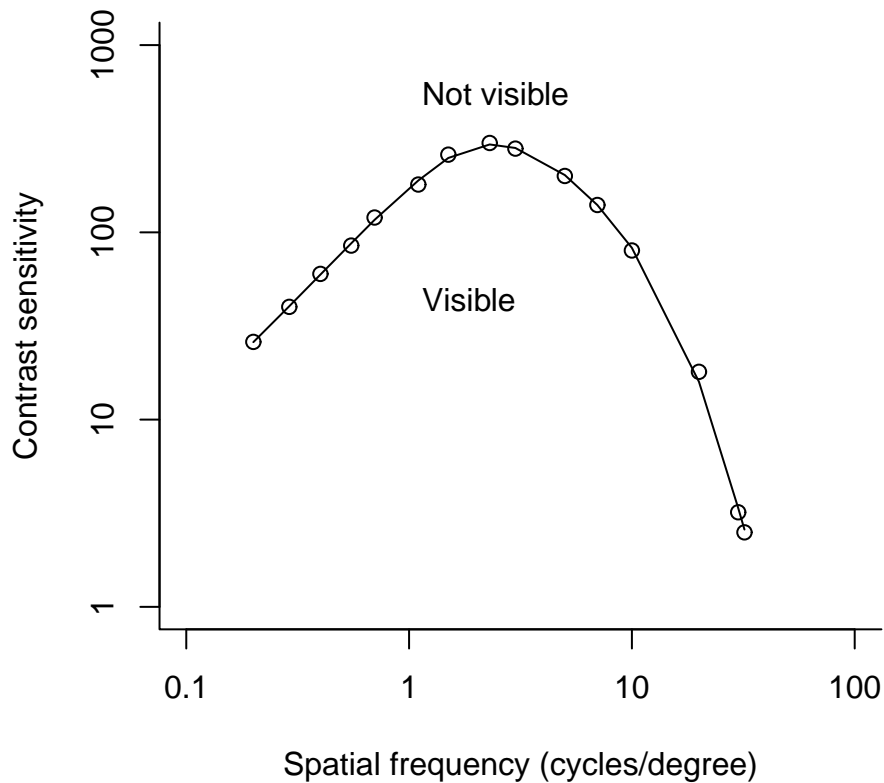
Figure 1.14: A typical contrast sensitivity function.

frequencies was expected and resulted from the deteriorating optical properties of the eye as the spatial frequency increased, whereas the fall off at lower spatial frequencies was the result of neural factors (Levin et al. (2011, p. 639); Wandell, 1995, p. 202).

The use of gratings became the predominant method for research into the spatial frequency response of the human visual system, providing a stimulus where the spatial frequency, spatial phase, orientation, and contrast can be independently varied. Spatial frequency refers to the number of luminance cycles, or bars (noting that each cycle consists of two bars, one light and one dark), that the grating repeats in one degree of visual angle; the greater the number of cycles, the higher is the spatial frequency. The spatial phase describes the shift of the grating with reference to a fixed point, for example, a grating that is a quarter of the way through its 360° cycle will have a phase shift of 90°. Orientation refers to the tilt of the grating which can be vertical, horizontal or oblique. Contrast is usually measured to determine the effect that one of the other parameters has on the sensitivity of the visual system. Examples of gratings are shown in figure 1.15
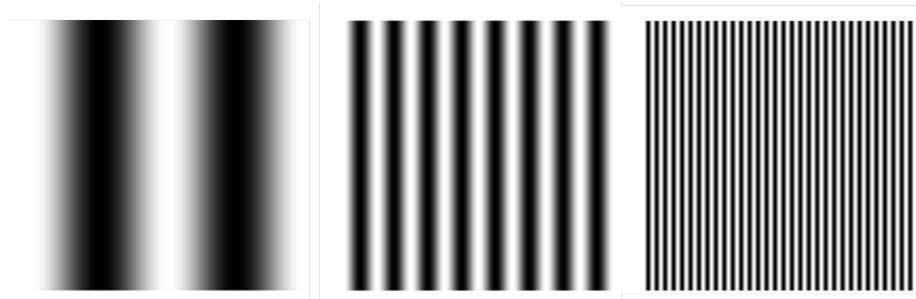
Figure 1.15: Examples of spatial frequency gratings; low frequency (left), medium frequency (centre) and high frequency (right).

1.5.1.3   Fourier analysis and multiple channels

Schade (1956) assumed that the contrast sensitivity function reflected the response of a single detector mechanism, sensitive to a broad range of spatial frequencies, with peak sensitivity at around 4 cycles per degree and tailing off at higher and lower frequencies. Campbell & Robson (1968) suggested that the function wasn't the result of a single channel but multiple channels each tuned to a specific frequency. They conducted trials using square wave gratings and from the Fourier analysis of a square wave (equation (1.61)) predicted that the fundamental component of the square wave would determine its visibility and it would have a contrast sensitivity $\frac{4}{\pi}$ greater than a same amplitude sine wave shown as follows.

Fourier series for a square-wave function:

$$f\left(x\right) = \frac{4}{\pi}\left\{\sin\frac{2\pi}{X} + \frac{1}{3}\sin 3\frac{2\pi}{X} + \frac{1}{5}\sin 5\frac{2\pi}{X} + ...\right\} \tag{1.61}$$

From (equation (1.61)), the fundamental component of the square wave is:

$$\frac{4}{\pi}\sin\frac{2\pi}{X}$$

And the same amplitude sine wave is:

$$\sin\frac{2\pi}{X}$$

This is exactly what they found with the ratio of $\frac{4}{\pi}$ consistently true down to a spatial frequency of 0.8 cycles per degree. They proposed that at higher spatial frequencies only the fundamental of the square wave would be visible, with the harmonics outside the spatial sensitivity range of the human visual system, producing a retinal image of a sine wave with an amplitude $4/\pi$ greater than a sine wave with the same amplitude as the original square wave. As the spatial frequency reduced they observed that the ratio of sensitivity to the square wave and the sine wave increased as the

69
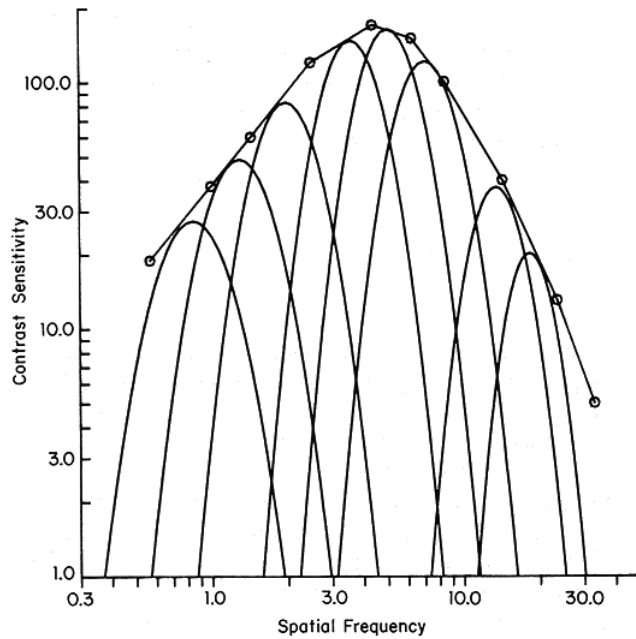
Figure 1.16: A schematic multiple channel model of spatial frequency detection.

higher frequency harmonics began to fall within the range of the visual system. The increased ratio, however, did not behave as predicted by the theoretical model of single detector with a broad band filter but rather, as a set of independent detectors each with its own narrow band filter and proposed this as evidence for the multiple channel hypothesis. A schematic example of a multiple channel model is shown in figure 1.16.

The multiple channel theory was supported by evidence from Blakemore & Campbell (1969) who used the method of neural adaptation to a specified stimulus as a test mechanism. They proposed that prolonged viewing of a high contrast pattern at a specified frequency would lead to neural adaptation and would reduce the sensitivity of the visual system to subsequently viewed patterns. The hypothetical result for a single detector mechanism would be seen as a uniform dip in sensitivity across the full spatial frequency range of the contrast sensitivity function, whereas, for a multiple channel detector mechanism, there would be a dip in sensitivity only in the range of the spatial frequency of the adaptation grating, as shown in the example in Figure 1.17. Blakemore & Campbell (1969) found the latter; adaptation only reduced the sensitivity in the vicinity of the spatial frequency of the adaption grating, thus providing strong evidence for the multiple channel hypothesis of Campbell & Robson (1968).

Further evidence in support of the multiple channel hypothesis is provided by Graham & Nachmias (1971) who used sine wave gratings with spatial frequencies separated by a ratio of 3:1, displayed separately as simple gratings and then together as complex gratings. They measured the

70
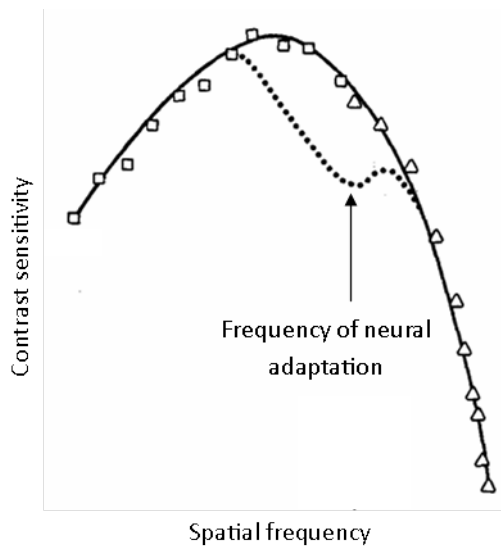
Figure 1.17: Example of neural adaptation showing the depressed contrast sensitivity around the frequency range of the adaptation stimulus.

contrast threshold for pattern detection for the simple gratings and then compared the contrast thresholds required for detection of the complex gratings. A single channel model for spatial detection would predict addition of the waveforms, and thus, a lower contrast detection for complex than simple gratings, even when the simple grating is below threshold. A multiple channel model would predict that each grating would be detected by a separate channel and, thus the complex grating would not be detectable unless one of its simple components was above threshold. The results supported the multiple channel theory with the ratio for the detection of a complex to a simple grating consistently falling close to one, supporting the idea that each frequency is detected separately (Graham & Nachmias, 1971).

The concept of independent channels has, more recently, been challenged with research suggesting a more complex picture of interacting neural networks (Wilson & Wilkinson, 2004, pp. 1062-1067). Theories such as contrast gain control (Heeger, 1992) and spatial pooling (Ellemberg et al., 1998; Wilkinson et al., 1997) suggest that there is interaction across visual channels to enhance individual channel responses. Nonetheless, the concept of multiple channels continues to provide a good model of the response to spatial frequency of the human visual system and, represents a suitable estimation for the purposes of this thesis.

1.5.2 The effect of temporal frequency on signal detectability

The introduction of an animated display mode in this thesis means that the visual system of the observer has to consider temporally modulated stimuli as well as spatially modulated. The temporal frequency of a stimulus represents how quickly that stimulus changes with time with a

low temporal frequency referring to a slow animation rate and a high temporal frequency referring to a high animation rate. In the same way that the response of the visual system to spatial stimuli can be characterised by a spatial contrast sensitivity function, its response to temporal stimuli can be represented by a temporal contrast sensitivity function, representing how the visual system responds to temporal variation at a range of frequencies.

1.5.2.1 The temporal contrast sensitivity function

The first temporal contrast sensitivity function was produced by de Lange (1958) who used close approximations to sinusoidal (the technology of the time made it impossible to produce pure sinusoids) and square waveforms that varied in their frequency and modulation. de Lange (1958) produced a temporal function, similar to the spatial function discussed in section 1.5.1.2, as shown in figure 1.18. de Lange (1958) found that sensitivity peaked at around 8Hz, falling rapidly as frequency increased until the critical flicker frequency (the frequency at which the flicker will appear to fuse into a continuous light) is reached between 50-70Hz. There is also a fall off in sensitivity as frequency decreases, however, this is not as steep as seen with increasing frequency.

As with spatial frequency, the temporal frequency contrast sensitivity function is thought to be underpinned by multiple channels with much of the evidence pointing towards three channels (Hess & Snowden, 1992; Mandler & Makous, 1984; Watson & Robson, 1981).

1.5.3 The spatio-temporal contrast sensitivity function

Whereas the temporal contrast sensitivity function as described by de Lange (1958) does not consider the effect of the spatial frequency characteristics of the signal, the studies by Hess & Snowden (1992); Mandler & Makous (1984); Watson & Robson (1981) investigating temporal channels did and recognised that there is an interaction between the spatial and temporal responses of the visual system. This interaction was investigated by Robson (1966) who measured the temporal contrast sensitivity functions at four spatial frequencies and showed that the visual system behaved differently to temporal modulation with different spatial frequencies. At high spatial frequencies, no decline was seen as temporal frequency reduced, however, at low spatial frequencies, a marked decline was observed (Robson, 1966). Kelly (1972, 1979) mapped the spatiotemporal function of the visual system from a range of measures demonstrating the interaction between the two characteristics.. Figure 1.19 shows the relationship between spatial and temporal characteristics, illustrating the visual system's band pass characteristic at low spatial frequencies and its low pass nature at high spatial frequencies. Whilst some have argued that the spatiotemporal contrast sensitivity function would suggest that the spatial and temporal mechanisms are inseparable (Kelly, 1984),
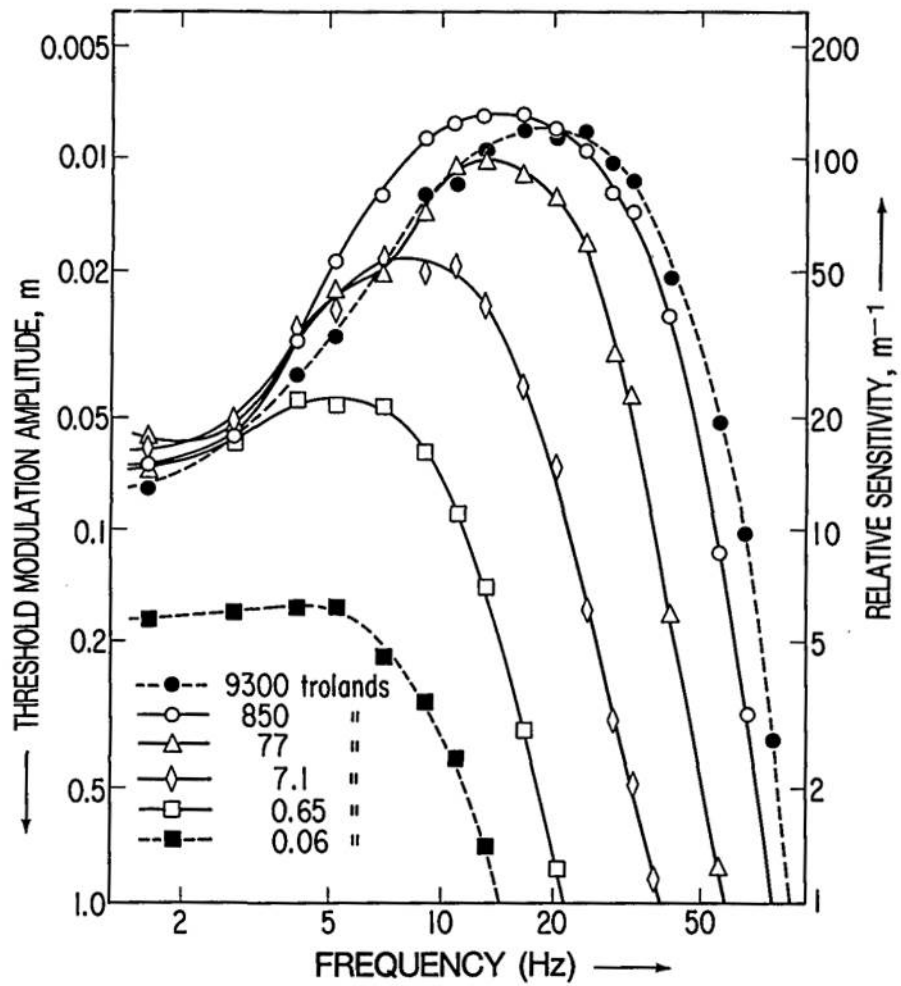
Figure 1.18: A typical set of temporal contrast sensitivity curves showing the variation of contrast sensitivity with temporal frequency measured at various mean background luminance levels (Stork & Falk, 1987).
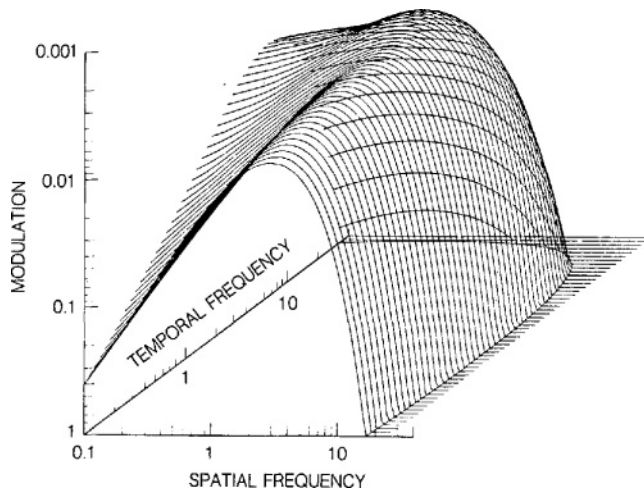
Figure 1.19: The human spatio-temporal contrast sensitivity function from Kelly (1979).

there is also a great deal of evidence to support the idea that spatial and temporal responses are mediated by separate mechanisms.

1.5.3.1 Sustained and transient channels

Tolhurst (1973) and Kulikowski & Tolhurst (1973) suggested that separate mechanisms could mediate the visual system's sensitivity to spatial frequency information and to temporal frequency information and they were among the first to use the sustained and transient description for these mechanisms. Kulikowski & Tolhurst (1973) used temporally modulated gratings that were either switched on and off or phase shifted by 180 degrees (alternated). The effect of this was to create a stimulus that was transient then static (on/off grating) or always transient (alternated grating). Because, the alternated gratings were always on, merely shifted by 180 degrees, in contrast to the on/off grating which was on for half of the time, the luminance with the alternated waveform was twice that of the on/off waveform and Kulikowski & Tolhurst (1973) argued that the observer should, therefore, be twice as sensitive to the alternated waveform. However, Kulikowski & Tolhurst (1973) found that this ratio only held true for the detection of flicker and not for the detection of stationary patterns and argued that this must be the result of two independent types of channel, which they referred to as movement analysers and form analysers, the first sensitive to transiency and the second, sensitive to form. Using waveforms modulated at 3.5Hz, Kulikowski & Tolhurst (1973) observed that at spatial frequencies below 5 cycles/degree the movement analysers were more sensitive and, above this, the form analysers became more sensitive. By repeating the experiments at different temporal frequencies, Kulikowski & Tolhurst (1973) were able to conclude that flicker and pattern were detected by separate channels, both with different responses to spatial and temporal frequency. They referred to these channels as sustained; most sensitive at

medium to high spatial frequencies with no fall off as temporal frequency reduced, and transient; most sensitive at low to medium spatial frequencies with a distinct fall off in sensitivity as temporal frequency reduced. The sustained channels were equally sensitive to alternated and on/off gratings as the channels are responsive to pattern, whereas the transient channels are more sensitive to the alternated grating as this is a continuously transient waveform, unlike the intermittent nature of the on/off grating.

Watson & Nachmias (1977) provided more formal definitions for the sustained and transient channels, proposing that the sustained channels were characterised by an impulse response of only one sign, either positive or negative and a maximum response at a temporal frequency of zero, whilst the transient channels had responses with equal positive and negative response and an integral of zero with no response at a temporal frequency of zero. As a purely sustained or purely transient system is unlikely to occur, they suggested that the channels would be either relatively sustained or relatively transient, dependent upon which characteristic predominated (Watson & Nachmias, 1977).

There remains considerable debate regarding whether the sustained and transient channels are separate or whether both spatial and temporal stimuli are processed by a single mechanism, however, as with the debates surrounding the independence of channels for spatial frequency detection, the concept of sustained and transient channels provides a suitable model of the response to spatial-temporal stimuli that meets the purposes of this thesis.

1.6 Types of Noise Backgrounds and Other Concepts Relevant to this Thesis

1.6.1 Gaussian white noise

So far, consideration has only been given to the detection of signal in Gaussian white noise, which is an important noise condition as, when used with the use of a signal known exactly paradigm, it affords the simplest detection problem; the "...simple versus simple hypothesis..." (Kay, 1998, p. 94). This enables the property of interest to be isolated and examined whilst minimising the possibility of confounds that could be introduced by more complex images. GWN is a noise where the pixel values have a normally distributed probability density function and are independent and identically distributed (uncorrelated) (Lu & Dosher, 2014, p. 34). Figure 1.20 shows an example of Gaussian white noise and its power spectral density showing an equal power distribution across all frequencies (showing just one dimension - the power spectral density is two dimensional and so
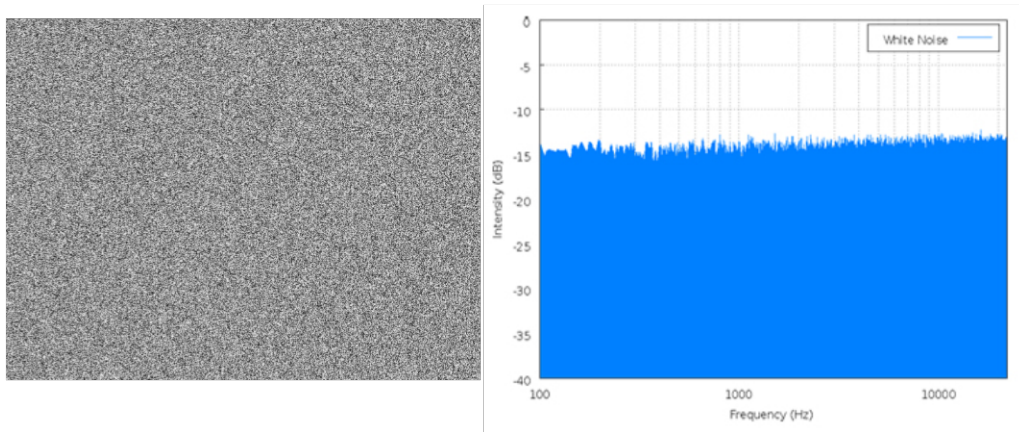
Figure 1.20: Example of Gaussian white noise (left) with its power spectral density plotted on a log-log scale (right).

the frequency components will be reflected about the Y-axis when plotted).

1.6.2   Signal known exactly with Gaussian white noise

Signal known exactly (SKE) refers to an experimental set up where the observer knows the signal's profile, size and location within an image (Abbey & Eckstein, 2010, p. 244). The Gaussian white noise / SKE model has been widely used in visual psychophysics to investigate precise visual properties such as the efficiency of human visual signal discrimination (Burgess et al., 1981), the relative efficiencies of first-order and second-order vision (Manahilov et al., 2005) and classification images (Abbey & Eckstein, 2002). These are just three examples from 1000s of other studies using a Gaussian white noise / SKE model, which include four of the most influential, already discussed, from Burgess and colleagues in their series of Visual Signal Detection papers (Burgess & Ghandeharian, 1984a,b; Burgess, 1985; Burgess & Colborne, 1988). There are fewer examples of a Gaussian white noise / SKE model being used in the more applied domain of mammography research, where the preference is for synthetic or real mammogram noise backgrounds, however, the value of the Gaussian white noise/SKE model is to establish the theoretical baseline for a phenomenon before testing it in a more realistic setting. This is the rationale taken by this thesis and closely reflects the approach taken by Reiser & Nishikawa (2006), who investigated the detection and discrimination of simulated microcalcifications using the performance achieved with a white noise background as a comparison to the performance for the same tasks using real mammogram backgrounds.

1.6.3 Synthetic signals

Lesions within a mammogram can have a variety of shapes and appearances (Zonderland & Smithuis, 2013) and, therefore, the use of real lesions as signals complicates the use of a SKE approach. Synthetic lesions are often used for this reason. Reiser & Nishikawa (2010) used spherically symmetric 3D designer nodules with radii ranging from 0.025 cm to 0.8 cm and, similarly, Suryanarayanan et al. (2005) generated two-dimensional mass like lesions for their research into the detection of simulated lesions on digital mammograms. Castella et al. (2009) and Castella et al. (2007) also used a spherical mass superimposed upon real mammogram backgrounds. Several researchers have used a signal with a Gaussian profile (Abbey & Barrett, 2001; Chakraborty, 2006; Johnson et al., 2002), sometimes referred to as a Gaussian bump or blob, and this is the signal profile that will be used throughout this thesis. The use of a single signal profile simplifies the experimental procedure making the use of naïve observers more practical and this profile is a good match to a typical lesion (a small mass) found in mammograms (see also section 2.9).

1.6.4 Power law noise

As outlined in section 1.6.1, Gaussian white noise enables the simplified analysis of the theoretical performance of the ideal observer; however, real mammograms contain a more complex noise spectrum and to enable extrapolation of results to real mammograms requires images that reflect the properties of real mammograms. There are two main components to the noise in x-ray mammograms; firstly, quantum noise, which can be described as background fluctuations arising from the finite number of x-ray quanta as well as non-signal noise introduced by the imaging system electronics and, secondly, anatomical noise, which refers to the background fluctuations resulting from the normal breast tissue and anatomy within the breast but not including the signal (Samei et al., 2000, p. 656; Bochud et al., 2004). Modern imaging systems minimise the level of quantum noise and, as a result, anatomic noise is the limiting and most important component of the overall noise and the component that determines signal detection performance (Samei et al., 2000, p. 658; Burgess et al., 2001). Burgess & Judy (2007) suggest that the anatomic noise is close to pure random noise following a power spectrum of the form $1/f^\beta$.

Bochud et al. (1999) found that the exponent β varies from 1.5 to 4 in their sample of mammograms and, along with other researchers, suggest that the average value for β is approximately 3 (Burgess et al., 2001; Kierkels et al., 2012). This would suggest that a random noise generated background with a power law spectrum of $1/f^3$ would be a suitable model for mammogram noise, however, Burgess (2010, p. 53) contends that real mammograms can only be considered as being partly random noise with the remainder being the non-random features of the breast tissue and
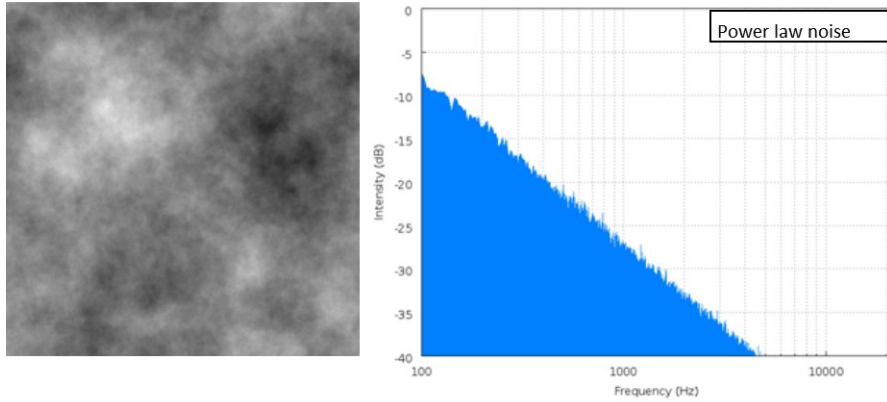
Figure 1.21: An example of power law noise (left) with its power spectral density plotted on a log-log scale (right)

anatomy and a number of studies support this view (Bochud et al., 2004, 1999; Burgess et al., 2001). Burgess et al. (2001) included a comparison of the performance of human observers when detecting simulated tumours in either real mammogram backgrounds or simulated mammogram backgrounds with a $1/f^3$ power law spectrum and found that, whilst the simulated tumour was detected with a lower contrast threshold in real mammograms than in simulated mammograms, there was a qualitative agreement in the way they performed. The conclusion drawn was that the real mammograms behaved as if they contained a mixture of random noise and deterministic components (the anatomic structure), which assisted the observers in the decision making process (Burgess et al., 2001). In the same study, when conducting a search task, the results indicated no difference between the contrast thresholds for real mammograms and simulated mammograms suggesting that for a search task the two backgrounds could be considered as equivalent (Burgess et al., 2001). Despite the differences observed in the detection experiments, Burgess et al. (2001) nonetheless concluded that the statistical properties of power law noise backgrounds are similar to those of real mammograms, making them suitable for theoretical research. This is supported by Reiser et al. (2013), whose study compared the performance of human and model observers carrying out a signal known exactly detection task in both power law noise backgrounds and real breast backgrounds. The real breast backgrounds were two dimensional 384 x 384 pixel sections from tomosynthesis acquired x-ray images and the simulated mammograms were generated using a power law noise model with a simulated designer nodule used as the signal to be detected (Reiser et al., 2013). Human observer performance for the detection of the signal as it varied in size was very similar for both simulated and real breast backgrounds and closely matched the performance of the pre-whitening observer model (which will be discussed in section 1.6.6), thus supporting the use of a power law noise background for simulating breast tissue (Reiser et al., 2013).

78

The research cited here suggests that noise backgrounds with a $1/f^3$ power law spectrum provide an acceptable level of similarity to real mammograms for detection and search experiments, as conducted in this thesis, and, therefore, this thesis will progress from using Gaussian white noise backgrounds to using noise backgrounds with a $1/f^3$ power law spectrum as a realistic alternative to real breast images.

1.6.5 Mammogram backgrounds

As discussed in section 1.6.4, the two main types of noise in medical images are quantum noise and anatomical noise with anatomical noise being the limiting and most important component of the overall noise. The anatomical features present in real mammograms can affect the decision making ability of the observer and Samei et al. (2000, p. 660-678) identify three key effects by which anatomical structure impacts upon the observers ability to detect abnormalities: the search effect, the global effect and the local effect.

The search effect refers to the way anatomical structure not only influences the observers overall impression of the image, but also the way they subsequently scan the image which may result in the failure to detect an abnormality, for example, areas of increased breast density have been shown to affect the way the radiologist scans the mammogram with more attention on denser areas than less dense areas resulting in poorer detection rates for abnormalities in the less dense areas (Mousa et al., 2014).

The global effect refers to the tendency of normal anatomic structure to mimic true abnormalities, thus creating false signals and evidence for this can be seen in the study by Zanca et al. (2007), where the characteristics of fatty tissue were shown to have an impact, significantly increasing the number of false positives in this type of breast tissue.

Samei et al. (2000) also suggests that local effects have a major impact on the detection of abnormalities and this refers to the way that local structure can obscure or affect the appearance of an abnormality such that it is camouflaged and Samei et al. (1998) conducted an experiment to quantify the effect of this. By varying the location of abnormalities in relation to the anatomic structure in chest radiographs, Samei et al. (1998) found a significant effect of the proximity of the anatomic structure and that varying the proximity had a similar effect to varying the size of the abnormality.

Although synthetic mammograms, as described in section 1.6.4, can provide close statistical representations of real mammograms, the evidence above suggests that, having established a sound theoretical basis for the detection of simulated lesions in synthetic mammograms, it is necessary to then assess the performance of the techniques with real mammograms. This thesis will, therefore,

progress onto testing both the symmetric, side-by-side presentation and animated display modes using real mammogram backgrounds.

1.6.6   Pre-whitening

For the detection of a signal in Gaussian white noise, the ideal observer was shown to cross-correlate the received image with a template of the signal Green & Swets (1966, p. 165) and use Bayesian reasoning to select the option with the highest *a posteriori* probability (Burgess & Ghandeharian, 1984a; Burgess, 1985). In power law noise, the theory is the same with one extra element; the power law noise has a low-pass spectrum and, therefore, is spatially correlated. Therefore, the ideal observer will first pre-whiten the noise field prior to cross-correlating with the signal template (Bochud, 2013, pp. 153-164; Burgess, 2010, pp. 38-40; Burgess & Judy, 2007). Pre-whitening (or simply whitening) a stimulus by decorrelating the data it contains has been proposed as a method of redundancy reduction, to remove unnecessary information and improve the efficiency of the visual system (Hyvärinen et al., 2009, p. 126). Pre-whitening, thus, reduces the spatially correlated noise to white noise (Hyvärinen et al., 2009, p. 126). This process was suggested as far back as 1961 by Barlow, who proposed that the human visual system simplified the natural scene by removing information that was positively correlated (Barlow, 2001). This information would be predictable and effectively redundant and its removal would enhance the efficiency of the human visual system. It would be expected that evolutionary pressures would lead the visual system to be maximally efficient at decorrelating stimuli within a power spectrum corresponding to that of natural scenes. Field & Brady (1997) showed that natural scenes exhibit a power spectrum ranging from $1/f^{1.2}$ to $1/f^{3.2}$ and, indeed, there is a long line of research that suggests that the human visual system has evolved to operate with maximal efficiency in natural scenes (Atick, 1992; Field, 1987; Tolhurst & Tadmor, 2000). There is also considerable research into the question of whether human observers can pre-whiten spatially correlated images and this will be discussed next.

Research carried out into pre-whitening suggests that the human visual system is able to pre-whiten images with a low pass spectrum (Abbey & Barrett, 2001; Abbey & Eckstein, 2007; Burgess, 1999; Burgess & Judy, 2007; Rolland & Barrett, 1992), such as encountered in natural scenes, but not with noise with a high pass spectrum (Myers, 1985; Myers et al., 1985). In the study carried out by Burgess (1999), the ability of the human observer to partially pre-whiten low pass filtered noise was demonstrated by comparing the performance of human observers against three observer models; the pre-whitening observer, the non pre-whitening observer and the partial pre-whitening (or Hotelling) observer. Burgess (1999) used both a low pass Gaussian filter and a low pass power

law filter to create noise backgrounds with a low pass Gaussian or power law noise spectrum and carried out a 2AFC trial for the detection of Gaussian profile nodules with the finding that the human observer results were much better fitted by the pre-whitening models than the non pre-whitening model. This can be seen in Figure 1.22, which shows one example of the data plotted by Burgess (1999) for observer performance in power law noise. The evidence from Burgess (1999) thus supports the suggestion that the human visual system is able to pre-whiten images with a low pass spectrum.
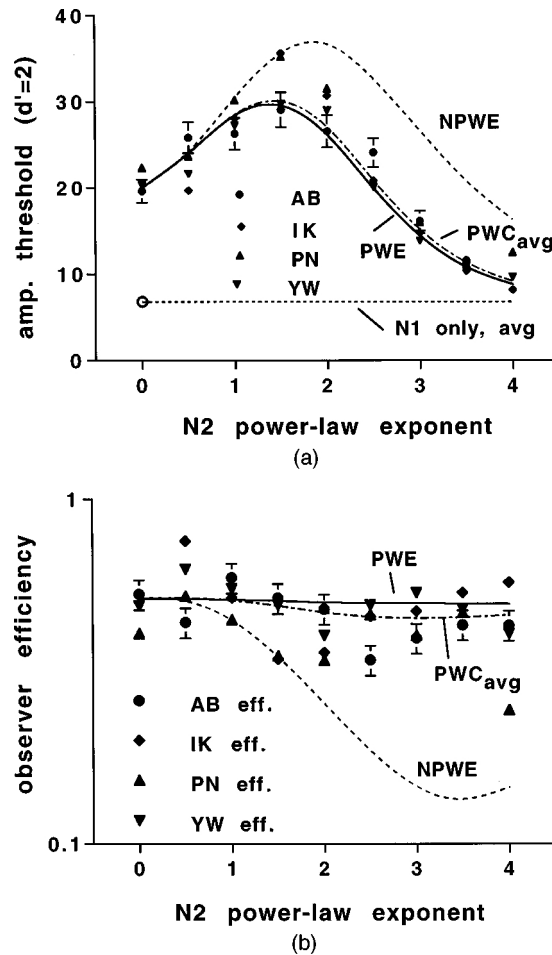
Figure 1.22: Example Figure reproduced from Burgess (1999) showing the variation of performance (top figure) and efficiency (lower figure) with the power law exponent (β). The theoretical performances of the non-pre-whitening observer ($NPWE$ - dashed line), the pre-whitening observer (PWE - solid line) and the partial pre-whitening observer (PWC_{avg}- dash-dot line) are shown. Human observer performance is shown by the filled symbols and it is clear that human performance is much better predicted by the pre-whitening models than by the non-pre-whitening model.

The findings of Burgess (1999) are further supported by the results of the second part of the study by Myers et al. (1985), where four transfer functions were used to create images with background noise that ranged from a low pass power spectrum through to increasingly high pass power spectra. Myers et al. (1985) used human observer efficiency, as defined by Burgess et al.

(1982) and as shown in equation 1.62, applied to the task of detecting a disc signal in a noise background.

$$\eta = \left(\frac{d'}{SNR_{ideal}} \right)^2 \tag{1.62}$$

where:

η = efficiency

d' = detectability index

SNR_{ideal} = signal to noise ratio for the ideal observer

Manipulating equation 1.62 as follows:

$$\eta = \left(\frac{d'}{SNR_{ideal}} \right)^2 \left(\frac{SNR_{npw\,ideal}}{SNR_{npw\,ideal}} \right)^2 \tag{1.63}$$

where:

$SNR_{npw\,ideal}$ = signal to noise ratio for the non pre-whitening ideal observer

$$\eta = \left(\frac{d'}{SNR_{npw\,ideal}} \right)^2 \left(\frac{SNR_{npw\,ideal}}{SNR_{ideal}} \right)^2 \tag{1.64}$$

$$\eta = \eta^{npw} \times observer\,reconstruction\,penalty \tag{1.65}$$

η^{npw} = efficiency of human observer relative to an ideal non pre-whitening observer

The overall efficiency (η) is the product of the efficiency of a human observer relative to an ideal non pre-whitening observer (η^{npw}) and a factor that is proportional to the ability of the observer to pre-whiten the noise. This manipulation enabled Myers et al. (1985) to observe the effect of changing the power spectra of the background noise in the images; if η^{npw} remained constant then they could show that any reduction in efficiency must result from a reduction in the observer's ability to pre-whiten the noise and this is exactly what they found. The results showed that human observers exhibited the same level of efficiency with power law noise with a low pass power spectrum as with the low pass Gaussian white noise, with their efficiency falling as the power spectrum became more and more high pass, strongly supporting the idea that human observers are able to pre-whiten images with a low pass power spectrum but not with a high pass spectrum (Myers et al., 1985).

The ability of the human observer to pre-whiten correlated backgrounds is not universally supported and research by Judy (1996) found that the performance of human observers for the detection of clusters of simulated calcifications was worse in lumpy backgrounds (a type of synthetic

background containing correlated noise) than in a uniform, uncorrelated, noise background. Judy (1996) found that the detectability index, d', was significantly lower with lumpy backgrounds suggesting that the observers were unable to pre-whiten the noise.

Whilst the evidence supporting the ability of the human observer to utilise a pre-whitening strategy is equivocal, a reasonable conclusion is that the human observer can, at least, partially pre-whiten correlated noise fields. As the images typically used in x-ray mammographic research, and as used in this thesis, fall within a low-pass spectrum, whilst accepting that the human observer may be able to only partially pre-whiten, thus reducing their efficiency, the ideal observer strategy of pre-whitening the noise fields is, nonetheless, consistent with this research.

1.7 The Presentation of Paired Images

1.7.1 The ideal observer strategy for detecting a signal in paired images

The use of an ideal observer approach requires that we identify the task to be carried out, identify the stimulus within which the task will be performed and, finally, identify the optimum computation for conducting the task (Geisler, 2011; Kersten & Mamassian, 2009). For this thesis, the task is to detect a signal in one image of a pair of images. The signal is a Gaussian blob and two paired images are presented with a measurable level of correlation between them. One image is signal plus noise; the other is noise alone. Perhaps the simplest, and most obvious, computation that could be carried out would be to subtract, or difference, the two images and then cross-correlate the difference image with the observer's template of the signal to be detected (Burgess & Colborne, 1988) (see section 1.3.1 for a discussion of cross-correlation). If the two images were identical (apart from the embedded signal to be detected), with an inter-image correlation of one, subtraction would remove all the background noise, leaving only the signal in one image and performance would be perfect. As the inter-image correlation declines, and the level of difference between the two images increases, less and less noise would be removed and performance would also decline.

A less obvious, but more sophisticated, method of removing redundant information is to decorrelate (or pre-whiten) the image pair (Hyvärinen et al., 2009, pp. 126-130). Pre-whitening, in relation to removing redundancy *within* an image, was discussed in section 1.6.6, however, the same technique of decorrelation can be applied to an image pair by decorrelating *between* the two images. It should be noted that decorrelation and pre-whitening (or simply whitening) refer to the same operation Hyvärinen et al. (2009, p. 126) and the two terms are interchangeable. However, in this thesis, to minimise any potential confusion, we will use the term pre-whitening to refer to the redundancy reduction carried out *within* an image and the term decorrelation will refer to the

redundancy reduction carried out *between* two images.

Each process can be understood by considering the task of detecting a signal known exactly in one image of a pair of Gaussian white noise images with an inter-image correlation of between 0 and 1. Figure 1.23 shows a simple 16 x 16 pixel white noise image consisting of 256 pixels.
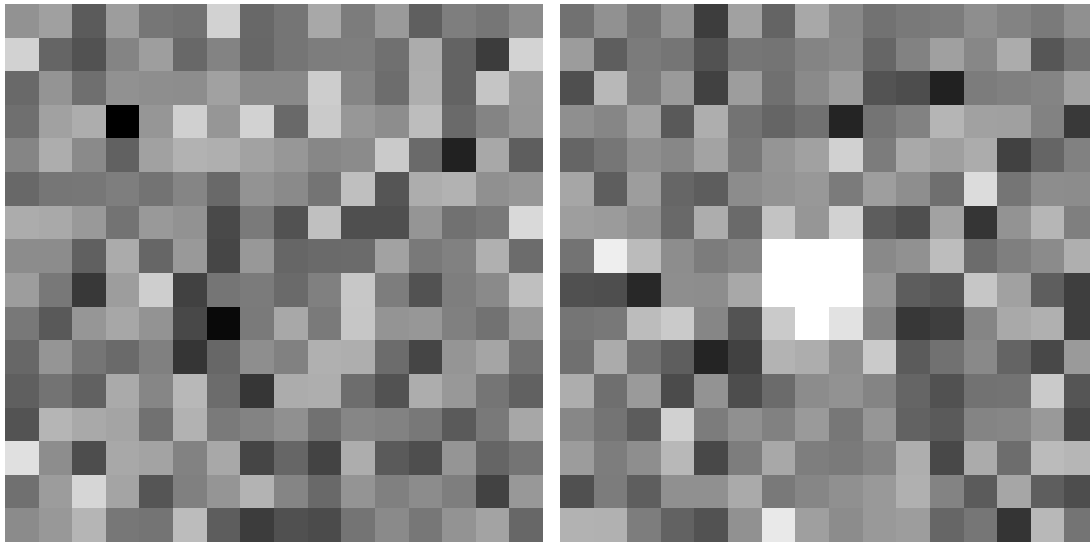
Figure 1.23: Example images for the illustration of the decorrelating process. Each image is a 16 x 16 pixel white noise image with 256 pixels and each pixel has a numerical value from 0 to 255 representing the intensity value of the pixel. The right image contains a blob signal, shown well above threshold.

Each image shown in Figure 1.23 can be displayed in a numerical format, as shown in Figure 1.25. Each pixel has now been replaced by a numerical value representing its grey level and can take a value from 0 to 255, where 0 is pure black and 255 is pure white, as shown in Figure 1.24.

Figure 1.24: Graphic showing the grey shade intensity represented by the range of pixel values from 0 to 255.

The signal has been placed in the right image and in Figure 1.23 can be seen as a bright white area which in Figure 1.25 can be seen as pixel values of, or very close to, 255.

146	160	91	156	118	114	209	104	117	167	124	154	95	126	118	139
209	101	82	132	158	104	131	103	117	125	126	112	172	99	60	211
105	148	111	145	142	140	161	137	137	205	133	110	174	99	196	152
111	161	172	0	150	208	150	210	105	202	151	141	188	106	132	151
133	173	138	97	161	178	174	162	151	135	139	202	106	34	168	95
104	118	119	126	115	133	105	147	138	116	191	85	172	178	144	150
172	169	153	115	152	145	73	122	82	192	80	79	149	113	121	217
140	140	96	171	102	153	71	152	102	106	107	162	121	129	176	108
157	120	56	157	205	65	117	123	106	128	198	125	82	126	138	190
120	89	150	167	147	72	10	122	168	122	198	148	151	128	114	151
103	149	117	106	128	53	103	142	129	176	174	108	69	149	164	114
95	115	97	170	134	184	109	54	172	173	110	82	173	152	117	98
83	181	168	164	113	178	122	131	144	114	134	130	120	90	121	167
223	141	77	168	163	130	168	69	102	67	173	91	78	149	101	116
112	156	214	164	85	128	149	178	133	105	147	128	140	126	66	152
139	152	180	119	116	187	93	60	80	76	116	138	120	147	163	103

113	145	118	149	61	161	101	168	136	114	121	124	141	131	122	144
156	94	123	117	82	118	116	131	138	102	130	160	135	172	86	114
79	184	125	154	66	157	112	138	157	83	77	34	124	128	132	162
143	134	162	89	174	115	101	113	37	116	130	181	161	159	130	58
101	118	143	135	163	120	149	161	208	124	169	159	172	65	101	129
166	94	157	102	93	140	147	152	123	158	142	111	219	117	141	140
158	155	143	106	173	106	194	150	209	93	79	162	53	147	182	127
115	237	187	141	124	134	254	255	255	137	145	189	108	127	139	175
80	79	41	142	141	167	255	255	255	149	93	87	197	160	94	62
117	120	188	201	134	83	201	255	226	133	55	62	133	169	176	62
112	171	114	94	36	66	179	173	143	202	91	113	137	101	72	154
172	111	154	75	146	77	107	140	146	132	102	82	114	115	202	85
135	116	92	208	124	144	129	154	119	150	98	90	132	134	153	72
156	125	142	183	72	125	168	126	122	132	174	72	171	109	187	182
80	124	94	144	144	169	121	134	144	153	175	131	92	166	94	77
178	177	125	99	83	145	233	158	140	153	157	102	118	52	183	116

Figure 1.25: The white noise images from Figure 1.23 displayed in numerical format. The shade of grey of each pixel has been replaced by a numerical value representing its intensity. The left image consists of a noise only background and the right image consists of a signal added to a noise background.

r_1	r_2
146	113
160	145
91	118
156	149
118	61
114	161
209	101
104	168
117	136
167	114
124	121
154	124
95	141
126	131
118	122
139	144
209	156
101	94
82	123
132	117
158	82
⋮	⋮
147	52
163	183
103	116

Figure 1.26: Each image represented as a vector with the pixel values placed in two columns each with their number of rows equal to the number of pixels in the image. Note: for clarity only 24 rows are shown, whereas the actual vector for these images would be 256 rows long. The left vector (r_1) consists of noise only and the right vector (r_2) consists of the signal added to a noise background.

Each image can be flattened into a vector with the pixel values placed in a column with the number of rows equal to the number of pixels in the image, as shown in Figure 1.26. The left vector (r_1) consists of a noise only background and the right vector (r_2) consists of the signal added to a noise background.

The observer therefore receives r, which consists of two columns, r_1 and r_2, and for the case of a signal in the right image this can be illustrated as follows:

$$r_1 = n_1 \tag{1.66}$$

and

$$r_2 = s + n_2 \tag{1.67}$$

where:

$n_1 =$ noise in left image

$n_2 =$ noise in right image

$s =$ signal

$r_1 =$ left stimulus image

$r_2 =$ right stimulus image

1.7.1.1 Differencing the images

Considering the method of redundancy reduction of differencing the two images by subtracting one from the other, the two columns of r_1 and r_2 are subtracted and the resulting differenced columns are then cross-correlated with the signal (Kay, 1998, p. 106). Thus the observer computes:

$$r_{1\,diff} = r_1 - r_2 \tag{1.68}$$

$$r_{2\,diff} = r_2 - r_1 \tag{1.69}$$

where:

$r_{1\,diff} =$ resulting r_1 image following subtraction of r_2

$r_{2\,diff} =$ resulting r_2 image following subtraction of r_1

Using Equation 1.44 on page 57 we can determine the value of the test statistic resulting from the cross correlation of each differenced column:

$$T = (r_{diff})^t\, s \qquad\qquad (1.70)$$

where:

T = Test statistic following cross correlation with r_{diff}

s = signal template

t denotes transpose

If $T > criterion$, decide yes, signal present, else, no, signal not present.

1.7.1.2 Decorrelation of the images

The columns of r_1 and r_2 are correlated, with an inter-image correlation between 0 and 1, and have a 2 x 2 covariance matrix K. r is a matrix having columns r_1 and r_2, as shown in Equations 1.66 and 1.67. To decorrelate we multiply r by the inverse of the covariance matrix K^{-1} and the resulting decorrelated columns are then cross-correlated with the signal, as in section 1.7.1.1, (Kay, 1998, p. 106). Thus the observer computes:

$$r_{dec} = rK^{-1} \qquad\qquad (1.71)$$

where:

r = matrix having columns r_1 and r_2

r_{dec} = decorrelated image of r

Using Equation 1.44 we can determine the value of the test statistic resulting from the cross correlation of each differenced column:

$$T = (r_{dec})^t\, s \qquad\qquad (1.72)$$

where:

T = Test statistic following cross correlation with r_{dec}

s = signal template

t denotes transpose

If $T > criterion$, decide yes, signal present, else, no, signal not present.

1.7.1.3 Simulations

Simulations were conducted to compare the theoretical performance of the differencing observer, the decorrelating observer and a non-decorrelating observer. The differencing and decorrelating observers are described above. The non-decorrelating observer is an observer that does not carry

out any redundancy reduction measures on the image pair, simply cross correlating the original image columns with the signal template.

For the simulations the images were Gaussian white noise with a standard deviation of 0.22 and the signal was a Gaussian blob. Simulations were run for each type of observer at inter image correlations of 0, 0.25, 0.5, 0.75, 0.9, 0.95, 0.99 and 1. Taking the level of d' at detection threshold to be 1, the contrast of the blob was adjusted for each simulation until a d' value of close to 1 was achieved for each type of observer. A minimum of 10 simulations with 6000 trials per simulation were carried out to obtain an accurate prediction for the contrast threshold for each type of observer at each correlation level.

Figure 1.27 shows the performance of each type of observer plotted against the inter image correlation. The non-decorrelating observer, who doesn't utilise the correlation between the images, not surprisingly, exhibits a very flat profile, with no improvement in performance as correlation increases. The decorrelating observer is able to use the image pair correlation. The decorrelating observer shows a performance level equal to that of the non-decorrelating observer at an image pair correlation of zero, however, shows a reduction in the contrast threshold required to detect the signal as correlation increases, demonstrating errorless performance with a correlation of 1. The differencing observer also exhibits errorless performance at a correlation of 1, and shows the ability to utilise correlation, however, at a correlation of zero, the performance of the differencing observer is worse than both the decorrelating and non-decorrelating observers.

1.7.1.4 Observer performance

Observer performance is calculated in the standard way (Green & Swets, 1966, p. 165; Kay, 1998, p. 102; Macmillan & Creelman, 2005, p. 121) using the following notation:

E = signal energy= $\sum s_1^2$

σ^2 = noise variance

d' = signal detection theory measure of performance

ρ = image pair correlation

Non-decorrelating observer performance The performance of the non-decorrelating and cross-correlating observer in the 2AFC detection task where the observer is unable to utilise the correlation between the columns of n is given by:

$$d' = \sqrt{\frac{2E}{\sigma^2}} \tag{1.73}$$

(Green & Swets, 1966, p. 165; Kay, 1998, p. 102; Macmillan & Creelman, 2005, p. 121)
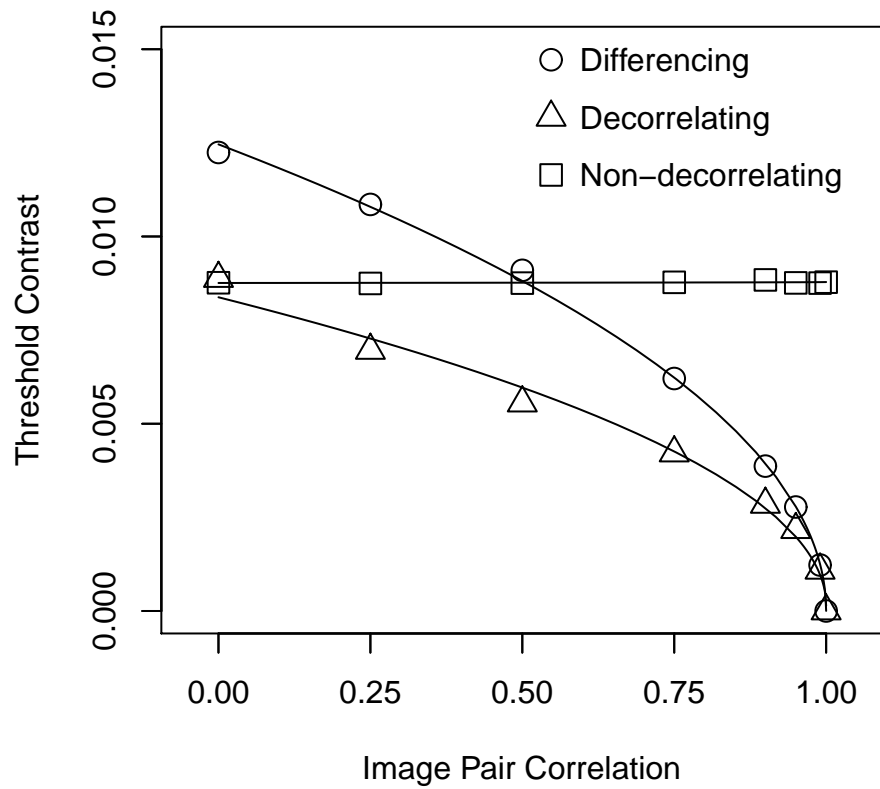
Figure 1.27: Results of simulations for each type of theoretical observer plotting observer contrast threshold against image pair correlation for the detection of a signal in paired images. The type of observer is shown in the legend.

Taking the level of d' at detection threshold to be 1, the energy threshold E_t is:

$$E_t = \frac{\sigma^2}{2} \tag{1.74}$$

The contrast threshold is equal to some constant a times the square root of the energy threshold. Therefore, for the non-decorrelating observer the contrast threshold C_t is:

$$C_t = a\sqrt{\frac{\sigma^2}{2}} \tag{1.75}$$

$$Non-decorrelating\ observer\ C_t = \frac{a\sigma}{\sqrt{2}} \tag{1.76}$$

Decorrelating observer performance   The performance of the decorrelating observer, when the noise has correlation ρ is given by:

$$d' = \sqrt{\frac{2E}{(1-\rho)\sigma^2}} \tag{1.77}$$

(Kay, 1998, p. 112)

Taking the level of d' at detection threshold to be 1, the energy threshold E_t is:

$$E_t = \frac{(1-\rho)\,\sigma^2}{2} \tag{1.78}$$

The contrast threshold is equal to some constant a times the square root of the energy threshold. Therefore, for the ideal observer the contrast threshold C_t is:

$$Decorrelating\ observer\ C_t = a\sqrt{\frac{(1-\rho)\,\sigma^2}{2}} \tag{1.79}$$

$$Decorrelating\ observer\ C_t = \frac{a\sigma\sqrt{1-\rho}}{\sqrt{2}} \tag{1.80}$$

Differencing observer performance From the simulation values, the contrast threshold of the differencing observer, when the noise has correlation ρ, appears to be a factor of $\sqrt{2}$ greater than the decorrelating observer. The values for the decorrelating observers contrast threshold can, therefore, be estimated from the following:

$$Differencing\ observer\ C_t = a\sqrt{(1-\rho)\,\sigma^2} \tag{1.81}$$

$$= a\sigma\sqrt{1-\rho} \tag{1.82}$$

Thus, for the differencing and decorrelating observers, as the correlation between the two noise fields increases, the detectability of the signal rises. In the limit, when the two noise fields are identical, d' is infinite and performance is perfect (Kay, 1998, p. 112). However, when measuring the contrast threshold, the differencing observer is less efficient than the decorrelating observer by a factor of $\sqrt{2}$. For the non-decorrelating observer, who is unable to use the image pair correlation, increasing the correlation between the two noise fields has no impact on the detectability of the signal.

From these simulations we conclude that the optimum strategy and, therefore, the strategy of the ideal observer, will be to decorrelate the image pair and then cross correlate the decorrelated image with the template of the signal to be detected, deciding signal present if the internal response exceeds some criterion.

1.7.2 Paired noise fields presented in a side by side display

Breast radiologists often view images in pairs so it is important in the context of this thesis to consider research that has been conducted using this mode of display. There does not appear to be any direct empirical evidence to support the presentation of mammograms in a symmetric side-by-side protocol, rather the practice seems to be based upon the intuitive notion that violations of symmetry will be easily detected, an area that will be explored further in section 2.5.

Not only does there appear to be no evidence to support the mirror symmetric presentation of mammograms, but neither does there appear to be much psychophysical research into the detection of signals in symmetrically paired images. One example is from Burgess & Colborne (1988), where the researchers used same and different noise fields (equating to paired noise fields with an inter-image correlation of one (same noise) and zero (different noise), respectively, in the context of this thesis) to estimate the ratio of internal to external noise (see section 1.1.5). The experiment

used this ratio to estimate the internal noise across a range of external noise strengths, however, more pertinent to the context of this thesis is the effect on observer performance of using same or different noise fields and the quoted value for one example value of signal to noise ratio, found $d'_1 = 2.0$ for the same noise fields and $d'_2 = 1.26$ for the different noise fields (Burgess & Colborne, 1988). Assuming this difference is significant (which wasn't specified), this result suggests that the observers were able to use (or, at least, partially use) the correlation between the images enabling their performance to improve as the similarity of the background fields improved.

Ahumada & Beard (1997) used the same methodology of same/different noise as Burgess & Colborne (1988) to investigate theoretical models for signal detection. They used three noise conditions, firstly a "fixed" condition where the images in each pair had identical noise backgrounds and the same pair were used throughout. Secondly, a "random-fixed", condition where each pair within a trial had the same noise background but different pairs were used across trials. Finally, for the "random" condition each image of a pair had different noise backgrounds and different pairs were used throughout the experiment. In relation to this thesis, the "random-fixed" condition would equate to the presentation of an image pair with a correlation of one between the images and the "random" condition would equate to the presentation of an image pair with a correlation of less than one between the images (it is surmised that the correlation would be closer to zero, however, this was not reported) (Ahumada & Beard, 1997). Unlike Burgess & Colborne (1988), Ahumada & Beard (1997) did not find any improvement in performance with the same noise images finding no significant difference between the two conditions which suggests that the observers were not able to use the correlation present between the images in the "random-fixed" condition, even though the images were identical, with a correlation of 1.

A similar approach, with similar findings was carried out by Watson et al. (1997) in their investigation into image quality and entropy masking. Watson et al. (1997) included a power law condition (random and twin conditions equating to different and same noise, respectively) and a white noise condition (white and fixed-white conditions equating to different and same noise, respectively) and once again, the images were presented side-by-side. As for Ahumada & Beard (1997), Watson et al. (1997) found no benefit on observer performance for the similarity of the noise background with either power law noise or white noise backgrounds, again, suggesting that the observers were unable to benefit from the correlation present between the images in the twin and fixed white conditions.

Thus the limited psychophysical research using paired noise backgrounds, presented side-by-side, for signal detection tasks is inconclusive. Whilst, as Burgess & Colborne (1988) suggest, the ideal observer will subtract the two background images, the experimental evidence for this is mixed,

with Ahumada & Beard (1997) and Watson et al. (1997) finding no effect for same compared to different noise fields and Burgess & Colborne (1988) finding an improvement for same compared to different noise fields. The research available relating to signal detection in paired and correlated noise fields is limited and the papers cited here have different goals to this thesis, however, this is an area that this thesis will explore further.

One alternative research strand to the psychophysical approach discussed above is the work done investigating a comparative visual search strategy. The side-by-side presentation of images, with the goal of identifying a subtle abnormality present in one image but not in the other, could be likened to the "spot the difference" puzzles popular in magazines and newspapers and the challenge of these puzzles demonstrates the difficulty faced by the observer.

Atkins et al. (2006) identified the radiologist's task in viewing paired images as a comparative visual search task and used eye gaze tracking to investigate the interaction of radiologists' with paired artificial images. They found that the search pattern occurs in two phases, firstly searching one image for a target and then making multiple saccades between images to confirm their finding, with longer fixations for more complex patterns, concluding from this that the repeated saccades are necessary to prevent cognitive overload because of the limitations of visual working memory (Atkins et al., 2006). Irwin (1991) referred to this as transsaccadic memory, suggesting it was similar to visual working memory with a limited capacity, and showed how the accuracy of participants in a pattern discrimination task was impaired as the complexity of the images increased. With the mechanism of comparative visual search being underpinned by repeated saccades Atkins et al. (2006), the findings of Irwin (1991) cast doubt on the capacity of transsaccadic memory (or, indeed, working memory) to retain the level of information contained in a mammogram to enable comparison across images.

Although Atkins et al. (2006) used side by side images, they were different views rather than symmetrical pairs, similar to presenting the cranio-caudal view along with the medio-lateral oblique view (see chapter 2 for an explanation of mammogram views). Hence, whilst illustrating the cognitive limitations for the human visual system in conducting search tasks in paired images, they do not address or refute the notion that targets present in one image of a symmetric pair will "pop out" rather than necessitating cognitively demanding visual search. Pomplun (1998) conducted comparative visual search studies for his PhD thesis, and two of the experiments carried out compared the performance of observers detecting a target in symmetrical paired images that were either a translational repeat or mirror repeat of each other; the only difference between the paired images being the presence of the target in one image of the pair (experiments A and E, Pomplun (1998)). Pomplun (1998) found no benefit of mirror symmetric presentation over

translational repeat presentation for the detection of the target and, using the criteria of search speed, area covered in the search and saccade length as measures of efficiency, found both search speed and area covered to be significantly higher with translationally symmetric images than for mirror symmetric images. Saccade length was significantly longer for the mirror symmetric images than for translationally symmetric images and Pomplun (1998) pointed out that the variability, and potentially longer distances between corresponding points in the mirror symmetric images, compared to the fixed distance between corresponding points in translational repeat images could lead to inaccurate programming of saccades with their end points likely to be in the wrong position. Although the stimuli used were very simple in comparison to a mammogram, the results from experiments A and E of Pomplun (1998) do not lend support to the notion that either translational or mirror symmetric presentation enables targets to "pop out" and, also, suggest that translational repeat symmetry is likely to confer better target detection performance than mirror symmetry.

The evidence presented does not lend support to the notion that displaying image pairs in a mirror symmetric format will confer any advantage to the observer for the detection of a signal in one of the images, either through utilising the correlation present or by some pre-attentive "pop out" mechanism.

1.7.3 Paired noise fields presented sequentially as an animated display

One alternative to presenting images in a side-by-side format is to display the images in an animated sequence, in the same location, so that any differences between the images will appear to flash on and off (flicker) or appear to move. Whilst there is extensive research into the sensitivity of the human visual system to temporally modulated stimuli (see section 1.5.2 for a short review), a large section of this research investigates the detection of flicker as a property in its own right (where the observer is discriminating between a flickering and non-flickering stimulus) and there is considerably less research investigating the use of animated displays as an aid to the detection of signals within those displays. The intuitive notion behind animated displays is that this mode of presentation will enable the visual system to make use of the correlation between the images facilitating redundancy reduction and better signal detection performance.

The use of animation to assist in visual search and detection is not a new idea, however, and examples of its application can be found in a number of diverse applications. In astronomy the blink comparator (Mayer & Phillips, 1983) has been used since the early 1900s to study astronomical photographs and it operates by rapid alternation of the images, enabling the astronomer to detect small differences between the images. When comparing two images of the night sky, taken several days apart, moving objects such as planets and comets will be observed to move and flicker, whereas

stars will remain stationary and, in 1930, the planet Pluto was discovered by Clyde Tombaugh, using a blink comparator to study photographic plates of a region in the constellation Gemini (Fraknoi, 2009).

Security is another application, using animated displays of images of security seals, taken at the initial sealing and at a later date, typically following transport or storage. Rapidly alternating the two images will cause any differences to flash on and off, which may be evidence of tampering (Lazerson, 1984). Research has also been carried out comparing the effectiveness of displaying two architectural drawings sequentially in the same location against displaying them side by side for the detection of differences, and possible errors in reproduction, between the two drawings (Fleury & Jamet, 2014). Although not strictly an animated display, presenting the drawings sequentially in the same location does bear some resemblance to the animated displays used in this thesis and Fleury & Jamet (2014) demonstrated improved effectiveness for the detection of errors (seen as differences between the two drawings) using this strategy.

A common factor across the application of animated displays in astronomy, security and architecture is the close similarity of the images being compared, with the aim of finding minor differences. The images compared in medical imaging, and in particular in mammography, rarely have such high levels of similarity and this may be one reason why there appear to be very few examples of the use of animated displays in this field. Notwithstanding this potential limitation, some studies have been conducted, such as the use of a blink comparator approach by Carlbom (1994), as a technique to identify mismatch between corresponding views of nerve cells to assist in registration and the construction of 3D images of the cells.

Animated displays were compared against the more traditional side by side display for the detection of tumours in brain images by Erickson et al. (2011). Erickson et al. (2011) used baseline and follow up images from 66 cases with a known malignant brain tumor and asked participants to identify changes between the baseline and follow up images. Three processing methods were tested; normal, where no processing was carried out, image subtraction and change detection using a computer algorithm. Each of these methods could be displayed in a traditional side-by-side mode or a flicker mode, alternately displaying the images in the same location, thus giving a total of six presentation methods (Erickson et al., 2011). Whilst image subtraction and change detection were found to be significantly better than the normal images for detecting subtle changes, the option of using animated displays did not improve detection for any of the presentation methods. The study used experienced neuroradiologists, although only one had any experience of animated displays and this observer did show an improvement with flicker, though it did not reach significance. Erickson et al. (2011) concluded that the lack of a significant effect for flicker may have resulted from a lack

of experience and training with the technique and, despite their findings, still believed that flicker has the potential to improve detection performance.

In relation to mammography, van Engeland et al. (2003) conducted a study to investigate the use of optimised displays for the detection of temporal change between previous and current mammograms. The study used real mammogram pairs without abnormality and pasted in lesions extracted from abnormal mammograms. The extracted lesions were resized to give seven new lesion dimensions and one resized lesion was pasted into one mammogram of the pair whilst the original size lesion was pasted into the other image of the pair. Two display modes were used; the traditional side-by-side display and the optimised display which allowed the observers to toggle between the two images of the pair. Four observers, who were physicists working in the field of mammography, were asked to select the image containing the largest lesion and a psychometric function of the fraction correct against resize factor was plotted showing a significant increase in performance for three out of the four observers when using the optimised display of toggling between the images (van Engeland et al., 2003). Whilst the toggle rate was manually controlled and determined by the observers and a different task was used, the study by van Engeland et al. (2003), nevertheless, hints at the strong potential for the effectiveness of an animated display mode such as will be tested in this thesis.

Honda et al. (2014) also conducted research into the use of an image toggle tool for the comparison of digital mammograms. The study used the retrospective images from 12 patients with suspected breast cancer and found that when the images were well aligned the use of the image toggling tool enabled easy detection of differences between the image pair, suggesting that the tool may be an effective aid to the radiologist (Honda et al., 2014). A similar study was conducted by Hasegawa et al. (2008), who found that with accurate registration, toggling between images was an effective method for finding subtle masses, asymmetries and architectural distortions within the breast. Whilst the research supporting the use of sequentially displayed images by toggling may be limited, the technique does have its advocates with a number of texts recommending toggling images as an aid to the detection of abnormalities (Dogan, 2012, p. 26; Kopans, 2007, p. 367).

Whilst not using sequential displays of mammogram pairs, another study investigating the effectiveness of utilising motion and flicker as an aid to the detection of microcalcifications in real mammograms was undertaken by Plett et al. (2007). Plett et al. (2007) used two types of motion to create the dynamic cues; firstly, spatial motion, where each pixel oscillated sideways with the rate of oscillation determined by the pixel intensity such that lighter pixels oscillated faster than darker pixels. Secondly, they induced temporal motion by oscillating the intensity of each pixel with the rate of oscillation determined by the original intensity of the pixel and, again, lighter pixels

oscillated faster than darker pixels. Diagnostic performance for the detection of microcalcifications in real mammograms was measured for five radiologists and five non-radiologists and analysed using receiver operating characteristic curves (ROC). In comparison to displays without dynamic cues, the area under the ROC curve (AUC) was 20.8% and 8.4% greater for non-radiologists and radiologists, respectively (Plett et al., 2007). Although the study by Plett et al. (2007) measured performance for the detection of microcalcifications, rather than masses (as this thesis will focus on) and single mammograms, rather than prior and current mammogram pairs (again as focused on in this thesis), it does, once again, show that motion and flicker are powerful cues that merit further investigation.

The evidence presented in section 1.7.3 gives an indication of the strong potential for animated displays to improve observer performance in signal detection tasks, however, the research cited is highly applied and there appears to be a dearth of research with a visual psychophysical approach that would provide the theoretical basis for an animated presentation method. One of the aims of this thesis is to to go some way to addressing this void by providing a theoretical basis for signal detection in paired noise fields and investigating this in laboratory conditions using Gaussian white noise images before addressing the applied use of animated displays to create a flickered stimulus in the field of mammography.

1.8   Chapter Summary

The aim of this chapter was to introduce the historical and theoretical background that underpins the research conducted in this thesis. The detection of visual signals in noise backgrounds builds upon a rich history of research spanning a broad base. From the early research determining the absolute threshold for seeing, researchers have investigated the various sources of noise that cause the variation in observer responses over and above this absolute threshold, using signal detection theory to determine the optimum possible performance for a given task and providing a benchmark against which other observers can be compared. The chapter has shown how this ideal observer approach not only sets a baseline performance level for comparison against, but, more importantly, reveals the optimum strategy for executing a given task. It is not an unexpected outcome that the human visual system will often fail to execute this optimum strategy, but this knowledge alone points the researcher towards the nature of the interface required to enable it to do so.

The detection of a tumour in one image of a pair of mammogram images is a complex task that encompasses a wide range of the historical and theoretical research and this chapter reviews a broad swathe of this, including the effect of internal and external noise, the response of the human

visual system to spatially extended patterns and the effect of the different types of background in which these patterns exist. Despite the complexity of the images, the ideal observer approach, nevertheless, enables an optimum strategy to be established and, for the detection of a signal in paired noise fields, this has been shown to be achieved by the decorrelation of the paired images, to remove redundant information. This is then followed by the cross correlation of the resulting images with a template of the signal to be detected, selecting the image that returns the highest internal response.

Real mammograms are complex images that introduce a level of uncertainty into the testing of a concept such as a decorrelating and cross correlating strategy. If the theoretical basis for this strategy can be established using the simplified conditions of Gaussian blob signals located in known locations in Gaussian white noise backgrounds, the theory can then be applied to more complex scenarios, such as power law noise backgrounds and, ultimately, to real mammograms. Thus, the use of both Gaussian white noise and power law noise backgrounds, in addition to real mammogram backgrounds has been introduced in this chapter and, this knowledge will be applied to the research in this thesis.

This chapter has suggested that there is limited evidence that the human visual system can execute this ideal strategy when the images are presented in a side-by-side format, as is the predominant practice in mammography, but introduces an alternative method of displaying the two images of the pair alternately in the same location in a continuous movie sequence. This establishes the rationale for this thesis; can the human visual system utilise the optimum strategy of decorrelating and cross correlating with paired images to optimise its performance for the detection of a signal in one of the images of the pair when they are presented side-by-side or, alternatively, when they are presented in a continuous, alternating movie sequence.

This rationale underpins each of the experimental stages of this thesis, however, before presenting the experiments, Chapter 2 will introduce mammography as an applied clinical discipline. Mammography has own its distinct terminology and practices along with a substantial body of clinical research, a small portion of which will be reviewed and the aim of Chapter 2 will be to enable the reader to ground the theoretical aspects of this thesis in the clinical application of mammography. Having done this, the broad aims of this thesis and a summary of each experiment will be given.

Chapter 2

Introduction to Mammography

2.1 Introduction

This thesis is concerned with the visual detection of signals embedded in noise and the background literature was reviewed in Chapter 1. My aim is to apply our knowledge of such visual detection to the applied problem of detecting lesions (abnormalities such as tumours) in mammograms (x-ray images of breasts). In this chapter I will show how, in certain defined circumstances, the radiologist's task of detecting a tumour in a mammogram can be considered as a case of detecting a visual signal in a noise background. Mammography is a highly complex and extremely challenging discipline with its own language and terminology. Whilst this thesis will be removed from the clinical application, it is, nevertheless, important to establish a basic understanding of what mammography is including the visual factors limiting the detection of tumours in real mammograms and the approaches used by radiologists to assist them in this task. This understanding will enable a better understanding of how the experimental results may impact on the clinical application and will help to place the thesis in the wider picture of mammography research.

2.2 Breast Cancer

Breast cancer is the most common cancer in the UK with a lifetime risk of diagnosis for women of 1 in 8 and in 2012, 11,643 women in the UK died from breast cancer (Cancer Research UK, 2014a). For women in the UK in 2012, 15% of all deaths from cancer were from breast cancer and it was second behind lung cancer as the most common cause of death from cancer in women (Cancer Research UK, 2014b). The breast is made up of adipose (or fatty) tissue, fibrous connective tissue and glandular tissues which consist of the milk producing lobules and the ducts to deliver the milk

Lobes

Milk ducts

Fatty tissue
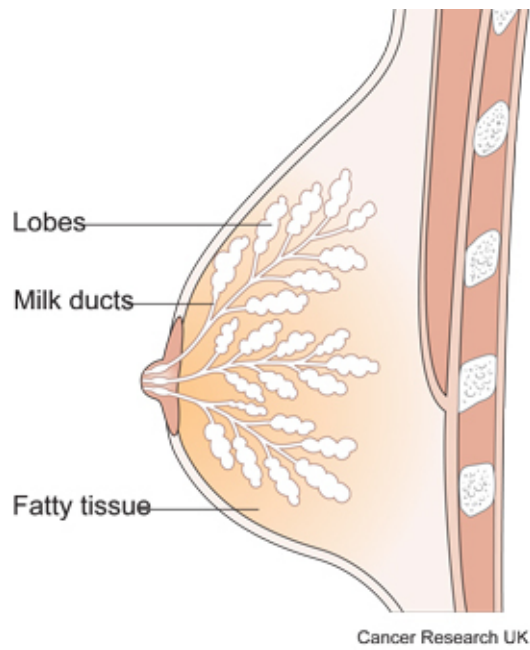
Cancer Research UK

Figure 2.1: Simplified breast image.
(Cancer Research UK / Wikimedia Commons, 2014)

to the nipple (figure 2.1). The lobule and duct are commonly referred to as the terminal ductal lobular unit (TDLU) and most tumours will originate here (Kopans, 2007, p. 30). The fibrous connective tissue and the glandular tissues together are often referred to as fibroglandular tissues and these are more dense and appear whiter on the mammogram than the fatty tissues (Bontrager & Lampignano, 2013, p. 751).

Breast cancer, like any other cancer, is an uncontrolled growth of cells leading to the formation of a tumour. The tumour may be benign or malignant. Benign tumours are slow growing, do not spread to other parts of the body and are not a threat to health. Malignant tumours can grow rapidly with the potential to spread to other parts of the body and become life threatening.

2.3   The Mammogram

A mammogram is an x-ray image of the breast and is carried out to detect and diagnose breast disease. It can be conducted as a diagnostic mammogram to investigate a specific problem or as part of a screening programme to give early indication of breast cancer. It can identify cancers that are too small for a woman or her doctor to see or feel. The mammogram is carried out by compressing the breast between an adjustable plate on top of the breast and a fixed x-ray plate underneath (figure 2.2). The compression of the breast functions to keep the breast still during the imaging process and enables a thinner layer of breast tissue to be x-rayed. This allows the

Figure 2.2: Woman undergoing a mammogram of the right breast (National Cancer Institute / Wikimedia Commons, 2006)

radiation dose to be minimised, reduces x-ray scatter and produces a clearer, sharper picture with a more uniform density of tissue across the image (Andolina & Lillé, 2010).

2.4 Breast Cancer Screening

The goal of any screening programme is the early identification of asymptomatic disease so that it can be treated and thus, reduce mortality from that disease. Mammograms enable the radiologist to detect tumours before any symptoms are present and so breast screening has the potential to reduce mortality from breast cancer. There is an extensive literature supporting the effectiveness of breast screening programmes and these are established in the vast majority of developed countries around the world (Youlden et al., 2012). They are regarded as an essential tool in reducing breast cancer mortality by both the European Union (Perry et al., 2006) and the World Health Organisation (World Health Organisation (WHO), 2002). Mammographic screening is the most widely adopted and only evidence based screening methodology for the early detection of breast cancer with numerous supporting studies showing mortality reductions in randomised controlled trials and service screening evaluations (see Hakama et al. (2008) for a summary). The National Health Service (NHS) England has operated a screening programme since 1988 (Advisory Committee on Breast Cancer Screening, 2006) and during 2012/13 1.97 million women aged 45 or over were screened, detecting 16,432 instances of cancer (Health and Social Care Information Centre, 2014).
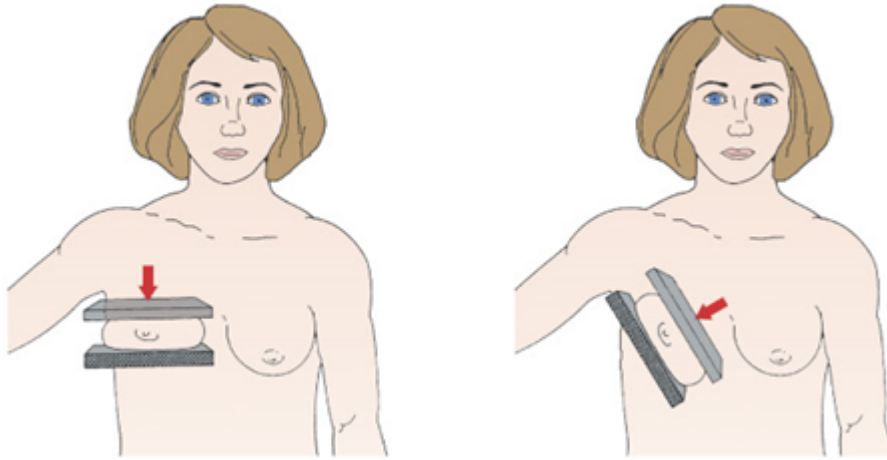
Figure 2.3: Cranio-caudal (CC) (left) and Medio-lateral Oblique (MLO) (right) positioning

In the UK, screening is carried out on women between the ages of 50 and 70 with three year intervals between each screening (Advisory Committee on Breast Cancer Screening, 2006).

A screening mammogram usually consists of four images, two from each breast with two views of each. The views used for routine screening mammograms are the cranio-caudal view and the medio-lateral-oblique view as shown in figure 2.3. The cranio-caudal (CC) view is a projection taken from above a horizontally-compressed breast and the medio-lateral-oblique (MLO) view is taken from the side and at an angle of a diagonally-compressed breast (Andolina & Lillé, 2010).

2.5   Symmetry and Asymmetry

Symmetry is an appealing quality and often associated with good health and attractiveness in humans (Livshits & Kobyliansky, 1991; Møller & Thornhill, 1998; Scheib et al., 1999; Tovée et al., 2000; Wade, 2010). Random deviation from perfect symmetry in bilaterally symmetric organisms, such as humans has been linked to qualities of health and sexual selection (Tomkins & Kotiaho, 2002) and asymmetry, therefore, represents an obvious choice as an indicator of disease or ill health. This is a concept that is widely supported in mammography with an extensive literature investigating asymmetry between breasts and the associated risk of developing breast cancer (Eltonsy et al., 2007; Scutt et al., 2006; Wang et al., 2011). However, whilst seemingly straightforward concepts, the terms symmetry and, in particular, asymmetry can have different meanings in mammography when compared to the mathematical definitions or even the layman's understanding of the terms. Even within mammography, asymmetry can be interpreted in several ways. It would, therefore, be prudent to clarify the various interpretations and how the terms will be used in this thesis.

2.5.1 Symmetry and asymmetry in visual science and mathematics

Wagemans (1995, p. 10) describes symmetry as "...self-similarity under a class of transformations...". This definition of symmetry includes transformations such as rotation or translation (Darvas, 2007) as shown in figure 2.4.

Figure 2.4: Types of symmetry - reflection (left) and translation (right)

Geometric symmetry can be mathematically defined with, for example, a function $f(x, y)$ having perfect mirror symmetry when viewed around a vertical axis y when:

$$f(x, y) = f(-x, y) \tag{2.1}$$

Where x and y are points, in each image, orthogonal and parallel to the axis of symmetry. There are a number of mathematical measures of the level of symmetry (or conversely, the level of asymmetry), such as the measurement of the difference between the two halves of an image, given by an equation of the form:

$$Level\ of\ asymmetry = \int_{x_0}^{n} \left[f(x, y) - f(2x_0 - x, y) \right]^2 \tag{2.2}$$

(Tyler, 2002)

Where:

$n = $ image size in pixels

$x_0 = $ position of the axis of symmetry

This thesis, however will use the Pearson correlation as a measure of the symmetry between the two halves of an image or of two paired images. The level of symmetry of functions $f(x_1, y_1)$ and $f(x_2, y_2)$ is given by:

$$\rho\left(\left(x_1, y_1\right), \left(-x_2, y_2\right)\right) = \frac{Covariance\left(\left(x_1, y_1\right), \left(-x_2, y_2\right)\right)}{\sqrt{Variance\left(x_1, y_1\right) Variance\left(-x_2, y_2\right)}} \tag{2.3}$$

When the Pearson correlation $\rho = 1$, the functions have perfect mirror symmetry and when $\rho = 0$, the functions are perfectly asymmetric. Values between $\rho = 1$ and $\rho = 0$,, indicate varying levels of asymmetry. This visual science or mathematical definition of symmetry will be referred to as "geometric symmetry".

2.5.2   Symmetry and asymmetry in mammography

As we will see in section 2.8, the breast radiologist often views mammograms in symmetric pairs and, by the scientific or mathematical approach described above, any differences would be classed as asymmetries. In some cases, radiologists do use a mathematical approach, however in other circumstances the use of the word "asymmetry" has a stricter definition relating to a category of abnormality (see section 2.6.2). These two, contrasting, approaches to the seemingly simple concept of asymmetry will now be discussed.

2.5.2.1   The use of asymmetry, as defined mathematically, in mammography

Most textbooks recommend that radiologists view mammograms in a mirror symmetric display to take advantage of the natural symmetry of the breasts, using deviations from symmetry, or asymmetry, as possible indications of disease (Andolina & Lillé, 2010; Bun, 2002; Harvey & March, 2013; Kopans, 2007; Sickles, 2007). The level of geometric asymmetry observed by the radiologist is subjective and the unquantifiable nature of this decision means that replicability can be difficult to achieve. Nonetheless, overall geometric asymmetry of the breasts, particularly in relation to the size of each breast, remains an important indicator for the breast radiologist.

A more objective approach, that does make more quantifiable estimates of the level of asymmetry is that of fluctuating asymmetry. Fluctuating asymmetry is a measure of biological asymmetry and refers to small random deviations from perfect symmetry in bilaterally paired structures such as may be present between the left and right sides of the human body (Tomkins & Kotiaho, 2002). In the context of mammography fluctuating asymmetry would represent differences between the left and right breasts and it can be measured in several ways. Eltonsy et al. (2007) and Scutt et al. (2006) measured the difference in the volume of each breast to estimate fluctuating asymmetry. Another measure uses the fluctuation of image pixel intensity between each breast to calculate density differences (Zheng et al., 2012, 2014). Feature based methods are also commonly used, particularly with computer aided detection (CAD) systems, whereby suspicious features are iden-

tified through image filtering techniques based upon image statistics and image qualities, such as texture and brightness and a comparison is then made to the opposite breast to identify asymmetric features (Lau & Bischof, 1990; Tahmoush & Samet, 2006; Wang et al., 2011).

Fluctuating asymmetry, therefore, represents a measure of the amount of deviation from perfect symmetry between the left and right breast as a whole, however, this is predominantly a research tool and is rarely, if ever, used by the practicing clinical radiologist. The clinical radiologist will use the term asymmetry in a more precise manner to refer to a specific class of abnormality as described below and in section 2.6.2

2.5.2.2 The use of asymmetry, as defined as a distinct class of abnormality, in mammography

When a clinical radiologist refers to the term "asymmetry", they are usually referring to a specific class of abnormality. While normal and other abnormal variations between breasts may contribute to the mathematical definition of asymmetry between two mammograms, the clinical radiologist will not usually call these features asymmetries unless they meet the strict definition of the class of abnormality known as asymmetry (Zonderland & Smithuis, 2013). Thus normal variation will be referred to as such and other abnormalities, such as a mass or an architectural distortion, will be identified by the class of abnormality that they belong to (see section 2.6 for a discussion of the types of abnormality encountered by the radiologist). A fuller discussion of asymmetric findings will be given in section 2.6).

2.6 Abnormalities – What is the Radiologist Looking For?

Breast cancer is an uncontrolled growth of cells leading to the formation of a mass of abnormal tissue known as a tumour. The tumour may be benign or malignant, however, before its malignancy or otherwise can be determined, it has to be found. This is the challenge for the radiologist. Whilst the tumour may not always be visible on the mammographic views taken, other signs such as asymmetry, architectural distortion or calcifications may be indicative of an underlying tumour and a brief description of each will be given to understand their relevance in the context of this thesis.

2.6.1 Mass

A mass is a three-dimensional feature with convex outward borders that is usually visible on two different views (Sickles et al., 2013). Masses can be round, oval or irregular in shape and can have

well-defined or indistinct edges that are microlobulated (small undulating circles along the edge of the mass), indistinct or spiculated (thin lines radiating from the mass) (Sickles et al., 2013; Zonderland & Smithuis, 2013).

2.6.2 Asymmetry

Zonderland & Smithuis (2013) define mammographic asymmetry as "Findings that represent unilateral deposits of fibroglandular tissue not conforming to the definition of a mass." Fibroglandular tissue refers to both the fibrous connective tissues and the glandular tissues of the breast and these will appear as the whiter areas on the mammogram (Bontrager & Lampignano, 2013, p. 751). The fibroglandular tissue will normally be bilaterally similar, however, when differences in the pattern of the fibroglandular tissue occur between the left and right breasts this will be reported as an asymmetry. Although a mass or architectural distortion can represent a difference in the parenchymal pattern from left to right breast they are not referred to as an asymmetry. A mass can be discriminated from an asymmetry by its spherical three dimensional shape, appearing denser at the center than at the periphery and it will be visible on two different projections whereas an asymmetry may only be visible on one projection, will have a more evenly dense appearance and will lack the more distinct convex borders of a mass (Sickles, 2007; Zonderland & Smithuis, 2013). An architectural distortion will be differentiated from an asymmetry by the pinched or indented pattern of the tissue (Shaw De Paredes, 2007, p. 370). The Breast Image Reporting and Data Systems (BIRADS) defines four sub-categories of asymmetry – asymmetry, global asymmetry, focal asymmetry and developing asymmetry (Zonderland & Smithuis, 2013), as summarised below (Sickles, 2007).

Asymmetry   represents an area of fibroglandular tissue that is only visible on one of the two standard views (either MLO or CC, but not both).

Global asymmetry   represents a greater volume of fibroglandular tissue, with no apparent mass, architectural distortion or calcification present, in one breast than in the opposite breast and corresponding to at least one quadrant of the breast.

Focal asymmetry   represents an asymmetry of fibroglandular tissue corresponding to less than one quadrant of the breast, visible on both standard views (MLO and CC) but lacking the distinct convex contours obvious characteristics of a mass.

Developing asymmetry represents a focal asymmetry that has grown larger or denser since the previous screening.

2.6.3 Architectural distortion

An architectural distortion is a localised disruption of the normal shape of the breast or of the internal pattern of the breast which is seen as an indentation or tightening of the features Shaw De Paredes (2007, p. 370) and it is this indentation or tightening that differentiates it from an asymmetry. For an abnormality to be classified as architectural distortion there should be no definite mass visible (Zonderland & Smithuis, 2013). Architectural distortion has a number of underlying causes, most of which are benign, such as post-surgical scarring, sclerosing adenosis (nodules of fibrous tissue or cysts), trabecular thickening (a thickening of the connective and supportive tissues) or fat necrosis (concentration and inflammation of fatty tissue usually resulting from surgery or trauma) but the distortion may also be indicative of a malignant breast lesion (Shaw De Paredes, 2007, pp. 363-444).

2.6.4 Calcifications

Calcifications are calcium deposits within breast tissue appearing as very small white spots or flecks on a mammogram, usually less than 1mm in size (Zonderland & Smithuis, 2013).

2.7 Use of symmetry and asymmetry in this thesis

In an effort to achieve consistency and clarity throughout this thesis the following definitions relating to symmetry and asymmetry will apply to this thesis:

Symmetry and symmetric will be used to refer to the type of presentation of the image pairs when they are presented side by side. Images that are reflected about a vertical axis will be referred to as having mirror symmetry or as being mirror symmetric. Images that are a same orientation translation of each other will be referred to as having repeat symmetry or as being repeat symmetric.

Fluctuating asymmetry refers to the differences between the left and right breast resulting from normal and abnormal variations (see sections 2.6 and 2.10).

Asymmetry will be used to refer to the specific class of abnormality as described in section 2.6.2.

2.8   Reading the Mammogram

Radiologists typically conduct a systematic search of each mammogram image for features of concern and make comparisons with the corresponding regions of any available images of the same breast from prior screenings and to the opposite breast from the same screening to identify differences that may indicate breast cancer (Harvey et al., 2008). To facilitate the latter, it is common practice for radiologists to view the mammograms in pairs – either bilateral pairs that consist of the same view of the left and right breast from the same screening or temporal pairs that consist of the same view of the same breast, taken at two different screenings (Kopans, 2007, p. 367).

2.8.1   Bilateral pairs

Bilateral pairs (sometimes referred to as contralateral pairs) consist of the same view (either CC or MLO) of the left and right breast from the same screening. As shown in figure 2.5, bilateral pairs are normally viewed in a mirror symmetric format. This protocol is recommended by most texts with the rationale that the radiologist can take advantage of the natural symmetry of the breasts (Andolina & Lillé, 2010; Bun, 2002; Harvey & March, 2013; Kopans, 2007; Sickles, 2007). Bilateral asymmetry is an important indicator for the radiologist as a predictor for breast cancer and this will be discussed later in this chapter.

2.8.2   Temporal pairs

Temporal pairs consist of the same view (either CC or MLO) of the same breast, taken at two different screenings. Temporal pairs aim to take advantage of the stasis of the breast over time and during the examination the radiologist will look for differences that indicate changes from the previous image. A feature that may, in itself, cause concern to the radiologist, could be discounted as a risk if it appears in, and is unchanged from, the previous mammogram. The comparison of temporal pairs has strong evidential support for reducing false positive decisions (Roelofs et al., 2007; Sumkin et al., 2003; Thurfjell et al., 2000), identifying developing asymmetry (Sickles, 2007), reducing recall rates (Callaway et al., 1997) and identifying subtle changes, such as the appearance or change of a small mass, that may be indicative of cancer (Frankel et al., 1995; White et al., 1994).
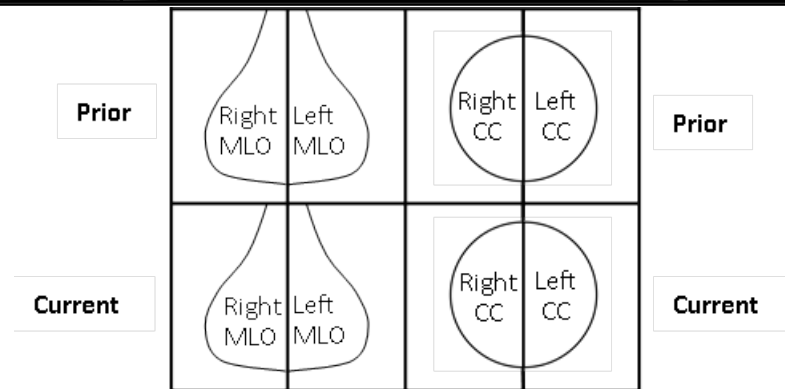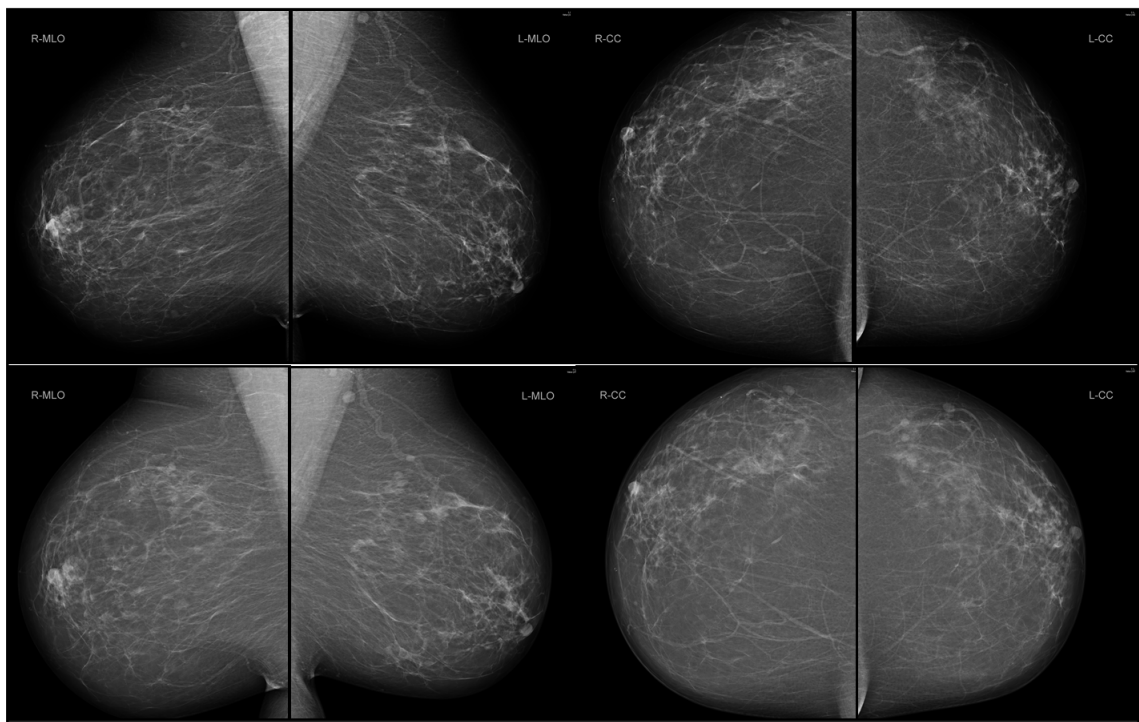
Figure 2.5: Standard mammogram opening hanging protocol with key

2.8.3   Mammogram workstation hanging protocols

As an aid in identifying abnormalities from the paired images, radiologists often use a mirror symmetric hanging protocol (Haygood & Dogan, 2013; Kopans, 2007, p. 367). Whilst, in the UK NHS, hanging protocols are not mandated, this is the usually recommended protocol such that bilateral pairs are displayed in a mirror symmetric format with the same four views from the prior screening situated above, as shown figure 2.5. The temporal pairs are, therefore, usually viewed as translational or repeat symmetric images of each other. Whilst it is technically possible for most mammography workstations to display the temporal image pairs in a mirror symmetric format, this is not usually done.

2.9   The Importance of Detecting Masses

Research by Venkatesan et al. (2009) suggests that a mass is the most likely indicator of malignant (or invasive) breast cancer. Venkatesan et al. (2009) found from their sample of 1287 instances of invasive breast cancer, that 68% were identified with a mass, compared to 5% identified with asymmetry, 6% identified with architectural distortion and 21% identified with calcifications. Other studies have also shown similarly high positive predictive value of both masses and calcifications; Gajdos et al. (2002) suggest that 95% of masses were caused by invasive cancers and McKenna (1994) concluded that masses, in particular when spiculated in shape and with suspicious calcifications were indicative of a high probability of breast cancer, but asymmetry was associated with a much lower probability.

This indicates the importance of identifying masses and this is further emphasised by the relationship between tumour size and mortality with the chances of survival being inversely correlated with the size of the tumour and ten year survival probability improving from 60% for a 2-5 cm tumour to 95% if the tumour is detected before it exceeds 1 cm (Tabár et al., 1999).

This is not to underestimate the importance of other indicators of breast cancer and geometric asymmetry, in particular, is recognised as an important sign for the radiologist, with a great deal of evidence linking asymmetry to the risk of breast cancer (Eltonsy et al., 2007; Scutt et al., 2006; Wang et al., 2011; Zheng et al., 2012, 2014). Indeed, symmetry or, asymmetry, is often cited as the radiologists' most potent weapon in the search for abnormalities (Andolina & Lillé, 2010) and, as previously stated, current practice recommends viewing mammograms back to back in a mirror image display to assist the radiologist in identifying deviations from symmetry (Harvey & March, 2013; Bun, 2002; Andolina & Lillé, 2010; Kopans, 2007). Thus, there is no question that asymmetry is a powerful tool and, in some cases, may be the only indication of breast cancer available to the radiologist (Sickles, 2007). Nevertheless, it should be clarified that the aim of this thesis is not to measure an observer's ability to identify whether two images are symmetric or not, rather the focus will be on the detection of discrete masses (simulated by a Gaussian blob) in paired images that contain a level of symmetry ranging between none and perfect symmetry. There is a large body of literature on the capability of the human visual system to detect symmetry and deviations from symmetry (Baylis & Driver, 1994; Julesz, 1971; Koning & Wagemans, 2009; Treder, 2010; Treisman & Patterson, 1984; Wagemans, 1995), however, the medical and psychophysical literature has much less to say on the effectiveness of symmetric displays in making a discrete mass more salient and testing this will be one of the goals of this thesis.

2.10   Normal Differences Between Paired Mammograms

Although, taking the subjective approach of the radiologist, the breasts are fairly symmetric structures, differences between breasts will always be evident on the mammograms. For bilateral pairs, this may result from natural differences between each breast and for temporal pairs the breast can change over time as a result of ageing, lifestyle changes or hormonal factors (Heine & Malhotra, 2002). For both bilateral and temporal pairs, differences can be introduced by process induced differences resulting primarily from the variation of breast positioning and compression within the scanner. The presence of differences between the images, whether they are a bilateral or temporal pair, is, therefore, normal, to be expected and is not usually indicative of a problem.

2.11   The Detection of Masses as a Psychophysical Task

Whilst, the detection of asymmetry and architectural distortion is clearly important and, in a clinical context, will clearly influence the radiologist's decision making strategy, from a psychophysical viewpoint, they could be considered to be part of the background noise limiting the observer's ability to detect a tumour. Thus, in the context of this thesis, where we are interested in an observer's ability to detect discrete masses in paired images, presented in a symmetrical, or alternative, protocol, asymmetries, architectural distortions and calcifications are not features to be detected, but part of the background noise limiting detection of the mass. Normal differences between the two images, whether they are bilateral or temporal pairs, will also contribute to this noise. Figure 2.6 illustrates this concept of a tumour representing a psychophysical signal and the breast tissue representing the background noise and figure 2.7 shows how this may look in a typical psychophysical experimental set up using real mammogram, Gaussian white noise and power law noise backgrounds.

When considering the psychophysical task of detecting a discrete mass from a paired noise background, the combination of differences between the images of the pair, resulting from abnormal and normal variations, creates a level of difference between the two images that can be measured by the Pearson correlation between them. The greater the difference is, the lower will be the level of correlation and vice versa. Hence, phrased in psychophysical terminology, when viewing paired images, the radiologist's task is, typically, to identify (recognise) a signal in paired noise backgrounds that have a level of correlation between them, where the signal is unknown in location, shape and size, etc.

This task description forms the basis for the experiments in this thesis, with initial experiments using a signal of known size, shape and location before progressing to signals of known size and
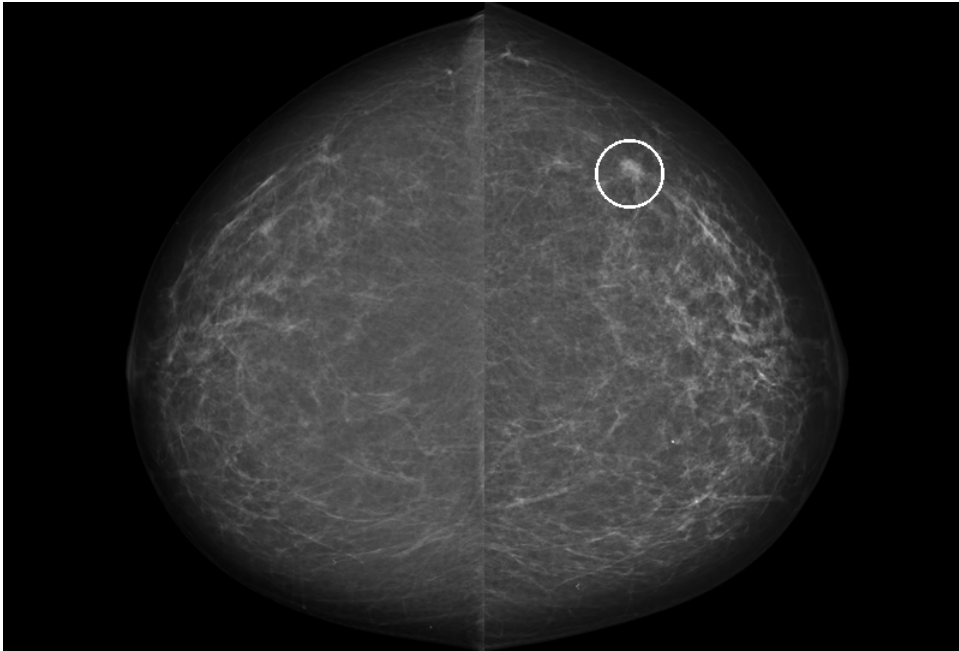
Figure 2.6: A CC mammogram pair, displayed in a mirror symmetric format, illustrating the concept of a signal (the tumour (circled)) in noise (the breast tissue - consisting of fat, fibroglandular tissue and blood vessels)

shape, but unknown location. For the Gaussian white noise and synthetic images used in this thesis, the level of difference between the two background images will be varied by adjusting the level of correlation between the two images. When these images are viewed side by side, varying the correlation varies the level of symmetry between the two images. This enables us to determine whether symmetric backgrounds, be they mirror or repeat symmetry, assist the radiologist in the detection of a signal embedded in one of the images.

2.12 Viewing Modalities

As already stated, current practice recommends that radiologists view image pairs in either a mirror symmetric format or a repeat symmetric format. This thesis will investigate the effectiveness of symmetry as an aid to the radiologist in detecting a tumour present in one image of the pair. An animated presentation method will also be introduced as an alternative viewing modality. To achieve this, two images of the pair will be displayed sequentially in the same location such that any differences between the two images will appear to flash on and off. The human visual system is highly sensitive to motion and flicker (Adelson & Bergen, 1986; Watson, 1986), with most cortical cells responding better to movement than to stasis (Hubel, 1995, pp. 78-79) and so an animated presentation method should be able to take advantage of this. This thesis will investigate the animated presentation method to determine its effectiveness in detecting a signal present in one
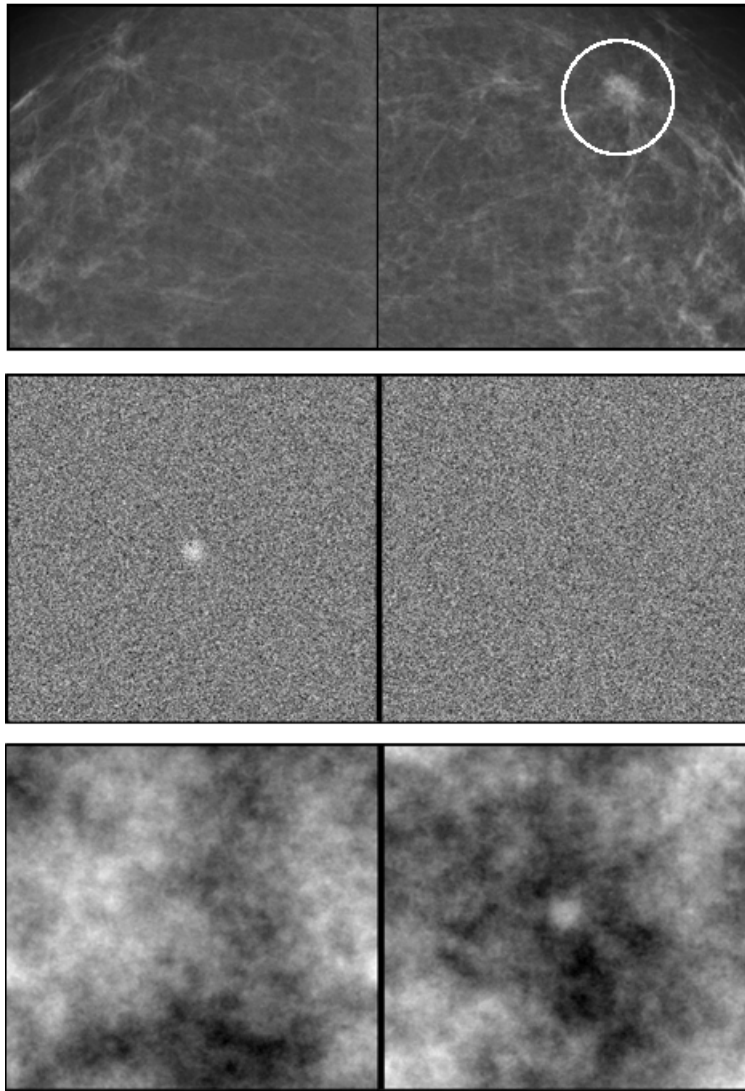
Figure 2.7: Examples of typical psychophysical type displays showing how a mammogram image pair with a tumour in one image (top) can be likened to a signal in a Gaussian white noise background (middle) or a power law noise background (bottom).

image of the pair and compare its performance against that of symmetric presentation.

2.13   Image Registration

Image registration is a technique that can be used to reduce the level of misalignment that can occur between the two images of a pair. Image registration can be defined as a process of transforming an image so that points in that image are brought into alignment with the corresponding points in a paired image (Wyawahare et al., 2009). Various techniques have been utilised to achieve registration algorithms. However, they basically fall into feature based or intensity based methods, or a hybrid of the two. Feature based methods utilise anatomical features such as the skin boundary,

the nipple, vasculature, the pectoral muscle (in MLO images) or fibro-glandular features. Intensity based methods use pixel intensity (Guo et al., 2006). The quality of registration is a measure of how closely aligned the two images in a pair are and a number of measures are available such as the sum of the squared differences, the normalised correlation coefficient and mutual information (Bozek et al., 2011a).

The effectiveness of registration for mammogram images, however, depends upon the initial similarity of the two images to be registered. It can be very effective for images that are similar but misaligned, such as in cases where the imaging process captures the same features of the breast but the breast has been shifted in a horizontal plane within the imaging scanner. Here, the two images can be brought into alignment by registration and high levels of correlation between the two images can be achieved. Registration is less successful where different features are captured as a result of changes in the breast or because of a different positioning of the breast in the scanner. For example, one image scan may capture the nipple whereas the second scan may compress the breast in a different vertical plane and not capture the nipple. In this example, where a feature (the nipple) is apparent in one image but does not appear in the second image, registration of that feature cannot be achieved. Notwithstanding this, even with an incongruity of features, registration will attempt to bring the breast boundary and any common features into alignment and, whilst the two images can never be identical in these cases, registration can increase the correspondence between them.

2.14 The Correlation Coefficient as a Measure of Image Matching

The similarity of the features present in the two images of a pair and the alignment of the corresponding features are both important attributes in the measurement of how closely two images are matched. How closely matched the two images are will determine the level of symmetry when the image pair are displayed in a symmetric format and the level of differences between the two images when they are displayed in a sequentially presented format. The level of symmetry and the level of difference between two images are the same measure - both reflecting the variable of how well matched the two images of a pair are and this variable is readily measured by the correlation coefficient (Pearson correlation) between the two images of the pair. Thus, the level of symmetry (when the images are presented side-by-side) and the level of difference (when the images are presented sequentially) reflect the same underlying variable and will both be measured by the correlation coefficient.

2.15   The Choice of a Target Signal

Breast cancer is a progressive disease which means that, unchecked, a malignant tumour will get bigger and will spread to other parts of the body. This progression can be halted by detecting the tumour early and the probability of a positive outcome is increased by early detection when the tumour is small (Michaelson et al., 2002; Tabár et al., 1999). It is important, therefore, to be able to detect small low contrast masses in one image of a bilateral or temporal pair. The appearance of masses can vary widely; round, irregular, distinct or indistinct, spiculated or smooth are just some of the descriptions attributed to a mass. For our aim of comparing viewing modalities in a laboratory setting, signal realism is a secondary consideration and a single, easily controlled signal is the primary goal. A small Gaussian blob was, therefore, chosen as the target signal giving good experimental control and an acceptable level of realism as discussed in section 1.6.3.

2.16   Chapter Summary

The aim of this chapter was to introduce the background information that establishes the context and relevance of this research in the clinical scenario. The chapter has also shown how specific aspects of the radiologists' task can be isolated and tested in a laboratory setting using a psychophysical approach. The psychophysical experiment has been identified as the detection of a signal in correlated paired noise backgrounds. Relating the experimental terminology to the clinical environment; the signal represents a tumour or mass that the radiologist aims to detect. The paired noise backgrounds are the mammogram pairs (bilateral or temporal) that the radiologist is viewing where the noise equates to the pixel values of the breast tissue. The correlation between the images of the pair will depend upon the differences between the two images and this will be the result of abnormal and normal variations. To isolate and test the specific task of detecting a small mass, in the experimental setting, the observer is only trying to detect the small mass and not the abnormal variations; in the context of this thesis, abnormal variation only represents a contribution to the correlation level between the images of a pair.

Chapter 3

The Aims of this Thesis

Chapter 1 introduced the broad range of literature that underpins the research conducted in this thesis and argued that the optimum outcome for the detection of a signal in paired noise fields can be achieved by an observer using a decorrelating and cross correlating strategy; this observer is known as the ideal observer. This strategy required that the paired images be decorrelated to remove redundant information before cross-correlating the resulting images with a template of the signal to be detected and selecting the image that returns the highest internal response as the image containing the signal. Chapter 1 argues that there is limited evidence that the human visual system can execute this decorrelating and cross correlating strategy when the images are presented in a side-by-side format but suggests that an alternative method of displaying the two images of the pair alternately in the same location in a continuous movie sequence may enable a human observer to operate in the same way as the ideal observer.

Whilst investigation of whether either mirror symmetric displays or animated movie displays can enable the human observer to operate in a manner akin to the ideal observer is a valid theoretical topic in its own right, this thesis also has a clinical rationale in the applied discipline of mammography. Chapter 2, therefore, introduces the terminology, practices and some of the relevant clinical research in the field of mammography to enable the reader to ground the theoretical aspects of this thesis in the clinical application of mammography. Chapter 2 establishes how the clinical task of the breast radiologist can be identified as the psychophysical task of detecting a signal in correlated paired noise backgrounds where the signal represents a tumour or mass that the radiologist aims to detect and the paired noise backgrounds represent the mammogram pairs that the radiologist is viewing.

This thesis will take a psychophysical approach to investigating the detection of a small mass in paired mammogram backgrounds, establishing the theoretical basis for the ideal observer strategy

in controlled conditions, before advancing closer to the clinical scenario. The aims of this thesis, therefore, are:

1. To investigate the effectiveness of symmetric displays, as used by breast radiologists, as an aid for the detection of a Gaussian blob in paired noise fields.

2. To investigate the effectiveness of animated displays as an aid for the detection of a Gaussian blob in paired noise fields.

3. To compare the effectiveness of animated and symmetric displays as an aid for the detection of a Gaussian blob in real mammogram pairs.

3.1 Summary of Experiments to be Conducted

3.1.1 Experiment 1: Weak Use of Symmetry in the Detection of Simulated Tumours in Paired Synthetic Mammograms.

The goal of Experiment 1 will be to address the first aim of this thesis: to investigate the effectiveness of symmetric displays, as used by breast radiologists, as an aid for the detection of a Gaussian blob in paired noise fields. The correlation between the paired images will be used as a measure of the symmetry between them and this will be varied between 0 and 1 to enable the effect of symmetry to be quantified. The experiment will first be conducted using Gaussian white noise fields to examine the theoretical basis for the use of symmetry, before introducing a power law noise, which better reflects the application of mammography.

3.1.2 Experiment 2: A Comparison of Side-by-Side Versus Animated Presentation of Images.

Experiment 2 will address the second aim of this thesis: to investigate the effectiveness of animated displays as an aid for the detection of a Gaussian blob in paired noise fields and compare against the effectiveness of symmetric displays. Again, the correlation between the image pair will be varied between 0 and 1 to investigate how varying the similarity of the images affects the observer's ability to detect the Gaussian blob signal. As in Experiment 1, the experiment will first be conducted using Gaussian white noise fields to examine the theoretical basis for the use of animation, before introducing a power law noise, which better reflects the application of mammography.

3.1.3 Experiment 3: Does the Rate of Alternation Affect the Observer's Contrast Threshold for the Detection of a Signal in Paired Noise Backgrounds?

The rate at which the displays are alternated will be investigated in Experiment 3. The aim of this experiment will be to determine what the optimal rate of alternation is for the task of detecting a Gaussian blob in paired noise fields and this will be investigated in both Gaussian white noise fields and power law noise fields.

3.1.4 Experiment 4: Comparison of Animated Presentation against Traditional Mirror Symmetric Presentation for the Detection of a Synthetic Tumour in Real Mammogram Backgrounds.

Experiment 4 will introduce real mammogram pairs with a synthetic Gaussian blob superimposed on one image. The aim of this experiment will be to investigate and compare the effectiveness of mirror symmetric and animated display types for the task of detecting a Gaussian blob in real mammogram image pairs.

3.1.5 Experiment 5: An Investigation into the use of Animated Presentation for the Detection of a Synthetic Tumour in Real Mammograms and Power Law Noise Backgrounds.

Experiment 5 will compare the effectiveness of mirror symmetric and animated display types for the task of detecting a Gaussian blob in paired real mammogram image sections and paired power law noise image pair sections. The experiment will use tightly controlled experimental conditions to ensure that the real mammogram image sections and power law noise image sections are matched on all specifications such that the only difference is the noise type. This will, thus, enable an unmitigated comparison of observer performance with each display type and each noise type.

Chapter 4

Experiment 1: Weak Use of Symmetry in the Detection of Simulated Tumours in Paired Synthetic Mammograms

4.1 Introduction

The introduction to mammography in Chapter 2 took one aspect of the breast radiologist's task and redefined it in psychophysical terms. Thus, the task of detecting a tumour or mass from the background of paired mammograms displayed in a mirror symmetric format is presented as the detection of a signal in correlated noise backgrounds displayed in a symmetric format. In the clinical environment, the correlation between the paired mammograms will vary dependent upon the level of normal and abnormal differences present, and the practice of viewing mammograms back to back in a mirror image display means that, as the level of correlation varies, the level of symmetry presented by the paired images will also vary.

Mirror symmetric presentation is widely recommended as an aid to making abnormalities more salient (Andolina & Lillé, 2010; Bun, 2002; Harvey & March, 2013; Kopans, 2007; Sickles, 2007), and there is an abundance of evidence to support this notion, showing that the human visual system is highly attuned to detect the occurrence of (and by inference, the violation of) visual symmetry (Baylis & Driver, 1994; Julesz, 1971; Koning & Wagemans, 2009; Treder, 2010; Treisman

& Patterson, 1984; Wagemans, 1995).

The effectiveness of mirror symmetric presentation for the detection of large violations, as would be caused by large abnormalities, is not disputed, however, there appears to be no empirical evidence to support the notion that mirror symmetric presentation assists the observer in the detection of a small localised violation, such as may be caused by a small mass or tumour. Indeed, research suggests that mirror symmetric presentation may have limitations for the detection of small localised violations in certain circumstances which may have implications for its use in mammography. First, there is evidence that violations of mirror symmetry may be less easily detected when the violation is remote from the midline of the symmetric display (Barlow & Reeves, 1979; Bruce & Morgan, 1975; Jenkins, 1982). This confers a substantial shortcoming for the use of symmetry in mammography as a localised mass may be present anywhere in the breast, not just close to the midline. Second, previous studies on symmetry detection providing evidence for the effectiveness of mirror symmetric displays as a tool for violation detection have used patterns that are relatively simple (e.g. Baylis & Driver, 2001), with a minimal number of features (e.g. Wenderoth, 1996) or a limited number of violations of the symmetric pattern (e.g. Locher & Wagemans, 1993). A typical pair of mammogram images, however, is not made up of such simple patterns, nor is it likely to be completely symmetric, with normal and abnormal variations in breast tissue potentially reducing the effectiveness of mirror symmetry as an aid to the detection of a localised mass when presented in the traditional side-by-side format.

While not agreeing on the underlying mechanism, several studies have shown the adverse effect of increasingly complex images on symmetry detection, providing explanations of increased levels of information within the images (Tapiovaara, 1990), the increasing density of that information (Rainville & Kingdom, 2002), or the increasing spatial frequency of the patterns within the images (Dakin & Herbert, 1998). In relation to the complexity of paired images, Huang & Pashler (2002), suggested that symmetry detection operates using coarse "binary maps" that filter individual features of an image, such as shape, size or colour, that are checked for symmetry (or for violations of symmetry). Huang & Pashler (2002) measured observers' response times to detect symmetric patterns and for all of the features presented found that response times increased as the complexity of the image increased concluding that symmetry detection is spatially inaccurate. When applied to a mammogram image pair, these findings suggest that symmetry detection would be a very coarse process and would be likely to miss minor violations of the symmetric pattern such as may be caused by a small tumour. This clearly has implications for the effectiveness of symmetry as an aid to the mammographer. Finally, although the occurrence of symmetry in nature is common, it is rarely perfect (Va'rkonyi & Domokos, 2006). This is demonstrated in examples commonly

thought of as symmetric but which rarely are, such as the human face (Lu, 1965) and the snowflake which, as Libbrecht (2006, p. 48) noted, "The vast majority show imperfect symmetry, if they show much symmetry at all". The implication here is that the human visual system may have optimally evolved to detect imperfect, natural symmetry but may be insensitive to minor violations of symmetry. This observation was supported by the Tjan & Liu (2005) study which found, using random dot patterns, that the visual system was disproportionately less well attuned to smaller rather than to larger departures from symmetry. The evidence from the study of Tjan & Liu (2005) that that small violations of symmetry are poorly perceived, gives us cause to doubt the effectiveness of such displays for the detection of small localised masses.

The preceding evidence suggests that the value of using symmetric displays to aid in the detection of small masses is, questionable and, whilst this evidence has focused on mirror symmetry, it should be noted that both mirror and repeat symmetric displays are used by the breast radiologist. Thus, the aim of this experiment was to test both mirror and repeat symmetric displays, however, before testing in a clinical scenario, this experiment will establish the theoretical basis by testing the effect of symmetry in a laboratory simulation. Thus, in this experiment we used synthetic images and synthetic tumours and presented two side-by-side noise images to simulate the conventional display of two mammograms side-by-side. Observers decided which image contained the synthetic tumour signal. Varying the level of correlation between the background images varies the level of symmetry presented by the images. By varying this during a signal detection task, the experiment is able to determine whether improvements in symmetry lead to improvements in observer performance. The question of whether symmetry helps in the detection of a signal can be analysed theoretically using an ideal observer approach (see Theory section of Experiment 1a on page 125). When attempting to detect a signal known exactly embedded in one of a pair of correlated noise patches, an ideal observer will decorrelate the two patches and cross-correlate a template of the signal with the decorrelated stimuli presented (<signal+noise> or <noise>) (Kay, 1998, p. 106). If the cross-correlation exceeds a criterion level, the observer says "signal present" and "signal absent" otherwise. Decorrelation effectively removes any correlated noise and with a correlation of 1, all the noise will be removed, leaving only the signal (Kay, 1998, p. 111). Thus, an observer who can decorrelate the noise patches will have much better performance than that of an observer who cannot perform such decorrelation. It is interesting to note that the correlation remains the same whether the image pair is presented in a mirror symmetric format or a repeat symmetric format and the ideal observer, therefore, performs the same for both types of symmetry.

Whether the human observer can perform decorrelation and use the symmetry of the background is an important question. The ideal observer provides the optimal benchmark against

which other observers can be compared and, even if humans perform poorly in relation to the ideal observer, the comparison provides an insight into the limitations of the human visual system and the requirements of the type of display that would enable the human observer to perform the task optimally. Real observer performance was assessed by measuring the contrast threshold for detecting the signal as a function of the degree of symmetry in the image pairs. The degree of symmetry was manipulated by varying the correlation between the two noise images. The correlation is defined as:

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{4.1}$$

where σ_{xy} is the covariance between pixel intensities in image x and image y, σ_x is the standard deviation of image x, and σ_y is the standard deviation of image y. If the correlation was zero, the noise was completely unrelated and the two images were completely asymmetric. If the correlation was 1.0, the two images had identical noise (though reflected about the vertical axis in the mirror condition) and the image pair had perfect symmetry. Intermediate levels of correlation produced pairs with partial symmetry. If symmetry helps the observer to detect the signal, as suggested by the performance of the ideal observer, the contrast threshold should decline as the correlation increases. In addition, if mirror symmetry as is commonly used in the clinic is helpful for detecting tumours and small masses, we expect that performance will be better for image pairs that have mirror symmetry rather than simply translational symmetry as presented in the repeat condition.

Experiment 1a used a Gaussian white noise background to enable a simplified examination of the theoretical basis for the ideal observer. Experiment 1b used a noise background with a $1/f^3$ power spectrum chosen for the similarity of its statistical properties with those of real mammogram backgrounds (Burgess et al., 2001). In both experiments the signal to be detected was a Gaussian blob signal, the characteristics of which are similar to the typical mass searched for in real mammograms, as discussed in section 1.6.3 and section 2.15.

The aim of both experiments was to investigate whether mirror and repeat symmetric presentations aid the observer in the detection of a signal in correlated noise fields.

4.2   Experiment 1a

In the first experiment the Gaussian blob signal to be detected was presented on one of two Gaussian white noise backgrounds. This noise permits a simplified formal analysis of the ideal observer (Abbey & Eckstein 2010, pp. 240-244; Kay, 1998, pp. 94-105), which we present in section 4.2.1. The experiment tested whether human observers can adopt the same strategy as the

ideal observer to take advantage of the image pair symmetry to assist in the detection of a signal in noise.

4.2.1   Theory

4.2.1.1   Theory for the detection of a signal in paired noise fields

The task is a simple signal known exactly (SKE) detection task, detecting a Gaussian blob signal which is placed in the centre of either the left or right image (analogous to the left and right breast images of a mammogram reading). The signal is added to a Gaussian white noise background. The noise within each image is white (having no spatial correlation), but is correlated between the images. The correlation value of the image pair (ρ) ranges from $\rho = 0$ (asymmetric) to $\rho = 1$ (symmetric). On each trial the observer receives two images, each of which we will flatten into a vector and place in a column. The observer receives:

$$r = s + n \tag{4.2}$$

where r, s, and n are matrices having number of rows equal to the number of pixels and two columns, r is the pair of displayed images, the columns of s contain either zero or the Gaussian blob signal (depending on whether the signal is on the left or right), and n is Gaussian noise. The columns of n have 2×2 covariance matrix K. Note that each column of n (each noise image) is Gaussian white noise – there is no spatial correlation within each image. However the two columns of n are correlated. The ideal strategy for deciding which image contains the signal has two stages. First the columns of r are decorrelated, by multiplying by the noise covariance matrix's inverse K^{-1}. Then the resulting decorrelated matrix is cross-correlated with the signal (Kay, 1998, p. 106). Thus the observer computes:

$$(rK^{-1})^t s_1 \tag{4.3}$$

where s_1 is the nonzero column of s and t denotes transpose. The observer decides "left" if the cross-correlation value of the first element is larger, and "right" otherwise.

The performance of this decorrelating and cross-correlating observer in the 2AFC detection task where the columns of n are uncorrelated is given by:

$$d' = \sqrt{\frac{2E}{\sigma^2}} \tag{4.4}$$

(Green & Swets, 1966, p. 165) where E is the energy of the signal $\sum s_1^2$, σ^2 is the noise variance,

and d' is the usual signal detection theory measure of performance. The performance when the noise in the two images has correlation ρ is:

$$d' = \sqrt{\frac{2E}{(1-\rho)\sigma^2}} \tag{4.5}$$

(Kay, 1998, p. 112)

Thus, as the correlation between the two noise fields increases, the detectability of the signal rises. In the limit, when the two noise fields are identical, performance is perfect. A simple way of visualising this is to think of the decorrelating algorithm as differencing the two images such that, if the noise in the two fields is identical, the noise will be completely removed, leaving only the signal and hence, performance is error-less. Note, however, that as the correlation between the two images declines, the effectiveness of a differencing algorithm, in relation to a decorrelation algorithm, will also decline.

Taking the level of d' at detection threshold to be 1, the energy threshold E_t is:

$$E_t = \frac{(1-\rho)\,\sigma^2}{2} \tag{4.6}$$

The contrast threshold is equal to some constant a times the square root of the energy threshold where the constant a is dependent upon the stimulus duration and signal area. Signal area remains constant throughout the experiment and, for real observers, temporal summation means that the signal energy quickly reaches a maximum and increasing the time function does not increase signal energy beyond this maximum, constant, value.

Therefore, for the ideal observer the contrast threshold C_t is:

$$C_t = a\sqrt{\frac{(1-\rho)\,\sigma^2}{2}} \tag{4.7}$$

The ideal observer is able to take advantage of increasing levels of symmetry of the two-field display, with the detection threshold declining as the image inter-correlation increases. If real observers are able to resemble the ideal, we would expect their thresholds to decline likewise.

4.2.1.2 Comparison of observer strategies using relative threshold contrast

Without a precise measure of the real observer's efficiency value and the viewing time for each, a direct comparison of the absolute values of contrast threshold calculated for the ideal observer against those actually recorded by the real observer is of limited value. The comparison of absolute contrast threshold values against the ideal observer is, however, unnecessary for the aim of determining whether the real observer can adopt the optimum strategy of decorrelation, as used

by the ideal observer, for the detection of a signal in paired noise fields. To compare the strategies used we need to compare the response curve for each observer (ideal and real) as the correlation between the paired images is varied from zero to one.

This can be achieved by measuring how the contrast threshold for the detection of the signal changes from no symmetry (a correlation of $\rho = 0$) to perfect symmetry (a correlation of $\rho = 1$). We would expect, if symmetry helps, that contrast threshold would fall as the level of correlation increases. This can most clearly be seen by representing the contrast thresholds for each correlation relative to the value at a correlation of zero, thus showing how the contrast threshold changes as the correlation level is increased, as shown in Figure 4.1.

From Equation 4.7, for the ideal observer, the contrast threshold equals:

$$C_t = a\sqrt{\frac{(1-\rho)\,\sigma^2}{2}} \tag{4.8}$$

At a correlation of zero, $\rho = 0$, therefore:

$$C_{zero} = a\sqrt{\frac{\sigma^2}{2}} \tag{4.9}$$

Therefore, for the ideal observer, the relative contrast threshold for each value of correlation (C_ρ), relative to the contrast threshold at $\rho = 0$ equals:

$$Relative\ Contrast\ Threshold = \frac{C_\rho}{C_{zero}} = \frac{a\sqrt{\frac{(1-\rho)\sigma^2}{2}}}{a\sqrt{\frac{\sigma^2}{2}}} \tag{4.10}$$

Which simplifies to:

$$Relative\ Contrast\ Threshold = \sqrt{1-\rho} \tag{4.11}$$

Thus, we can see that the ideal observer will exhibit a falling contrast threshold as inter-image correlation (symmetry) increases. For real observers, we can determine their rate of change of contrast threshold, and hence, their ability to use the correlation between the two noise fields, by fitting the model:

$$Relative\ Contrast\ Threshold = (1-k) + k\sqrt{1-\rho} \tag{4.12}$$

The constant k measures the degree to which an observer's performance is improved by the correlation between the two noise fields and, from equations 4.11 and 4.12, we can see that for the ideal observer $k = 1$. We shall refer to k as the symmetry improvement factor. For the real observer,
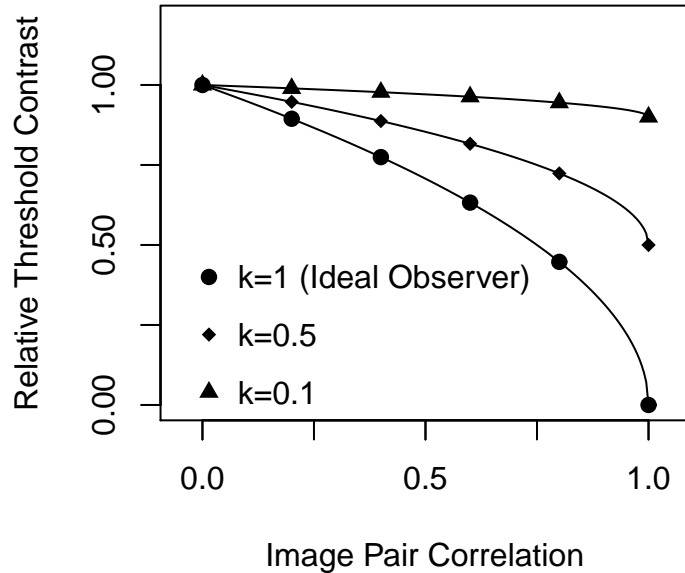
Figure 4.1: A plot of threshold contrast relative to the threshold contrast for an image pair correlation of zero for the detection of a signal in a noise background against the inter image correlation from $\rho = 0$ to $\rho = 1$. Three plots are shown for three values of the symmetry improvement factor (k): $k = 1$ (the ideal observer), $k = 0.5$ (an observer able to make moderate use of the correlation between the images) and $k = 0.1$ (an observer able to make poor use of the correlation between the images).

if $k = 1$ the observer performs like the ideal observer and contrast threshold declines maximally as the image pair becomes more and more symmetrical. If $k = 0$ the observer's performance is not affected by the symmetry and so the function is flat. For illustration, examples of functions of relative threshold contrast versus correlation for k values of 0.1, 0.5 and 1 are shown in Figure 4.1.

4.2.2   Method

4.2.2.1   Apparatus

Stimuli were presented on an LG 15EL9500-ZA OLED monitor with ultra-short persistence, refreshed at 120 Hz. Fine control of the stimulus contrast (1786 grey levels) was achieved using bit-stealing (Tyler, 1997), and the contrast was calibrated using a Minolta LS-110 photometer. The room lighting was dim and held constant across runs and observers such that the black of the monitor had a luminance of 0.02 cd/m2. Viewing was binocular from a chin-rest at a distance of 52 cm from the monitor screen; at this distance the width of the display was 27.20 deg.

4.2.2.2   Stimuli

Each image contained Gaussian white noise, and one of the two images (randomly either on the left or right) had a Gaussian blob signal added to it. The Gaussian white noise background is generated within the trial programme as a 256 x 256 pixel matrix with each pixel set to a value drawn randomly from a normal distribution of pixel values between 0 and 255, where 0 is pure black and 255 is pure white. The standard deviation of the noise was 0.22. Thus, each noise field was a square patch of 256 x 256 pixels subtending 10.87 deg square, and was surrounded by a grey region having the mean luminance of the noise patches of 25 cd/m2. Two noise patches were generated for each trial and the correlation between the two noise patches was set using a Cholesky transformation. The two fields were separated by 0.22 deg. The signal was a Gaussian blob truncated at ± 3 standard deviations with a spatial SD of 0.43 deg. The blob was always centred in the square noise background region (signal known exactly (SKE)). A small red fixation square was situated at the centre of the screen between the two images. Although each image background was white Gaussian noise, the two backgrounds had a correlation of 0.0, 0.25, 0.5, 0.75, 0.9 or 1.0. Two modes of image presentation were used. In the repeat condition, one noise background was a translated version of the other. In the mirror condition, one background was a reflection about the vertical axis of the other. The duration of the stimulus was dependent upon the response of the observer, with a response terminated display.

4.2.2.3   Procedures

The experiment used a two alternative forced choice (2AFC) paradigm and on each trial the observer was presented a left and a right image, as shown in Figure 4.2. In all experiments, the contrast threshold corresponding to 75% correct in the 2AFC task (d' = 1) was measured. Each pairing of correlation and type of symmetry (repeat or mirror) was presented in one 60 trial block. In one session all conditions were run in random order. In the first session, the observer was given a number of practice runs to become familiar with the procedure and the signal characteristics to facilitate a SKE protocol. On each trial, the observer was presented with a pair of images, and the observer chose the image containing the Gaussian blob signal by button press. The image pair remained on the screen until a response was given. Feedback for wrong responses was indicated by a pulsed enlargement of the central red square. At the start of a block of trials, the contrast was set well above threshold and was always on the right, in order to refresh the observer's memory about the signal. The contrast was varied from trial to trial using a staircase (Levitt, 1971) converging on 71% correct that decreased the contrast after two consecutive correct responses and increased it after one incorrect response. Each observer conducted a minimum of four sessions, where each
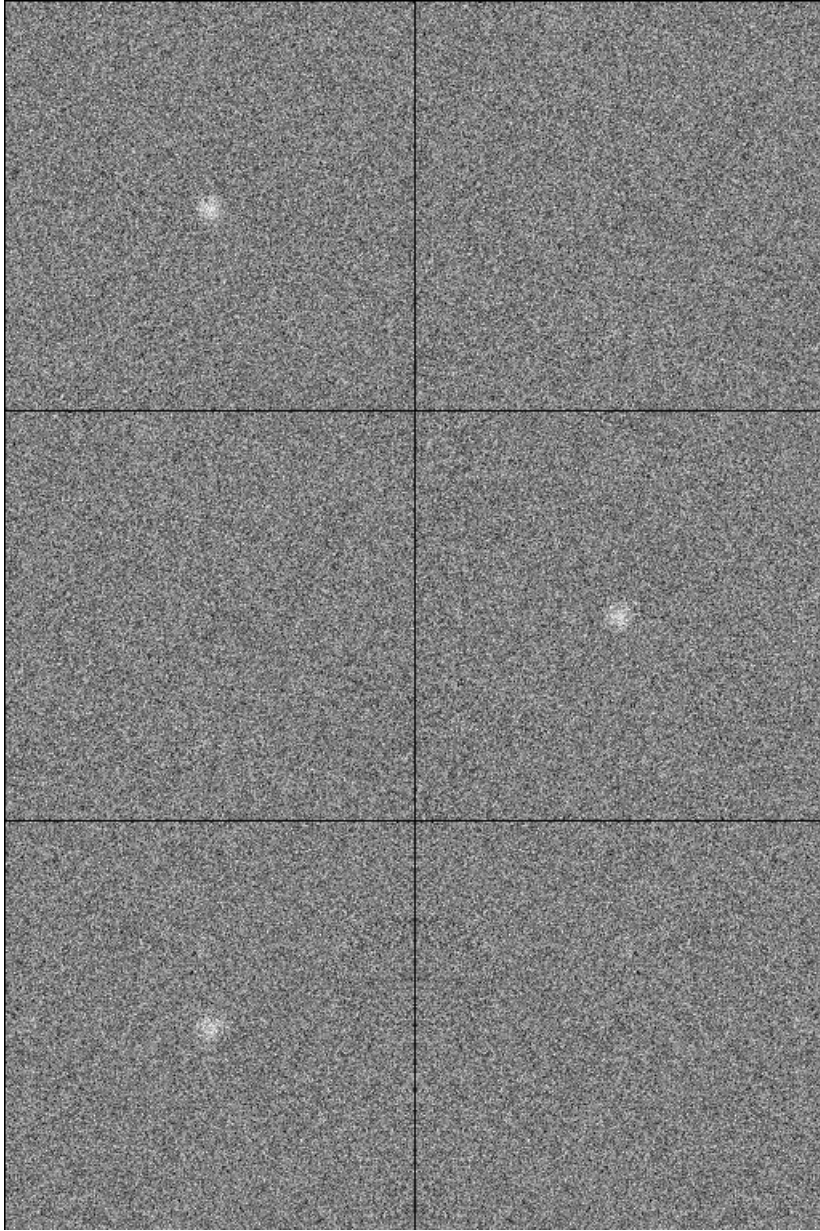
Figure 4.2: Example mirror-symmetric image pairs containing a Gaussian blob signal in Gaussian white noise. From top to bottom, the inter-image correlations are 0, .75, and 1. The blob is well above threshold and is either on the left or right.

session included both symmetry modes and all six correlation levels and took approximately one hour to complete.

4.2.2.4   Observers

Six observers participated in Experiment 1a (three male and three female). Four were inexperienced observers (JR, SP, JN, and KJ) but received training prior to commencing the study. GR and WS were the author and the author's PhD supervisor and both were experienced psychophysical observers. No observer had any background in radiology or medical physics. All observers had normal or corrected-to-normal vision.

4.2.3   Results and discussion

Contrast thresholds for detecting a Gaussian blob were measured as a function of the cross-correlation between the two noise fields presented to the observer. The cross-correlation was a measure of the level of symmetry present. Thresholds for each correlation value were calculated from the 75% correct point of each observer's psychometric function fitted using probit regression. The response on each trial was correct or incorrect, and the probit regression used these binary values. Thus each threshold represents a fit to at least 4 blocks of 60 trials = 240 points. Using the relative values of contrast threshold, determined as shown in section 4.2.1.2, the experiment aimed to investigate whether symmetry, either mirror or repeat, aids the observer in the detection of a signal in correlated Gaussian white noise fields.

4.2.3.1   Mirror symmetric displays - does symmetry aid the observer in the detection of a signal in mirror symmetric noise fields?

Figure 4.3 shows the relative threshold contrast plotted against correlation for the detection of a signal in mirror symmetric paired Gaussian white noise backgrounds for the six observers. The thresholds and 95% confidence intervals obtained by probit regression are shown. Curves were fitted using least squares regression of Equation 4.12 and the fitted parameters are given in Table 4.1. It is clear from Figure 4.3 that real observers' performance does not improve as the correlation between the noise fields increases. The fitted curves are very flat compared to the performance of the ideal observer, as shown by the dotted curves.

In terms of Equation 4.12, the ideal observer has a k value of 1 and the real observers have values close to zero, or even negative. As shown in Table 4.1, the fitted parameter k has a value that is not significantly different from zero for observer JN ($k = 0$, 95% CIs [-0.04, 0.04]), slightly positive for observers JR, SP and GR ($k = 0.15$, 0.18 *and* 0.18 respectively, 95% CIs [0.08, 0.22],

Figure 4.3: Contrast thresholds for detecting a Gaussian blob relative to that obtained when the correlation is zero plotted as a function of the correlation between the two white noise fields. The solid curves are least squares fits of Equation 4.12. The noise fields in the image pair had mirror symmetry. Results for six observers are shown. Error bars show 95% confidence intervals. The dotted curve shows the performance for the ideal observer. The real observers' thresholds do not decline as correlation increases, unlike those of the ideal observer.

Gaussian White Noise			
Mirror Symmetric Display			
Observer	k	LCL	UCL
JR	0.15	0.08	0.22
SP	0.18	0.10	0.26
GR	0.18	0.14	0.22
JN	0.00	-0.04	0.04
WS	-0.45	-0.67	-0.23
KJ	-0.29	-0.33	-0.25

Table 4.1: The symmetry improvement factors (k) and 95% confidence limits for the mirror symmetric displays with a signal in Gaussian white noise for six observers.

[0.10, 0.26], and [0.14, 0.22], respectively), and negative for WS and KJ ($k = -0.45\ and\ -0.29$ respectively, 95% CIs [-0.67, -0.23] and [-0.33, -0.25], respectively). A negative value of k means that performance actually gets worse as the correlation increases. It is clear that for mirror symmetric displays with Gaussian white noise images real observer's performance does not improve, or at best, improves minimally, as the correlation between the noise fields increases. While the ideal observer will take advantage of the noise correlation and, thus, exhibit declining thresholds as noise correlation increases, real observers are not able to decorrelate the image pair.

4.2.3.2 Repeat symmetric displays - does symmetry aid the observer in the detection of a signal in repeat symmetric noise fields?

Whilst bilateral mammograms are usually viewed in a mirror symmetric format, temporal mammograms tend to be viewed across the workstation, effectively in a repeat symmetric format (unreflected about the vertical axis). The experiment, therefore also looked at the effect of symmetry for the repeat displays in the same way as described above. The results of the repeat symmetric condition are shown in Figure 4.4.

The pattern of results is similar to that in Figure 4.3 for the mirror symmetric noise, with fits of Equation 4.12 being markedly flat compared to that of the ideal observer (dotted curve).

The fitted values and 95% confidence intervals for parameter k are shown in Table 4.2. Once again, whereas the ideal observer has a k value of 1, reference to Table 4.2 shows that real observers have much lower values. As shown in Table 4.2, the fitted parameter k has a value that is not significantly different from zero for observers JR and KJ ($k = 0.07$ and -0.04 respectively, 95% CIs [-0.06, 0.20] and [-0.15, 0.13], respectively). Observer JN shows worsening performance as correlation increases ($k = -0.17$, 95% CIs [-0.32, -0.02]) and observers SP, GR and WS show modest improvements in performance ($k = 0.36, 0.21\ and\ 0.15$, 95% CIs [0.20, 0.52], [0.14, 0.28] and [0.01, 0.29], respectively). In all cases, the k values are far from the ideal observer's value of 1 and it is clear that for repeat symmetric displays with Gaussian white noise images real observer's
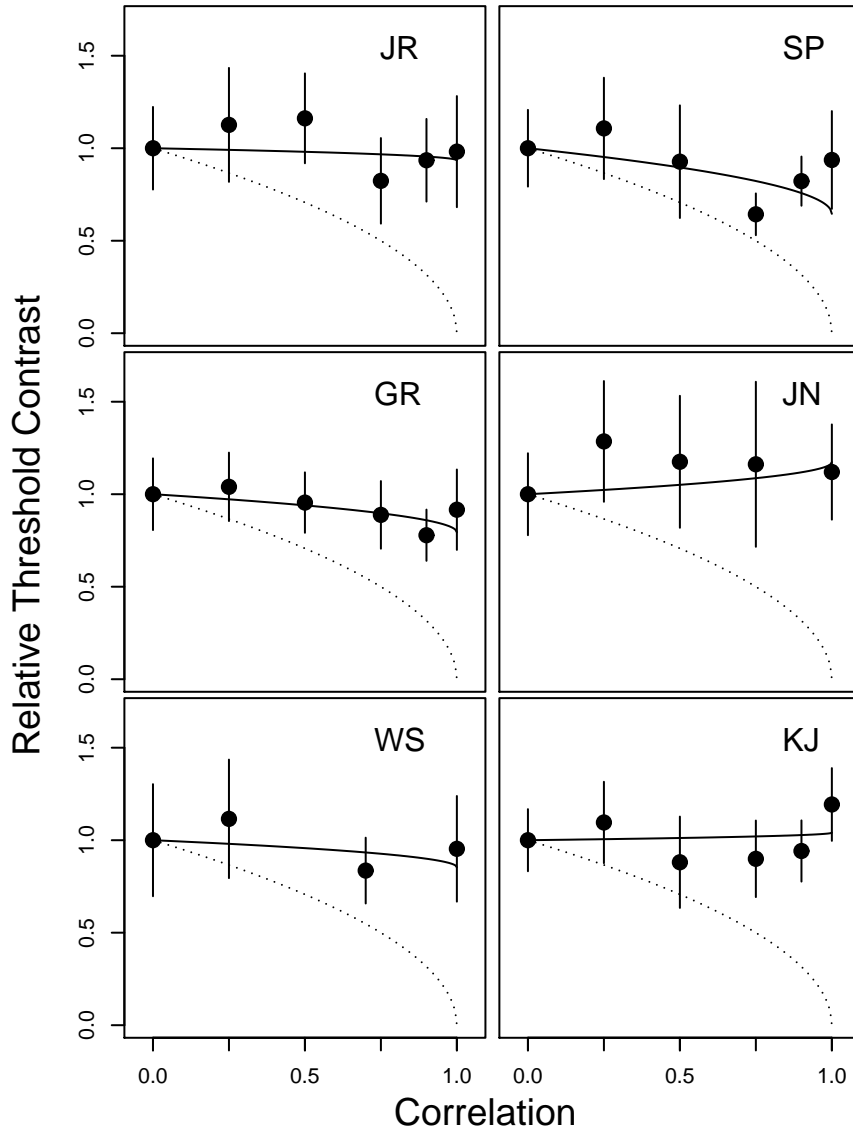
Figure 4.4: Contrast thresholds for detecting a Gaussian blob relative to that obtained when the correlation is zero plotted as a function of the correlation between the two white noise fields. The solid curves are least squares fits of Equation 4.12. The noise fields in the image pair had repeat symmetry. Results for six observers are shown. Error bars show 95% confidence intervals. The dotted curve shows the performance for the ideal observer. The real observers' thresholds do not decline as correlation increases, unlike those of the ideal observer.

Gaussian White Noise			
Repeat Symmetric Display			
Observer	k	LCL	UCL
JR	0.07	-0.06	0.20
SP	0.36	0.20	0.52
GR	0.21	0.14	0.28
JN	-0.17	-0.32	-0.02
WS	0.15	0.01	0.29
KJ	-0.04	-0.15	0.07

Table 4.2: The symmetry improvement factors (k) and 95% confidence limits for the repeat symmetric displays with a signal in Gaussian white noise for six observers.

performance does not improve, or at best, improves moderately, as the correlation between the noise fields increases. While the ideal observer will take advantage of the noise correlation and, thus, exhibit declining thresholds as noise correlation increases, real observers are not able to decorrelate the image pair.

As a whole, the results of Experiment 1a show that increases in noise correlation between the paired backgrounds produce a modest or no decline in detection thresholds for both mirror and repeat displays. The results do not, therefore, provide support for the usefulness of either mirror or repeat symmetry in detecting targets in Gaussian white noise. The intuition that viewing mammograms in mirror-symmetric pairs should help in the detection of tumours is not supported by this experiment.

4.3   Experiment 1b

4.3.1   Introduction

Gaussian white noise in each of the two images was used in Experiment 1a (section 4.2 on page 124) due to the simplicity of the ideal observer. However, in order to make inferences about the usefulness of symmetry in mammogram reading, it would be helpful to present observers with images that are closer to mammograms in appearance, whilst retaining a measure of control over image statistics. As discussed in section 1.6.4 noise backgrounds with a $1/f^3$ power law spectrum provide a suitable level of similarity to real mammograms for detection and search experiments and their use means that we are better able to draw conclusions about performance with real mammograms from out results. To that end, in Experiment 1b we used noise with a $1/f^3$ power spectrum and we will refer to this as "power law noise". The experiment will measure contrast thresholds for detecting a Gaussian blob as the symmetry of the side-by-side display (correlation between the two noise fields) increases. As before, the aim of the experiment is to determine whether symmetry, be it mirror or repeat, aids the observer in the detection of a signal in correlated noise fields, which

in this experiment consist of power law noise.

4.3.2 Theory

For an ideal observer we expect that performance will improve as the symmetry increases. The theory is as given in Experiment 1a Theory section 4.2.1, with one extra element. The power law noise has a low-pass spectrum and therefore is spatially correlated. Therefore the ideal observer will first pre-whiten each noise field (Burgess, 2010, pp. 26-46; Bochud, 2013, pp. 153-164) prior to decorrelating the image pair and cross-correlating with the signal template. Since the first step of pre-whitening reduces the spatially correlated noise to white noise, there are no changes to the theoretical development given in Experiment 1a Theory section 4.2.1.

4.3.3 Method

The apparatus, stimuli and procedures for Experiment 1b were identical to those used in Experiment 1a with the exception that a low-pass $1/f^3$ power law noise background was used instead of a white noise background (Figure 4.5). A set of 150 pairs of power law images was generated for pairwise correlation value. These were generated by filtering pairs of correlated Gaussian white noise images. The actual correlation of each resulting noise pair was measured, and pairs having correlation values more than 0.01 away from the nominal level were discarded. The standard deviation of the noise in the images was the same as for the white noise, 0.22. On each trial, a random pair of images was selected from the pool, and a Gaussian blob was added to one of them.

4.3.4 Observers

Six observers participated in Experiment 1b (two male and four female). Four were inexperienced observers (JR, CA, JN, and AW) but received training prior to commencing the study. GR and WS were the author and author's PhD supervisor and both were experienced psychophysical observers. No observer had a background in radiology or medical physics. All observers had normal or corrected-to-normal vision and each observer conducted a minimum of four sessions, where each session included both symmetry modes and all six correlation levels and took approximately one hour to complete.

4.3.5 Results and discussion

Contrast thresholds for detecting a Gaussian blob were measured as a function of the cross-correlation between the two noise fields presented to the observer. The cross-correlation was a

136
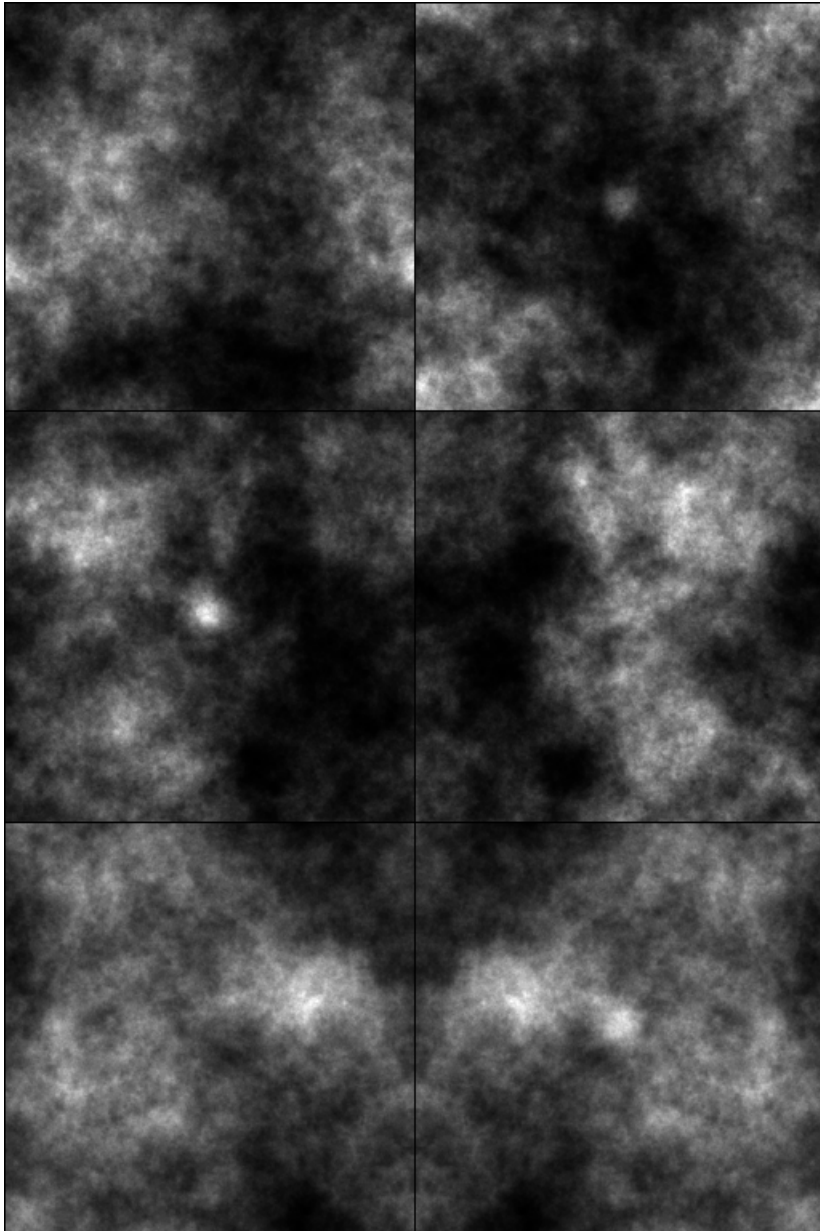
Figure 4.5: Example mirror-symmetric synthetic mammogram image pairs with low-pass $1/f^3$ power spectrum noise. From top to bottom, the inter-image correlations are 0, .75, and 1. A Gaussian blob having contrast well above threshold is shown on the left or right.

measure of the level of symmetry present. Thresholds for each correlation value were calculated from the 75% correct point of each observer's psychometric function fitted using probit regression. The response on each trial was correct or incorrect, and the probit regression used these binary values. Thus each threshold represents a fit to at least 4 blocks of 60 trials = 240 points. Using the relative values of contrast threshold, determined as shown in section 4.2.1.2, the experiment aimed to investigate whether symmetry, either mirror or repeat, aids the observer in the detection of a signal in correlated power law noise fields.

4.3.5.1 Mirror symmetric displays - does symmetry aid the observer in the detection of a signal in mirror symmetric noise fields?

Figure 4.6 shows the relative threshold contrast plotted against correlation for the detection of a signal in mirror symmetric paired power law noise backgrounds for the six observers. The thresholds and 95% confidence intervals obtained by probit regression are shown. Curves were fitted using least squares regression of Equation 4.12 and the fitted parameters are given in Table 4.3. It is clear from Figure 4.6 that, whilst observers JR and GR do show some improvement, for most real observers' performance does not improve as the correlation between the noise fields increases. Their fitted curves are very flat compared to the performance of the ideal observer, as shown by the dotted curves.

Power Law Noise			
Mirror Symmetric Display			
Observer	k	LCL	UCL
AW	0.10	0.02	0.18
JR	0.39	0.31	0.47
CA	-0.03	-0.09	0.03
GR	0.52	0.45	0.59
JN	-0.03	-0.09	0.03
WS	-0.11	-0.21	-0.01

Table 4.3: The symmetry improvement factor k and 95% confidence limits for the mirror and repeat conditions with a signal in low-pass $1/f^3$ power law noise for six observers.

As with Gaussian white noise in Experiment 1a, the symmetry improvement factors are much lower than the ideal value of $k = 1$. As shown in Table 4.3, the fitted parameter k has a value that is not statistically different from zero for observers JN and CA ($k = -0.03$ and -0.03 respectively, 95% CIs [-0.09, 0.03] and [-0.09, 0.03], respectively) and slightly positive for AW (($k = 0.1$, 95% CIs [0.02, 0.18]). Observers JR and GR show moderately positive values ($k = 0.39$ and 0.52 respectively, 95% CIs [0.31, 0.47] for observer JR and [0.45, 0.59] for observer GR, and observer WS shows a slightly negative ($k = -0.11$, 95% CIs [-0.21, -0.01]). A negative value of k means that performance gets worse as the correlation increases. From these data it is clear that mirror
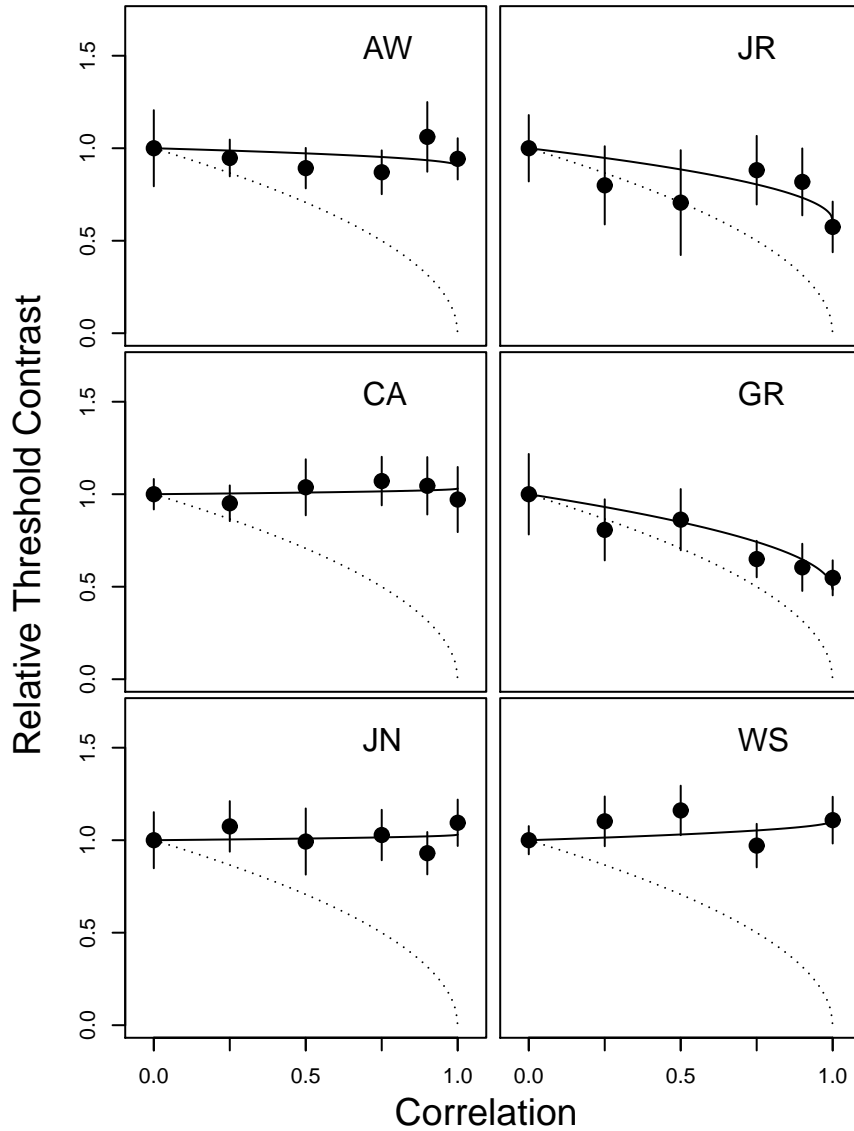
Figure 4.6: Contrast thresholds for detecting a Gaussian blob relative to that obtained when the correlation is zero plotted as a function of the correlation between the two power law noise fields. The solid curves are least squares fits of Equation 4.12. The noise fields in the image pair had mirror symmetry. Results for six observers are shown. Error bars show 95% confidence intervals. The dotted curve shows the performance for the ideal observer. The real observers' thresholds do not decline as correlation increases, unlike those of the ideal observer.

symmetry provides, at best modest, but in most cases, little or no help when trying to detect a signal in paired noise fields.

4.3.5.2 Repeat symmetric displays - does symmetry aid the observer in the detection of a signal in repeat symmetric noise fields?

The experiment also looked at the effect of symmetry for the repeat displays in the same way as described above. The results of the repeat symmetric condition are shown in figure 4.7. The pattern of results is similar to that in figure 4.6 for the mirror symmetric noise, with fits of Equation 4.12 being markedly flat compared to that of the ideal observer (dotted curve).

The values of the symmetry improvement factor (k), derived from the linear regression analysis of the variation of relative threshold contrast with correlation for the detection of a signal in paired power law noise fields displayed in a repeat symmetric format are shown in Table 4.4.

Power Law Noise			
Repeat Symmetric Display			
Observer	k	LCL	UCL
AW	0.09	0.04	0.14
JR	0.48	0.27	0.69
CA	0.06	0.00	0.12
GR	0.48	0.38	0.58
JN	-0.13	-0.21	-0.05
WS	0.11	0.08	0.14

Table 4.4: The symmetry improvement factor k and 95% confidence limits for the mirror and repeat conditions with a signal in low-pass $1/f^3$ power law noise for six observers.

Once again, the symmetry improvement factors are much lower than the ideal value of $k = 1$. As shown in Table 4.4, the fitted parameter k has a value that is not statistically different from zero for observer CA ($k = 0.06$, 95% CIs [0.00, 0.12]) and slightly positive for AW and WS (($k = 0.09$ and 0.11 respectively, 95% CIs [0.04, 0.14] and [0.08, 0.14] respectively). Observers JR and GR show moderately positive values ($k = 0.48$ for both observers, 95% CIs [0.27, 0.69] and [0.38, 0.58], respectively and observer JN shows a slightly negative ($k = -0.13$, 95% CIs [-0.21, -0.05]). A negative value of k means that performance gets worse as the correlation increases. In all cases, the k values are well below the ideal observer's value of 1. From these data it is clear that repeat symmetry provides, at best modest, but in most cases, little or no help when trying to detect a signal in paired power law noise fields. This tells us that symmetry, be it mirror or repeat, does not provide much help to the observer for the detection of a signal in one of the images.
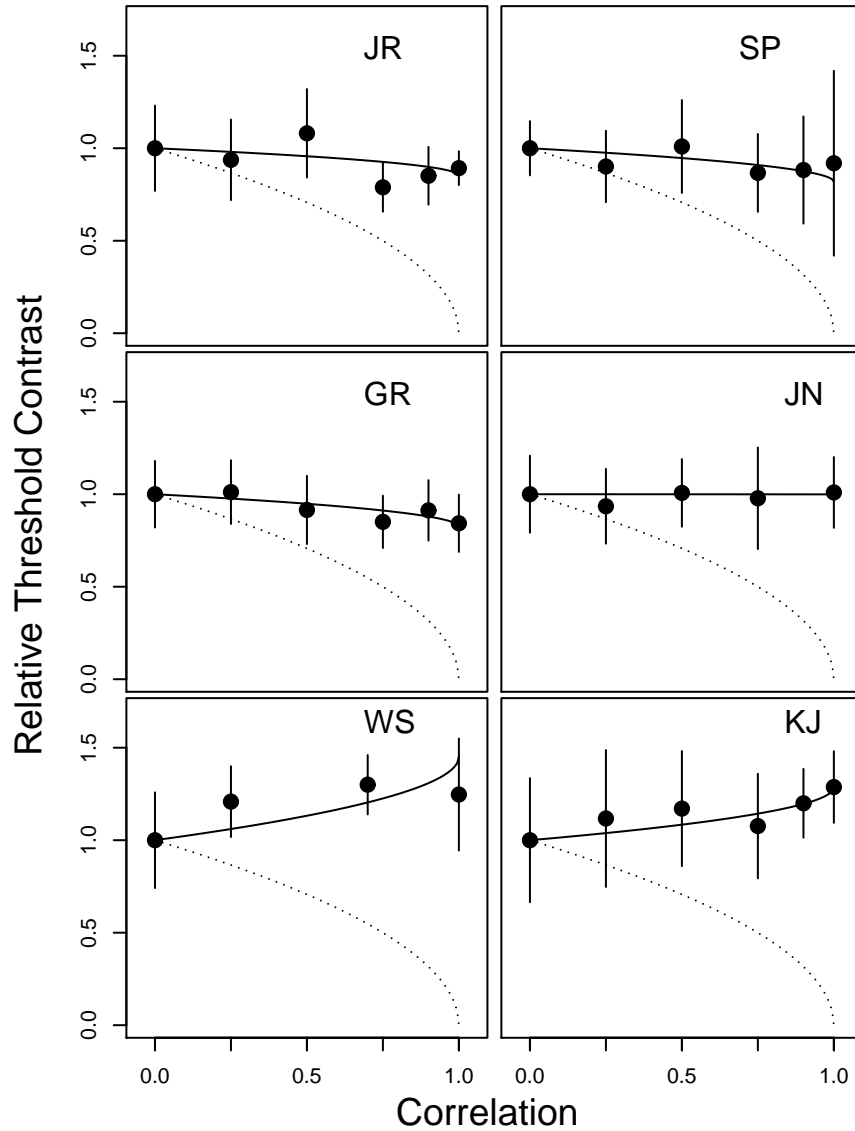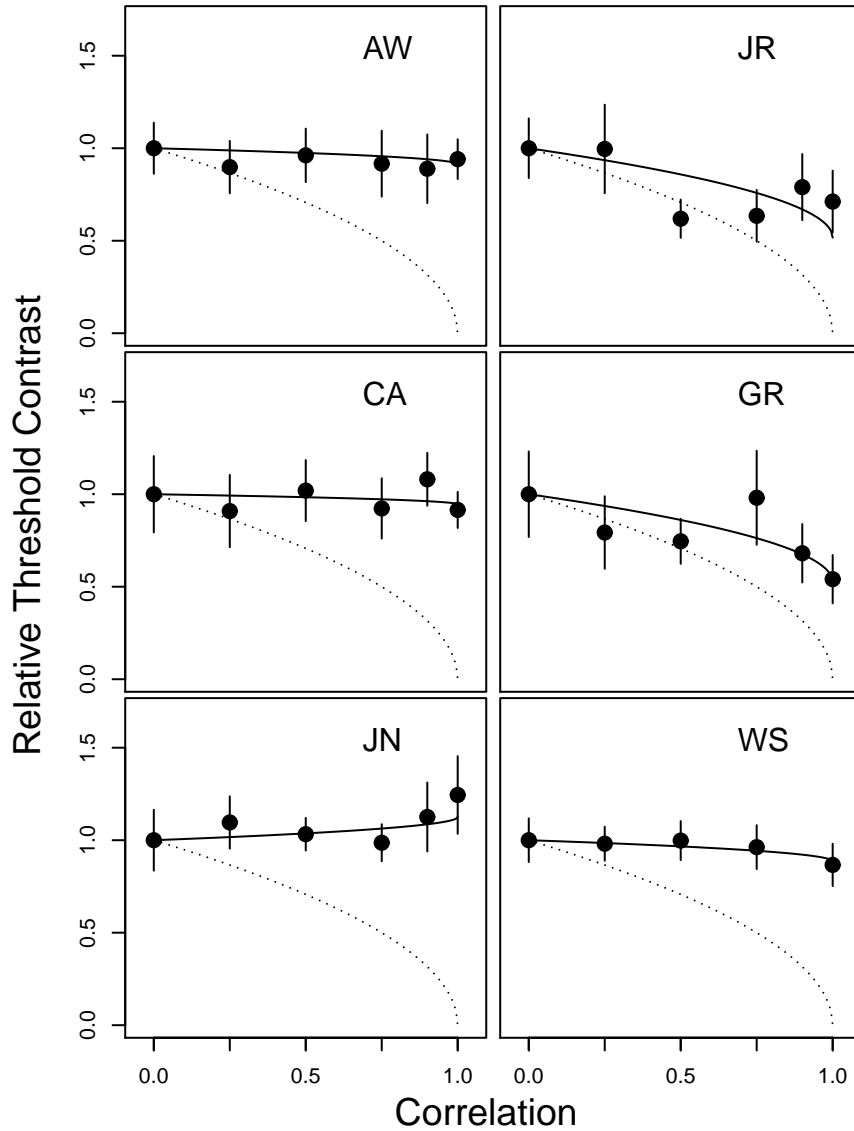
Figure 4.7: Contrast thresholds for detecting a Gaussian blob relative to that obtained when the correlation is zero plotted as a function of the correlation between the two power law noise fields. The solid curves are least squares fits of Equation 4.12. The noise fields in the image pair had repeat symmetry. Results for six observers are shown. Error bars show 95% confidence intervals. The dotted curve shows the performance for the ideal observer. The real observers' thresholds do not decline as correlation increases, unlike those of the ideal observer.

4.4   General Discussion

Radiologists conventionally view mammograms in a mirror symmetric side-by-side display, with the intuition that an abnormality will stand out against the symmetrical background. This is a time honoured clinical technique for large differences, guiding the radiologist in further search. However, it is not known how much the detection of more localised lesions is aided when they are embedded in symmetric noise backgrounds. The purpose of this study was to examine, using controlled laboratory stimuli, whether observers can use the symmetry present between the image pair to improve their performance in the detection of a small Gaussian blob representative of a small tumour. The symmetry was manipulated by varying the correlation between the two noise fields presented on each trial. An ideal observer's contrast thresholds decline in proportion to $\sqrt{1-\rho}$ where ρ is the correlation between the two noise fields. We found that real observers were very unlike an ideal observer, showing little or no benefit of symmetry when detecting a low contrast Gaussian blob (simulated small mass).

The first experiment was conducted with a Gaussian white noise background and it was predicted that the ideal observer would decorrelate the two images prior to cross-correlating the noisy stimulus received with a signal template. It was, therefore, expected that increases in noise correlation resulting from greater symmetry of the alternative image pairs would confer an advantage, producing a reduction in contrast threshold. The second experiment was conducted with a power law noise background and it was predicted that the ideal observer would first pre-whiten each image before utilizing the same strategy as with Gaussian white noise, of decorrelating the image pair prior to cross-correlating the noisy stimulus with a signal template.

To gain an intuition into decorrelation, consider an observer who differences the two images. This is a form of decorrelation, though it is inferior to a method that multiplies the images by the inverse of their covariance matrix. As the noise fields in the image pair become more and more similar, the difference becomes smaller and smaller. In the limiting case of identical noise in the two images, the difference is zero and so the noise is removed altogether, leaving only the signal. The aim of each experiment was to determine whether the human observer would utilise a strategy similar to that of the ideal observer when detecting a blob in Gaussian white noise and power law noise backgrounds. The results of this study show that human observers do not behave like the ideal observer; their thresholds for detecting the signal were essentially unaffected or, at best, only moderately affected by the correlation of the two noise fields (degree of symmetry).

The literature on the use of correlated noise backgrounds to aid in signal detection is limited (see section 1.7.2 on page 91 for a review). Burgess & Colborne (1988) report that for one observer, 2AFC detection was done with $d' = 1.26$ for totally uncorrelated white noise fields and $d' = 2.00$ for

identical noise fields (our repeat condition). Assuming that this difference is statistically significant, the Burgess & Colborne study suggests that in some cases observers can decorrelate the paired noise fields. This was not true for our observers with white noise fields that were uncorrelated or identical, as is shown in Figure 4.3 – the relevant points are at correlations of 0.0 and 1.0, and none of these observers shows a benefit of correlated noise fields. Ahumada & Beard (1997) also conducted a study similar to the current study and, whilst their aims were different to ours, they also reported no significant difference in detection threshold levels between image pair backgrounds that were symmetric or non-symmetric (identical or not), a result that is consistent with the current study. Ahumada & Beard (1997) used a white noise, which has equal power at all frequencies including high spatial frequencies, and this was similar to the stimuli used in the current experiment. This may constitute a key factor in the observers being unable to decorrelate the noise fields. Pre-whitening is a form of decorrelation and, as discussed in section 1.6.6, human observers have been found to be unable to pre-whiten high frequency noise (Myers, 1985; Myers et al., 1985), as contained in the white noise used in the study by Ahumada & Beard and in the current study.

Pre-whitening is usually associated with the simplification of noise with a power law spectrum to improve the efficiency of the human visual system and it may be worth, at this juncture, reminding the reader why we are attempting to pre-whiten noise that already has a white noise power spectrum. Whilst each image is, itself, white, there is, nonetheless, a correlation between the two images and it is this correlation between the two images that the ideal observer takes advantage of. The process of pre-whitening is discussed in section 1.6.6 and, just as the evidence presented in section 1.6.6 supports the notion that the human visual system is able to use pre-whitening to remove redundancy within a scene, it is also suggested that it will also be able to, similarly, remove redundant information between two adjacent scenes. Whilst pre-whitening is the process that underpins both redundancy reduction within a scene and redundancy reduction between two scenes, to maintain clarity, the former will be referred to as pre-whitening and the latter as decorrelation. Decorrelation is, thus, a method of redundancy reduction that can be applied to paired images and its effect is to remove correlated noise between the two images, therefore making the signal easier to see. It is interesting to note that two observers (JR and GR) did show a modest improvement in their ability to use the correlation in Experiment 1b (section 4.3), where the images were power law noise, with both mirror and repeat displays. Although the other observers that participated in both experiments were unable to utilise the correlation in either experiment, the modest improvements shown by JR and GR may be indicative of a greater ability to pre-whiten noise with a low pass spectrum (Abbey & Barrett, 2001; Abbey & Eckstein, 2007; Burgess, 1999; Burgess & Judy, 2007; Rolland & Barrett, 1992) and, therefore, make use of the

correlation between the image pair by removing correlated (and, therefore, redundant) information.

Notwithstanding the improvement in the ability to use the correlation between the images seen in power law noise as exhibited by observers JR and GR, this is still relatively modest in relation to the ideal observer. As a whole, both Experiment 1a and 1b demonstrate that real observers are not able to use the correlation present between paired images when they are presented in a mirror symmetric or repeat symmetric format.

One observation made during the experiments, though not measured, was that observers GR and JR appeared to deliberate longer than other observers. Both observers suggested that they were consciously discounting features that appeared in both images and, hence, seemed to be carrying out a post-attentive decorrelation strategy rather than, perhaps, a pre-attentive decorrelation strategy used by other observers. The results in Gaussian white noise suggest that this strategy did not confer any advantage, but, in power law noise both GR and JR performed significantly better than any other observer. The lower spatial frequency of power law noise with its more distinct "clumpy" features may have enabled their post-attentive decorrelation strategy to be easier to carry out and hence more effective at discounting noise. The anecdotal nature of these observations and the experimental procedure of unlimited viewing times and no fixation or limitation on eye movements means that it is not possible to address this point. The use of this protocol was chosen to more closely replicate the task of the radiologist and is a recognised methodology (Burgess & Ghandeharian, 1984a,b; Burgess, 1985; Burgess & Colborne, 1988), however, to address this point and to rule out a post-attentive decorrelation strategy would require a repeat of the experiment with limited viewing time and fixation to minimise eye movements.

The inability to match the performance of the ideal observer may also stem from the two stage nature of the detection process: decorrelation of the image pair followed by cross-correlation. In the first stage the image pair is viewed as a whole to facilitate decorrelation and in the second stage the observer is required to view each half of the image separately to obtain a cross-correlation value for each half. This is a statistical exercise for the ideal observer and, therefore, achievable without loss. However, if the real observer is carrying out a similar, two stage process, then the process of forming an accurate internal representation of the decorrelated image, holding it in iconic memory and cross-correlating each half without loss would be a challenging task for the visual system. Whilst iconic memory is a high capacity memory store, retention time is very short (typically <1sec) and this may degrade performance as the internal representation fades.

Although only apparent for two observers (WS and KJ in Gaussian white noise; WS in power law noise) it is not clear why they showed an improvement in performance with repeat over mirror symmetry, particularly as the predictions for an ideal observer suggest that both conditions will

see similar advantages of the correlated noise and thus will have similarly declining thresholds as noise correlation increases. It is interesting that no observers show the opposite trend of improved performance in mirror symmetry over repeat symmetry. Despite the comparative improvement seen in these two observers in repeat displays, their function was still essentially flat and did not indicate any ability to use the correlation present between the images of the pair.

Overall, the results show that mirror symmetry and repeat symmetry do not confer any more than weak benefits in the detection of small synthetic blobs in correlated noise fields. Whilst, some observers did show an advantage in signal detection in symmetric over non-symmetric displays, for both types of symmetry, this was no more than a modest improvement. The fact that this occurred more predominantly in power law noise than Gaussian white noise where, as discussed in section 1.6.6, the human visual system is more efficient at pre-whitening (decorrelating) images with a low-pass spectrum, suggests that it may be image pair decorrelation, rather than symmetry per se, that determines this improvement. In addition, this improvement was predominantly shown by only two observers and may have been the result of a post-attentive discounting strategy; further research would be necessary to investigate this. The results also show that mirror symmetric presentation confers no benefit over translational repeat symmetry in the detection of small synthetic blobs in correlated noise fields. he current study has used synthetic noise backgrounds and synthetic signals and, thus, does not enable us to draw conclusions about the performance of radiologists searching for tumours in real mammograms. Having established a theoretical baseline, this thesis will progress to test our findings in a more clinical scenario. However, based on the current results, if we were to extrapolate our findings with synthetic mammograms and synthetic tumours to the clinical world of real mammograms and real tumours, it would suggest that some small fraction of radiologists may show limited benefit from symmetric (repeat and mirror) mammogram presentation, although most will not. It is clear that symmetric displays do not enable the observer to use the correlation between the images and, therefore, if a display method can assist observers in decorrelating image pairs, then improvements in performance may be possible. The next stage of this thesis, before progressing to the use of a more clinical experimental set up, will be to investigate an alternative, movie, presentation mode. The aim of this will be to determine whether a movie presentation mode can facilitate the use of the correlation present between image pairs to improve observer performance for the detection of a signal in one image of the pair.

Chapter 5

Experiment 2: A Comparison of Side-by-Side Versus Animated Presentation of Images

5.0.1 Introduction

If mirror symmetric presentation is beneficial to the observer in detecting a signal in one image of a pair then we would expect signal detection performance to improve as the level of symmetry improved. We found this not to be the case and the results of Experiment 1a (see section 4.2 on page 124) and 1b (see section 4.3 on page 135) suggest that mirror symmetric presentation does not confer any more than limited benefits to the observer. The observers' performance with a mirror symmetric display was no better than with a repeat symmetric display and both showed only weak to moderate improvements in performance as the level of symmetry improved.

Whilst the ability of the human visual system to use symmetry as an aid to signal detection may, therefore, be questionable, its sensitivity to motion and flicker is less so (Adelson & Bergen, 1986; Watson, 1986). With species survival dependent upon the detection of predators and prey it is of little surprise that motion is a highly salient property and it is, as stated by Hubel (1995, pp. 78-79) "no wonder that most cortical cells respond better to a moving object than to a stationary one". Flicker is an integral aspect of motion detection and it is, therefore, of little surprise that flicker is also a highly salient visual property in its own right.

Spalek et al. (2009) showed that flicker is able to grab attention in the same way as colour and motion. They found that a flickering target 'popped out' from a field of distractors producing a

flat search profile as a function of the number of distractors, supporting their assertion of flicker as a primitive visual feature. Similar findings have been reported by Franconeri et al. (2005) who found that luminance based transients, such as flicker, captured attention and Ludwig et al. (2008) whose results demonstrated that flicker was particularly disruptive in drawing attention away from a specified target, illustrating its effectiveness in capturing attention. Despite the large body of literature associated with flicker in general, there appears to be little of this related to signal detection, however, it presents an interesting alternative to symmetric presentation to make tumours salient. Instead of placing mammograms in a mirror symmetric configuration side-by-side, they can be presented sequentially in the same location in a two-frame animation sequence such that any difference between the images will result in highly salient motion and flicker. For example, a tumour that is present in one image and absent in the other will be seen to flash on and off. The continuous two-frame animation sequence can be likened to a movie and will be referred to as movie presentation throughout this study. By contrast the mirror symmetric displays are static and will be referred to as static displays. This study is interested in testing whether the movie presentation of a pair of images can aid the observer in the detection of a signal in one of the images and whether it can confer an advantage in this task over the observer using a static presentation technique.

Of course, any differences, other than a potential tumour, will also be made salient and this is an important observation in a clinical scenario as no two breasts, nor even two mammograms of the same breast taken at different times, will be identical. These differences arise from natural variations between left and right breast and natural changes in breast composition between screenings as well as differences introduced during the screening process as a result of the deformability of breast tissue and breast positioning in the scanner. Intuitively we would expect the task of detecting a signal to become more difficult as the two images become less similar and this is what the ideal observer analysis shows (see Theory section in Experiment 1a, section 4.2.1 on page 125). In a clinical situation, with real image pairs, misalignment or skewing between the images can be compensated for, to a certain extent, by image registration (Guo et al., 2006). Image registration is a technique used to reduce the level of misalignment between two images and a brief description of registration can be found at section 2.13. As discussed in section 2.13 registration cannot compensate for an incongruence in the features present in each image. Thus, the extent to which two images match is a result of the differences in features present between the images and any misalignment between corresponding features and is typically measured by the correlation between images. The effect of varying how well the two images of a pair are matched needs to be quantified in assessing both a static presentation technique and a movie presentation technique and this can

be done by varying the level of correlation between the image pairs. By presenting the same images in two alternative ways, the effect of varying the correlation level is seen in different ways; when presented side-by-side in a static display the symmetric appearance varies and when presented in an animated movie display the extent and amount of flicker and apparent motion varies.

We can quantify the effect of mis-match between two images by varying the level of correlation between the two images of a pair (as defined in Experiment 1a, section 4.2.1 on page 125) and measuring the contrast threshold necessary for the detection of a signal in one image. If the correlation was zero, the noise was completely unrelated and would indicate no correspondence between the two images at all. If the correlation was 1.0, the two images had identical noise and the image pair would have perfect correspondence. Intermediate levels of correlation produced pairs with varying levels of correspondence reflecting the varying levels of correspondence found in real image pairs. Experiment 2a used a Gaussian white noise background to enable a simplified examination of the theoretical basis for the ideal observer. Experiment 2b used a noise background with a $1/f^3$ power. In both experiments the signal to be detected was a Gaussian blob signal.

The Theory section of Experiment 1a (section 4.2.1 on page 125) shows that the ideal observer achieves optimum performance for the detection of a signal in correlated noise fields, taking advantage of the correlation by decorrelating (or whitening) the two noise fields to increase signal to noise ratio. Experiment 1 demonstrated that static (mirror symmetric) presentation does not enable the real observer to do this. Movie presentation may enable the real observer to decorrelate the paired noise fields and, hence, make use of the correlation between them to improve signal detection performance. Therefore, the aim of Experiment 2a and 2b is to determine whether movie presentation enables the real observer to decorrelate (or whiten) paired noise fields and thus, use the correlation present between the images. Experiment 2a and 2b also replicate the testing done in Experiment 1a and 1b to determine whether static (mirror symmetric) presentation enables the real observer to decorrelate (or whiten) paired noise fields and thus, use the correlation present between the images

5.0.2 Theory

The theory for Experiment 2a is the same as for Experiment 1, as described in the Theory section of Experiment 1a (see section 4.2.1 on page 125).

Figure 5.1: Example two frame animation sequence showing left and right displays, each containing two images displayed sequentially in the same location. One image contains a Gaussian blob signal in Gaussian white noise and is either in the left or right display pair. The blob is shown well above threshold.

5.0.3 Experiment 2a

5.0.3.1 Method

Apparatus   The apparatus used in Experiment 2a was the same as for Experiment 1a.

Stimuli   For the static presentation, the stimuli were the same as the mirror symmetric displays used in Experiment 1a, however, repeat symmetric presentation was not used and a movie presentation mode was used in Experiment 2a. To create the movie stimuli, the noise backgrounds were presented as the same orientation images in the same location, displayed sequentially with frame duration of 250 ms. As with static displays, movie displays were also presented side by side with one side (randomly either the left or right display) containing the noise field with the superimposed blob, as shown in Figure 5.1. In both presentation modes the Gaussian blob signal was the same as used in Experiment 1a.

Procedures   The procedures were the same as used in Experiment 1a for static displays but with a movie display instead of a repeat symmetric display. As in Experiment 1a, the observer was presented a left and a right image, with an example static display as shown in Figure 4.2 and an example movie display shown in Figure 5.1. For both presentation modes, viewing time was unlimited, with a response terminated display.

Observers   Six observers participated in Experiment 2a (three male and three female). Five were inexperienced observers (JR, SP, JN, CH and KJ) but received training prior to commencing the study. GR is the author and an experienced psychophysical observer. No observer had any background in radiology or medical physics. All observers had normal or corrected-to-normal vision.

5.0.4   Results and Discussion

Contrast thresholds for detecting a Gaussian blob were measured as a function of the correlation between the two noise fields presented to the observer. The correlation was a measure of the level of correspondence present between the two images. Thresholds for each correlation value were calculated from the 75% correct point of each observer's psychometric function, fitted using probit regression. The response on each trial was correct or incorrect, and the probit regression used these binary values. Each threshold represents a fit to at least 4 blocks of 60 trials = 240 points. The study aimed to determine whether movie presentation enabled the real observer to decorrelate (or whiten) paired noise fields and thus, use the correlation present between the images. The study also aimed to replicate the work done in Experiment 1a to determine whether static (mirror symmetric) presentation enabled the real observer to decorrelate (or whiten) paired noise fields and thus, use the correlation present between the images.

These aims have been addressed by making a comparison against the ideal observer to determine whether human observers are able to make use of the correlation between the two noise fields for the detection of a signal in one of those noise fields, in a similar manner to the optimum strategy of the ideal observer. As discussed in Experiment 1, whilst a direct comparison of the absolute values of contrast threshold calculated for the ideal observer against those actually recorded by the real observer is of limited value, a comparison of the response curves for each observer (ideal and real) enables the strategies of the respective observers to be compared. This gives an indication of whether the real observer is able to use the correlation present between the images, as exemplified by the ideal observer. We would expect, if the correlation helps, that contrast threshold would fall as the level of correlation increases from a correlation of $\rho = 0$ to a correlation of $\rho = 1$.. This can most clearly be seen by representing the contrast thresholds for each correlation relative to the value at a correlation of zero, thus showing how the contrast threshold changes as the correlation level is increased. This process is explained in Experiment 1a Theory (section 4.2.1 on page 125) but here, we will refer to correlation rather than symmetry.

To clarify, the relative threshold contrast equals:

$$Relative\,Threshold\,Contrast = \frac{C_t\,at\,\rho_x}{C_t\,at\,\rho_0} \tag{5.1}$$

where:

$C_t\,at\,\rho_0$ = contrast threshold at zero correlation

$C_t\,at\,\rho_x$ = contrast threshold at correlation x

where x = inter image correlation = 0, 0.25, 0.5, 0.75, 0.9 or 1.0
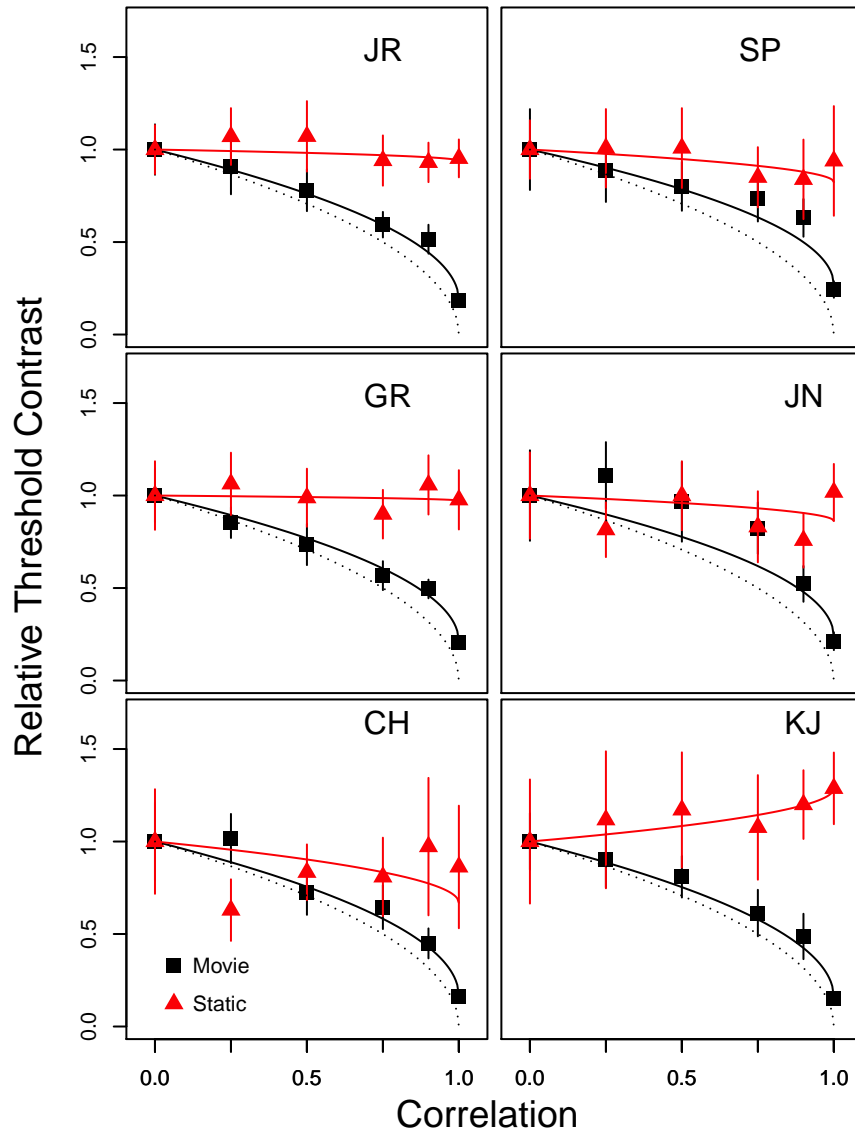
Figure 5.2: Contrast thresholds for detecting a Gaussian blob relative to that obtained when the correlation is zero, plotted as a function of the correlation between the two Gaussian white noise fields. The solid curves are least squares fits of Equation 4.12. The noise fields in the image pair were presented as a static display (filled triangle) or as a movie display (filled square). Results for six observers are shown. Error bars show 95% confidence intervals. The dotted curve shows the performance for the ideal observer.

Using the linear model (Experiment 1a, Equation 4.12 on page 127):

$$Relative\,Threshold\,Contrast = (1-k) + k\sqrt{1-\rho} \qquad (5.2)$$

Where k is the slope of the function and will be called the correlation improvement factor. If $k = 1$, the observer performs like the ideal observer, and threshold declines maximally as the correlation between the image pair goes from $\rho = 0$ to $\rho = 1$. If $k = 0$, the observer's performance is not affected at all by changes in correlation and so the function is flat. The correlation improvement factors for static presentation and movie presentation are shown in Table 5.1 on the next page and the fitted curves for each observer are shown in Figure 5.2.

Figure 5.2 clearly shows the contrast in performance strategies between static presentation, where the fitted curve is very flat, in comparison to the movie presentation, which shows a profile much closer to that of the ideal observer. This can also be seen by reference to Table 5.1 on the following page, which shows that the results for static presentation replicate the results from Experiment 1a. The values of the correlation improvement factor (k) are close to zero, demonstrating very little, or no, improvement of performance as correlation improved. As shown in Table 5.1, for static presentation the fitted parameter k has a value that is not statistically different from zero for observers GR and JN ($k = 0.03$ and 0.14 respectively, 95% CIs [-0.04, 0.09] and [0.00, 0.28], respectively) and slightly positive for JR, SP and CH (($k = 0.06$, 0.18 and 0.33 respectively, 95% CIs [0.02, 0.10], [011, 0.25] and [0.05, 0.62], respectively). Observer KJ shows a negative value ($k = -0.29$, 95% CIs [-0.33, -0.25]). A negative value of k means that performance gets worse as the correlation increases. In every case the fitted curves have a value of k whose 95% confidence interval spans a range that is close to zero and very far from 1.0. From these data it is clear that static presentation provides little or no help when trying to detect a signal in paired noise fields. This confirms the findings of Experiment 1a, showing that mirror symmetry does not enable the real observer to take advantage of the correlation between paired Gaussian white noise fields and provides little or no help when trying to detect a signal in them.

As can be seen in Figure 5.2, the results for movie presentation demonstrate a very different response. Reference to Table 5.1 on the next page shows that the values of the correlation improvement factor (k) are much higher and much closer to the ideal observer's value of $k = 1$, demonstrating a clear improvement of performance as correlation increases. As shown in Table 5.1, the fitted parameter k has a value ranging from $k = 0.73$, 95% CIs [0.70, 0.78] for observer SP to $k = 0.84$, 95% CIs [0.83, 0.86] for observers CH and KJ. From these data it is clear that the real observer using movie presentation behaves much more like the ideal observer, able to take advantage of the noise correlation, exhibiting declining thresholds as noise correlation increases in

contrast to the real observers using static presentation, who are not able to decorrelate the image pair.

| Observer | Gaussian White Noise | | | | | |
| | Static | | | Movie | | |
	k	LCL	UCL	k	LCL	UCL
JR	0.06	0.02	0.10	0.81	0.80	0.82
SP	0.18	0.11	0.25	0.73	0.70	0.78
GR	0.03	-0.04	0.09	0.79	0.78	0.80
JN	0.14	0.00	0.28	0.76	0.70	0.82
CH	0.33	0.05	0.62	0.84	0.83	0.86
KJ	-0.29	-0.33	-0.25	0.84	0.83	0.86

Table 5.1: Correlation improvement factors (k) and 95% upper and lower confidence limits for the static and movie conditions with a signal in Gaussian white noise for six observers.

5.0.5 Experiment 2b

5.0.5.1 Introduction

Gaussian white noise in each of the two images was used in Experiment 2a due to the simplicity of the ideal observer. However, in order to make inferences about the usefulness of static or movie presentation in mammogram reading, it would be helpful to present observers with images that are closer to mammograms in appearance. To that end, in Experiment 2b we used noise with $1/f^3$ power spectrum, which has been found by other authors to have a power spectrum similar to that of real mammograms (Burgess et al., 2001). We will call this "power law noise". As before, the aim of the experiment is to measure contrast thresholds for detecting a Gaussian blob as the correlation between the two noise fields varies.

For an ideal observer we expect that performance will improve as the correlation increases. The theory is the same as for Gaussian white noise, with one extra element. The power law noise has a low-pass spectrum and therefore is spatially correlated. Therefore the ideal observer will first pre-whiten each noise field (Burgess, 2010, pp. 26-46; Bochud, 2013, pp. 153-164) prior to decorrelating the image pair and cross-correlating with the signal template. Since the first step of pre-whitening reduces the spatially correlated noise to white noise the theoretical results are the same as for Gaussian white noise.

5.0.5.2 Method

The apparatus and procedures for Experiment 2b were identical to those used in Experiment 2a. For the static presentation, the stimuli were the same as the mirror presentation used in Experiment 1b with a low-pass $1/f^3$ power law noise background. A repeat presentation was not
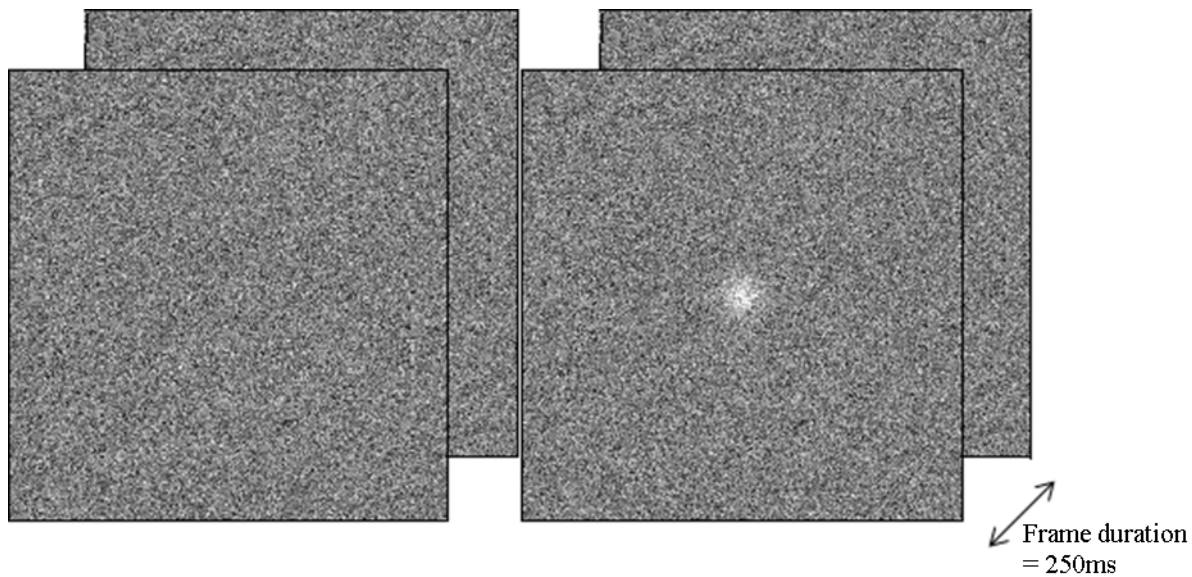
154
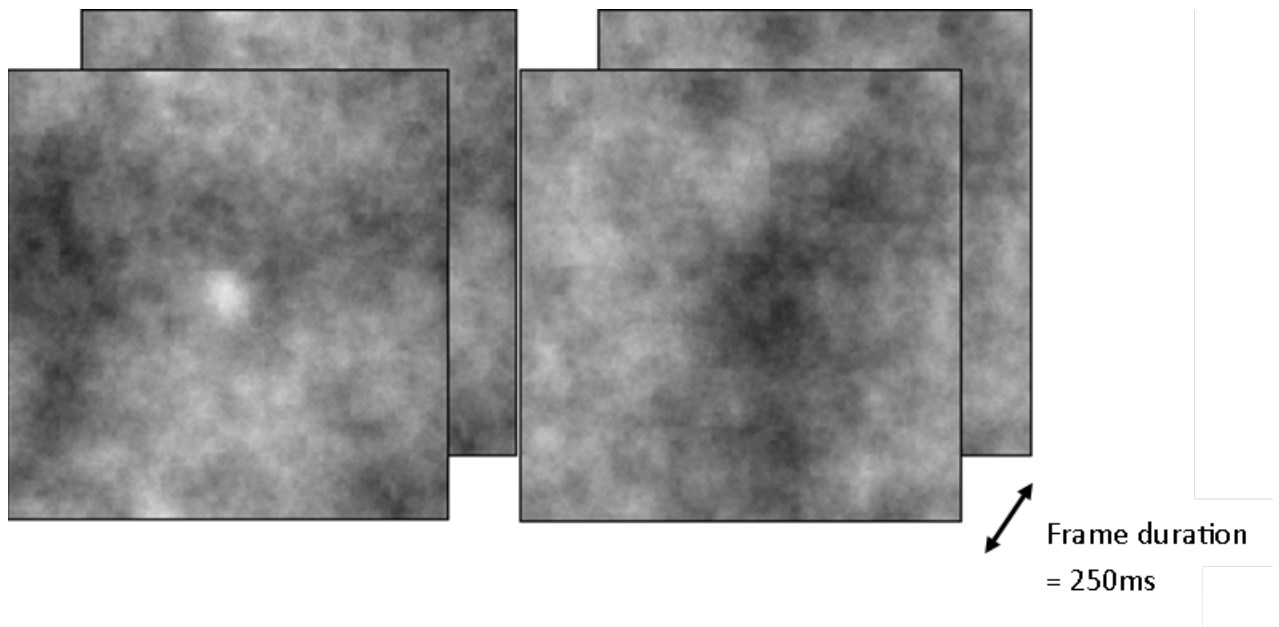
Figure 5.3: Example two frame movie sequence showing left and right displays, each containing two images displayed sequentially in the same location. One image contains a Gaussian blob signal in power law noise and is either in the left or right display pair. The blob is shown well above threshold.

used in Experiment 2b, however a movie presentation was used, again, with a low-pass $1/f^3$ power law noise background (Figure 5.3). The power law noise images were generated as described in Experiment 1b. Six observers participated in experiment 2b (three male and three female). Five were inexperienced observers (JR, CA, JN, AW and KJ) but received training prior to commencing the study. GR is the author and an experienced psychophysical observer. No observer had any background in radiology or medical physics. All observers had normal or corrected-to-normal vision.

5.0.5.3 Results and discussion

Experiment 2b aimed to address the same aim investigated in Experiment 2a; does movie presentation enable the real observer to decorrelate (or whiten) paired noise fields and thus, use the correlation present between the images? Experiment 2b also replicated the testing done in Experiment 1b to determine whether static (mirror symmetric) presentation enables the real observer to decorrelate (or whiten) paired noise fields and thus, use the correlation present between the images. The results were analysed in the same way as Experiment 2a (see section 5.0.4 on page 151) and will use the correlation improvement factor as presented in Experiment 2a as a measure of how well the real observer can use the correlation between the two images. This gives an insight into the strategy being used by the real observers when using movie and static presentation and enables

| Power Law Noise | | | | | | |
| | Static | | | Movie | | |
Observer	k	LCL	UCL	k	LCL	UCL
JR	0.39	0.31	0.47	0.97	0.96	0.97
GR	0.50	0.43	0.57	0.94	0.94	0.95
CA	-0.03	-0.09	0.02	0.42	0.36	0.48
KJ	-0.31	-0.41	0-.21	0.96	0.95	0.97
AW	0.10	0.02	0.08	0.67	0.60	0.74
JN	-0.03	-0.09	0.03	0.65	0.60	0.69

Table 5.2: Correlation improvement factors (k) and 95% confidence limits for the static and movie conditions with a signal in power law noise for six observers.

a comparison against the strategy used by the ideal observer. An ideal observer, using a decorrelating strategy will have a correlation improvement factor of $k = 1$, whereas, a non-decorrelating observer will have a correlation improvement factor closer to zero.

Reference to Figure 5.4 shows that the results for the real observers using static displays follow a similar pattern to that found in Experiment 1b. We see that the fitted curves for the real observers using static displays are essentially flat with only modest reduction in contrast threshold as the inter-image correlation increases. Also as seen in Experiment 2a, Figure 5.4 shows that for the real observers using movie displays, contrast threshold shows a much greater reduction as the inter-image correlation increases.

As in Experiment 2a, we can use the correlation improvement factor, k, to address the aim of the study and compare the performance curves of real observers using static displays and movie displays against those of the ideal observer to determine whether those real observers are adopting a strategy that enables the use of the correlation between the paired images, in a similar manner to the optimal strategy of the ideal observer. The k value is calculated as shown in Experiment 2a and the results are shown in Table 5.2.

For static presentation, two observers show moderate improvements in contrast threshold as correlation increases, with k values for JR and GR of 0.39 and 0.50 respectively, 95% CIs [0.31, 0.47], [0.43, 0.57], respectively. A weak improvement is seen by observer AW ($k = 0.10$, 95% CIs [0.02, 0.08]) and no improvement by observers CA and JN with correlation improvement factors that do not differ significantly from zero with k values equal to -0.03 for both observers, 95% CIs [-0.09, 0.02] and [-0.09, 0.03], respectively. KJ showed a negative k value, indicating that contrast threshold increased and performance got worse as correlation increased (k = -0.31, 95% CIs [-0.41, -0.21]). Overall, with static displays, in power law noise, we see a range of responses to increases in correlation from performance getting worse, through little or no change, to a moderate improvement in performance. However, the real observers using a static display are not able to approach the optimum use of the correlation between the two images, as shown by the ideal
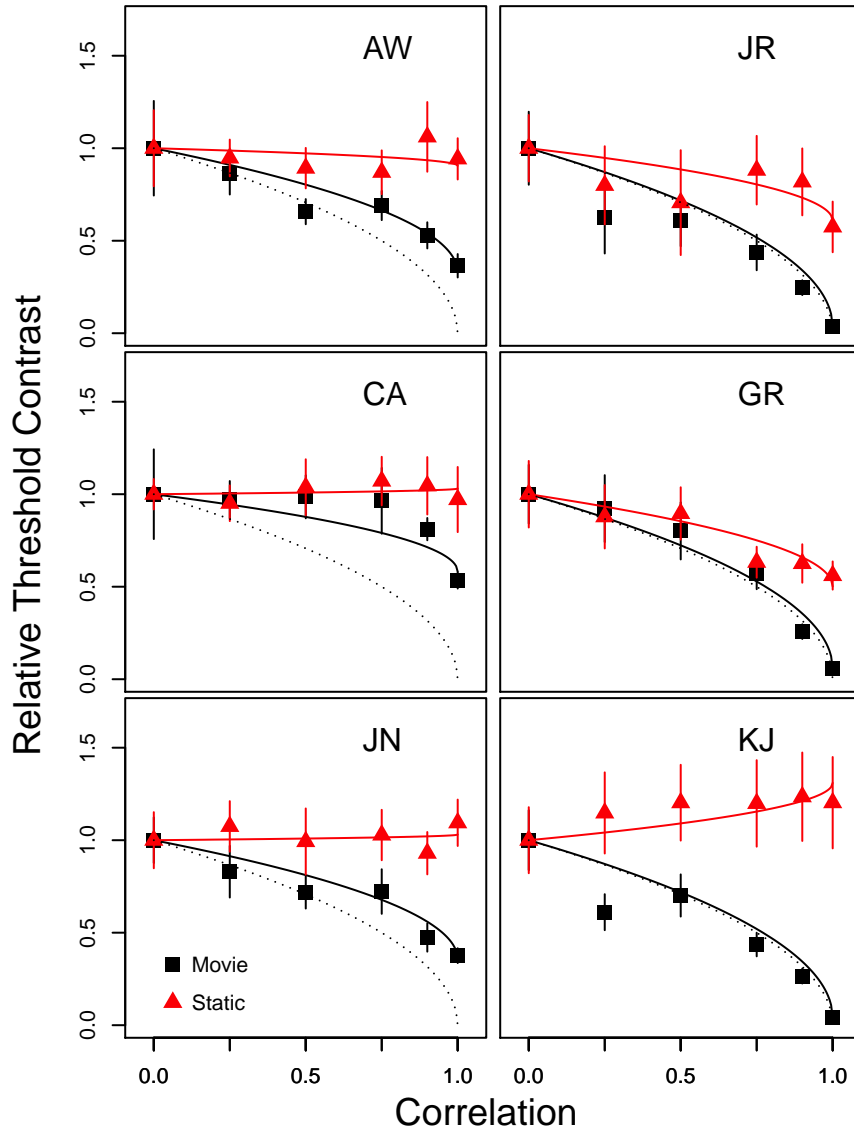
Figure 5.4: Contrast thresholds for detecting a Gaussian blob relative to that obtained when the correlation is zero plotted as a function of the correlation between the two power law noise fields. The solid curves are least squares fits of Equation 4.12. The noise fields in the image pair were presented as a static display (filled triangle) or as a movie display (filled square). Results for six observers are shown. Error bars show 95% confidence intervals. The dotted curve shows the performance for the ideal observer.

observer's correlation improvement factor of $k = 1$.

As we saw with Gaussian white noise in Experiment 2a, the movie presentation mode shows a very different picture with all six observers achieving k values far greater than their corresponding value with static displays, as seen in Table 5.2. Once again, there are a range of k values. A moderate improvement in contrast threshold as correlation increases is seen with CA, $k = 0.42$, 95% CIs [0.36, 0.48] and stronger improvements are shown by AW and JN with k values of 0.67 and 0.65, respectively, 95% CIs [0.60, 0.74], [0.60, 0.69], respectively. Three observers, JR, GR and KJ, show very high levels of improvement in contrast threshold as correlation increases, with values approaching that of the ideal observer ($k = 0.97$, 0.94 and 0.96, respectively, 95% CIs [0.96, 0.97], [0.94, 0.95], and [0.95, 0.97], respectively). The higher k values shown for movie displays over static displays demonstrate that the real observer is able to use improvements in correlation between the image pairs better when the images are displayed in a movie format than if they are displayed in a static format. Once again, this shows that the real observer using movie presentation is able to utilise the correlation between the image pair in a similar manner to the ideal observer, exhibiting declining thresholds as noise correlation increases when trying to detect a signal in paired noise fields, unlike the real observer using static presentation, whose thresholds decline, at best moderately and, at worst, not at all. The suggestion here is that symmetry, as exhibited by static displays, provides little or no help when trying to detect a signal in paired power law noise fields, whereas animation, as exhibited by movie presentation, does.

5.0.6   General Discussion

Radiologists conventionally view mammograms in a static, mirror symmetric side-by-side display, with the intuition that deviations from symmetry, as caused by an abnormality in one breast, will become salient. This is a widely used strategy; however, it may have limitations under certain conditions as discussed in section 1.7.2 on page 91 and in the introduction to Experiment 1 (section 4 on page 121). One visual property that is highly salient to the human visual system is motion and flicker and this has been exploited in this study by the movie display method.

The correlation between the two noise fields presented on each trial was varied, and contrast thresholds measured. An ideal observer who can decorrelate the two noise fields perfectly has contrast thresholds that decline in proportion to $\sqrt{1-\rho}$, where ρ is the correlation between the two noise fields. The concept of the ideal observer is a powerful tool that enables the optimum performance for a task to be specified. Its power stems as much from the ability to reveal what the human visual system cannot do as from revealing what it can do and, hence, gives an indication of a potential algorithm for a display system that may enable the human observer to perform the

task in the optimum manner.

Experiment 2a used a Gaussian white noise background and the theoretical discussion proposed that the ideal observer would decorrelate the two images prior to cross-correlating the noisy received stimulus with a signal template. This decorrelation process enables the ideal observer's performance to improve as the noise pair correlation increases. Experiment 2b was conducted with a power law noise background and the theoretical discussion proposed that the ideal observer would first pre-whiten each image before utilising the same strategy, as with Gaussian white noise, of decorrelating the image pair prior to cross-correlating each noisy image with a signal template.

The aim of this study was to examine whether real observers, when viewing images in a movie display mode, behave like the ideal observer and utilise the correlation present between the images. The results of this study show that when using the traditional static, side-by-side mirror presentation the human observers do not behave like the ideal observer; their thresholds for detecting the signal were essentially unaffected or, at best, only moderately affected by the correlation of the two noise fields. The flat performance profile across the range of correlation shows that, unaided, humans are poor at utilising the correlation between the paired noise fields. In contrast, when the human observer is aided by the movie display they do behave in a manner more like the ideal observer; their thresholds for detecting the signal reducing markedly as the correlation between the two noise fields increased. The suggestion here is that presentation of the images in a movie format provides a display that enables the human visual system to utilise the correlation present between the two images in a manner similar to an ideal observer and, thus, gives the human observer the potential to perform closer to the optimum level for the task.

This experiment has used an SKE paradigm with Gaussian white noise to enable a simplified analysis of the ideal observer and synthetic images using a power law noise background as a step towards the more realistic situation of mammography. Naive observers have also been used. The results, whilst illustrating the potential benefits of a movie presentation technique must be treated with caution and further research is necessary using real mammogram images, real tumours with unknown positions and trained readers before any practical implications can be assessed. The results do, nonetheless, suggest that movie displays do enable the human observer to utilise the correlation between paired images and give us confidence to progress the research into these more clinical scenarios.

Chapter 6

Experiment 3: Does the Rate of Alternation Affect the Observer's Contrast Threshold for the Detection of a Signal in Paired Noise Backgrounds?

6.1 Introduction

As discussed in Experiment 2, there is abundant evidence showing the sensitivity of the human visual system to motion and flicker (Adelson & Bergen, 1986; Hubel, 1995; Watson, 1986) and the salience of flicker has been demonstrated in a number of studies (Franconeri et al., 2005; Ludwig et al., 2008; Spalek et al., 2009). Flicker can easily be generated by sequentially presenting two images with minor differences in a two frame animation sequence, such that any differences between the two images will appear to flash on and off or appear to move. This is the technique used in Experiment 2, where observers attempted to detect a signal placed in one image of the pair with varying levels of correlation between the two images. The aim of Experiment 2 was to establish, using controlled laboratory stimuli, whether a two frame animation sequence could be a viable presentation mode for displaying mammogram pairs to improve the detection of a lesion or abnormality in one image.

The rate of sequential presentation of each image in Experiment 2 was chosen following a review of research across various fields to ascertain what rate of alternation would be most salient, but primarily on the basis of trial and error across a number of practice trials and what "felt" right. This study aims to provide some scientific grounding for the choice of alternation rate. Reference to Figure 1.18 on page 73 suggests that the human visual system is most sensitive to stimuli animated at a rate between approximately 5 Hz to 10 Hz, however, this is dependent upon the spatial characteristics of the signal, as shown in Figure 1.19 on page 74. The signal used in Experiment 1 and Experiment 2 was a Gaussian blob truncated at $\pm$ 3 standard deviations with a spatial SD of 0.43 deg, thus giving a signal size of 2.58 degrees and, taking the blob to be a single cycle, this equates to approximately 0.4 cycles per degree. Reference to Figure 1.19, suggests a peak sensitivity for a signal of this size of between 4Hz and 8Hz.

Research into attention has also found a range of animation or flicker rates suggested to be optimum. Spalek et al. (2009) using a visual search task found maximum sensitivity to flicker at about 10 Hz, whereas Huang et al. (2011), also conducting visual search tasks, here relating to advertising on web pages, found that search accuracy was best for a flicker rate of 0.5 Hz. Animated warnings are important in safety related fields such as motor vehicle safety and a report commissioned by the United Kingdom Department of the Environment, Transport and the Regions found that flash rates of 4 Hz facilitated the most effective detection of vehicle mounted warning signals (Cook et al., 2000). For warnings in general, Sanders & McCormick (1992, p. 151) recommended flash rates of between 3 - 10 Hz to draw attention to the signal's presence with 4 Hz considered to be the best rate. For alarms and warnings in buildings the European Standard (BS EN 54-23 - summarised at the Building Services Building Services Index (2015)) suggests the optimum flash rate for detection should be between 0.5 Hz and 2 Hz.

The ultimate goal of this thesis is to investigate the efficacy of flicker generated using animated displays as a technique that can be used in medical imaging. The particular area of interest for this thesis is mammography, where the use of paired images is common and an animated display would be feasible, however, the range of research with animation and flicker in this field is limited. Erickson et al. (2011) compared a flicker technique against a traditional side by side presentation of brain images, although the flicker rate was not specified and van Engeland et al. (2003) used a technique of toggling between mammogram images to induce an element of flicker but here the flicker rate was varied by the operator's toggling rate.

The research cited earlier into the use of animation and flicker in vision research, industry and road safety supports a wide range of flash or flicker rates, perhaps not surprisingly given the varied applications involved, however, it gives a sensible band of animation frequencies across which our

Single frame length (ms)	Cycle length - (ms)	Alternation rate (Hz)
50	100	10
83.33	166.67	6
133.33	266.67	3.75
166.67	333.33	3
250	500	2
500	1000	1

Table 6.1: Alternation rates and frame durations used in Experiments 3a and 3b.

own application of flicker can be tested. Thus, having established the viability of animated displays as a presentation mode in Experiment 2, Experiment 3a and Experiment 3b aimed to establish an empirical basis for the choice of animation rate and to determine what animation rate would be most effective for the detection of a signal in one noise field of two noise fields presented as a pair.

6.2   Theory

The theory for Experiment 3 is the same as for Experiment 1, as described in the Theory section of Experiment 1a (see section 4.2.1 on page 125).

6.3   Experiment 3a

6.3.1   Method

6.3.1.1   Apparatus

The apparatus used in Experiment 3a was the same as for Experiment 1a.

6.3.1.2   Stimuli

Presentation of the stimuli was the same as for the alternated presentation in Experiment 2a except that only one correlation between the backgrounds was used, this being 0.9, and instead of a single frame duration of 250 ms, six frame durations were used to give six alternation rates. The alternation rates were calculated as shown in Table 6.1.

6.3.1.3   Procedures

The procedures used for the alternated presentation were the same as used in Experiment 2a.

6.3.1.4   Observers

Three observers participated in experiment 3a (two male and one female). Two were inexperienced observers (JR and KJ) but received training prior to commencing the study. GR is the author and

an experienced psychophysical observer. No observer had any background in radiology or medical physics. All observers had normal or corrected-to-normal vision.

6.3.2   Results and discussion

Contrast thresholds for detecting a Gaussian blob were measured as a function of the alternation rate of the two noise fields presented to the observer. Thresholds for each correlation value were calculated from the 75% correct point of each observer's psychometric function fitted using probit regression. The response on each trial was correct or incorrect, and the probit regression used these binary values. Thus, each threshold represents a fit to at least 6 blocks of 60 trials = 360 points.

The study aimed to answer the question of what alternation rate is most effective for the detection of a signal in one noise field of two noise fields presented as a pair with Gaussian white noise backgrounds. Reference to Figure 6.1 shows how threshold contrast varied with variation of the alternation rate. The slope of the function gives an indication of whether the observer is able to make use of changes in alternation rate. The general picture emerging from Figure 6.1 is of a gradual decrease in contrast threshold as alternation rate decreases to a minimum of 2Hz for observers GR and JR and 1 Hz for KJ.

Table 6.2 aims to show whether the difference between the lowest threshold contrast and the highest threshold contrast is significant, and hence, whether change of alternation rate has a significant effect on the performance of the observer as measured by their contrast thresholds for detecting the signal. To achieve this a t-test was conducted between the alternation rate with the highest contrast threshold and the alternation rate with the lowest contrast threshold. By testing whether a significant difference exists, it can be concluded whether the change of alternation rate has a significant effect on the performance. For example, for observers GR, JR and the pooled data, the minimum measured contrast threshold is at 2Hz and the maximum measured contrast threshold is at 10Hz and, therefore, a t-test is conducted between the threshold contrast values at 2Hz and 10Hz. For observer KJ, the minimum measured contrast threshold is at 1Hz and the maximum measured contrast threshold is at 10Hz and, therefore, the t-test is conducted between the threshold contrast values at 1Hz and 10Hz.

Reference to Table 6.2 shows that JR, GR and the pooled data show a significant difference between the highest and lowest contrast thresholds indicating that alternation rate has a significant effect on observer performance ($t = 3.65$, $p < .001$; $t = 6.35$, $p < .001$; and $t = 3.50$, $p < .001$ for observers JR, GR and pooled data, respectively). KJ does not show a significant difference ($t = 1.62$, $p = .11$). With the exception of observer KJ, the optimum alternation rate appears

164
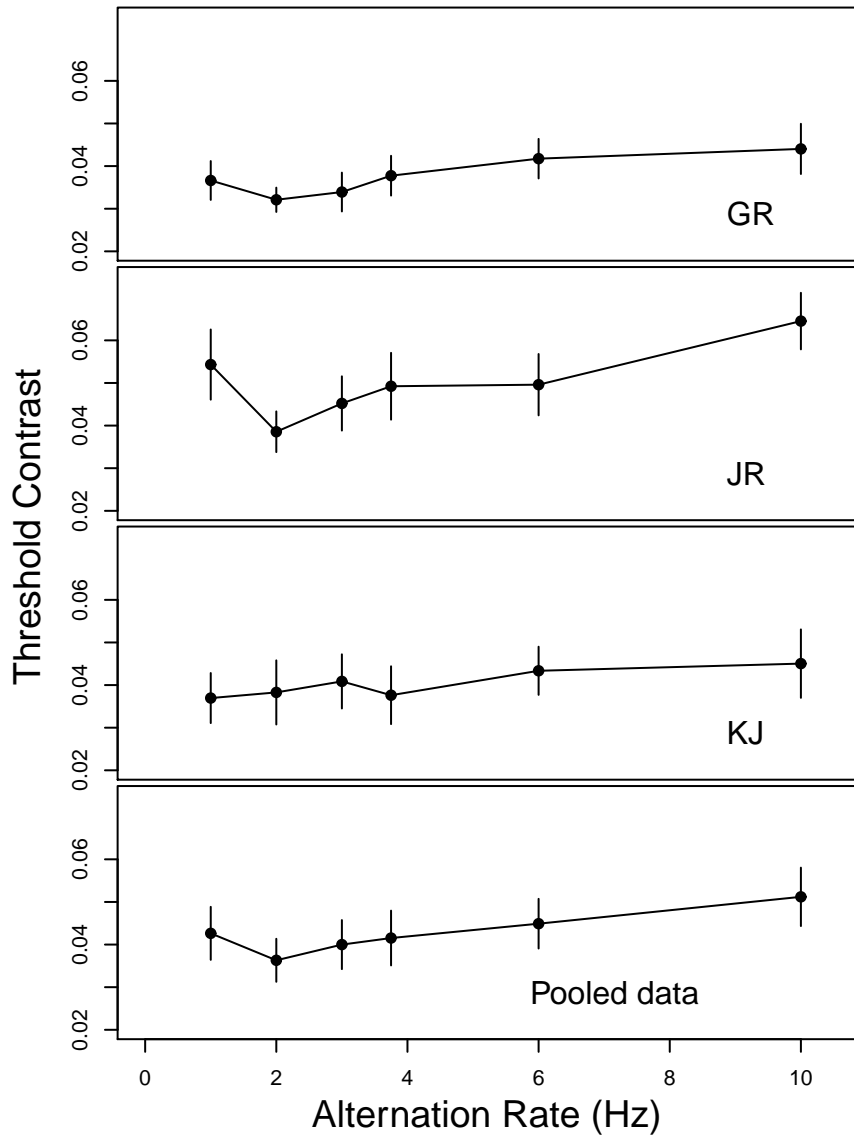
Figure 6.1: Plots of contrast threshold against alternation rate (Hz) for the animated condition with a signal in Gaussian white noise for three observers. Error bars show 95% confidence intervals.

		Gaussian White Noise						
	Range (Hz)	Lowest C_T	SE	Highest C_T	SE	t value	Significance	Effect Size (d)
GR	2-10	0.032	0.003	0.044	0.006	3.65	$<.001$**	0.29
JR	2-10	0.039	0.005	0.065	0.007	6.35	$<.001$**	0.58
KJ	1-10	0.037	0.006	0.045	0.008	1.62	0.11	0.12
Pooled data	2-10	0.037	0.005	0.051	0.007	3.50	$<.001$**	0.17

** indicates significant at .001 level $\qquad\qquad$ C_T =Contrast Threshold

Table 6.2: Table of difference between highest and lowest points for the functions at figure 6.1.

to be at approximately 2Hz. However, notwithstanding the significance of the differences, they, nonetheless, exhibit small effect sizes for GR, and the pooled data and a medium effect size for JR, suggesting that, at best, varying the alternation rate has only a small to medium effect on observer performance and, at worst, has no effect at all.

6.4 Experiment 3b

6.4.1 Introduction

As in previous experiments, to enable inferences about the usefulness of animation or flicker in mammogram reading, the observers were presented with images with a background which had a $1/f^3$ power spectrum, which has similar characteristics to those of real mammograms (Burgess et al., 2001). As in Experiment 3a, the aim of the experiment was to determine what alternation rate would be most effective for the identification of a signal in one noise field of two noise fields presented as a pair but, in this case, using images with a background that had a $1/f^3$ power spectrum.

6.4.2 Method

The apparatus, stimuli and procedures for Experiment 3b were identical to those used in Experiment 3a except that a low-pass $1/f^3$ power law noise background was used. The same three observers who participated in Experiment 3a were used in Experiment 3b.

6.4.3 Results and discussion

Contrast thresholds for detecting a Gaussian blob were measured as described in section 6.3.2. Once again, each threshold represents a fit to at least 6 blocks of 60 trials = 360 points. The study aimed to answer the question of what alternation rate is most effective for the identification of a signal in one noise field of two noise fields presented as a pair but, in this case, using images with a background that had a $1/f^3$ power spectrum.
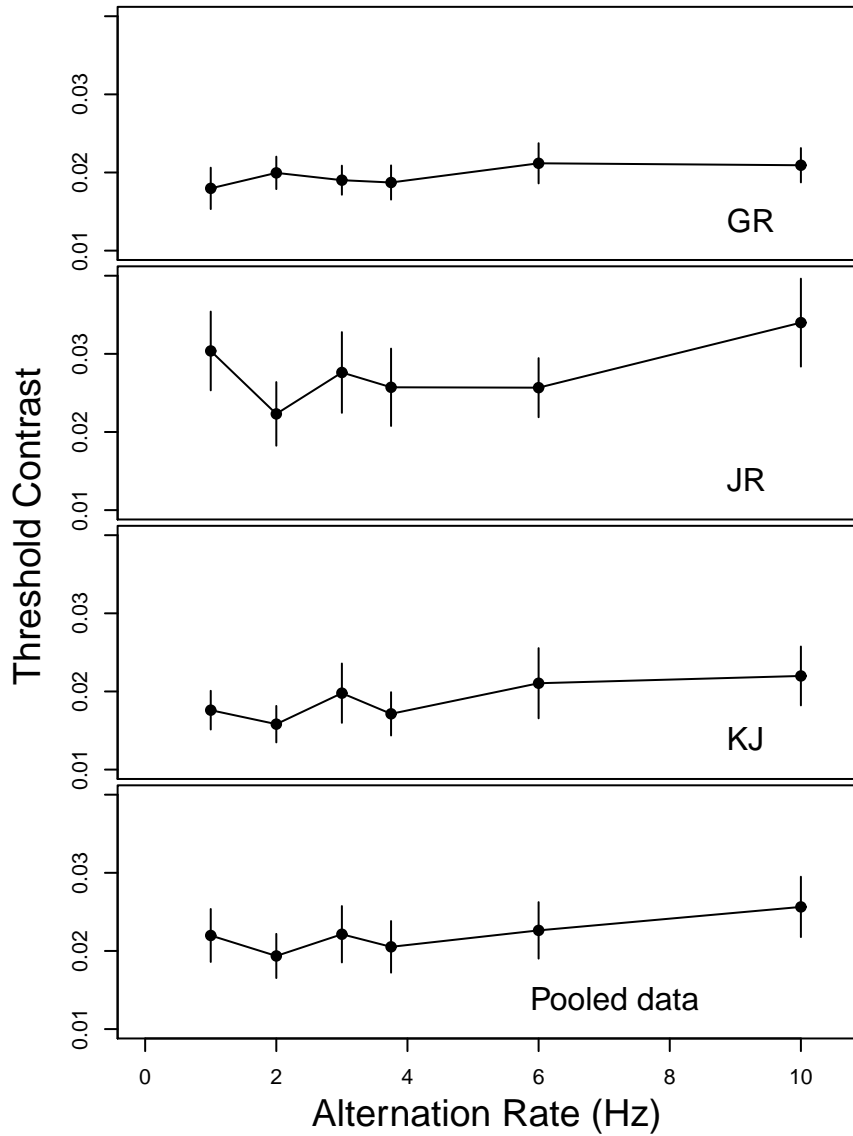
Figure 6.2: Plots of contrast threshold against alternation rate (Hz) for the animated condition with a signal in power law noise for three observers. Error bars show 95% confidence intervals.

Reference to Figure 6.2 shows how threshold contrast varied with variation of the alternation rate. The slope of the function gives an indication of whether the observer is able to make use of changes in alternation rate. The general picture emerging from Figure 6.2 is of a very gradual decrease in contrast threshold as alternation rate decreases to a minimum of 2Hz for observers JR and KJ and 1 Hz for GR.

Table 6.3 aims to show whether the difference between the lowest threshold contrast and the highest threshold contrast is significant, and hence, whether the change of alternation rate has a significant effect on the performance of the observer as measured by their contrast thresholds for detecting the signal. To achieve this a t-test was conducted between between the alternation rate with the highest contrast threshold and the alternation rate with the lowest contrast threshold. By

	Power Law Noise							
	Range (Hz)	Lowest C_T	SE	Highest C_T	SE	t value	Significance	Effect Size (d)
GR	1-6	0.018	0.003	0.021	0.002	1.74	0.08	0.17
JR	2-10	0.022	0.004	0.034	0.006	3.37	<.001**	0.31
KJ	2-10	0.016	0.002	0.022	0.004	2.79	0.006*	0.25
Pooled data	2-10	0.022	0.003	0.026	0.004	2.62	0.009*	0.22

* indicates significant at .05 level; ** indicates significant at .001 level

C_T = Contrast Threshold

Table 6.3: Table of differences between highest and lowest points for the functions at figure 6.2.

testing whether a significant difference exists, it can be concluded whether the change of alternation rate has a significant effect on the performance.

For observer GR, the minimum measured contrast threshold is at 1Hz and the maximum at 6Hz, therefore a t-test is conducted between the threshold contrast values at 1Hz and 6Hz. For observers JR, KJ and the pooled data the minimum measured contrast threshold is at 2Hz and the maximum at 10Hz, therefore a t-test is conducted between the threshold contrast values at 2Hz and 10Hz.

Reference to Table 6.3 shows that JR, KJ and the pooled data show a significant difference between the highest and lowest contrast thresholds indicating that alternation rate has a significant effect on observer performance ($t = 3.37$, $p < .001$; $t = 2.79$, $p = .006$; $and\ t = 2.62$, $p = .009$ for observers JR, KJ and pooled data, respectively). GR does not show a significant difference ($t = 1.74$, $p = .08$). With the exception of observer GR, the optimum alternation rate appears to be at approximately 2Hz, however, notwithstanding the significance of the differences, they, nonetheless, exhibit small to medium effect sizes, suggesting that, at best, varying the alternation rate has only a small to medium effect on observer performance and, at worst, has no effect at all.

6.5   General Discussion

The aim of this study was to investigate what the optimum alternation rate was for detecting a Gaussian blob in paired noise fields presented in an alternating display format. The results show that with both Gaussian white noise and power law noise backgrounds, varying the alternation rate has a significant but generally small to medium effect with an optimum alternation rate at approximately 2Hz. However, in both backgrounds there was a single observer that did not appear to benefit from the variation in alternation rate.

The small to medium effect sizes may be considered to be a reflection of the broad range of frequencies proposed as being optimum as cited in the introduction. The results also validate our own choice of alternation rate of 2 Hz as a suitable display rate, albeit, if only as a confirmation

that it is not significantly worse than any of the other alternation rates between 1 Hz and 10 Hz. On most observer plot functions, there did appear to be a directional change from a decreasing trend of contrast threshold to an increasing trend at 2 Hz, although this step change was only significant in one observer (JR), and not present in all of the plots. However, it does lend some support to the use of 2 Hz.

As previously discussed, the theoretical background would, perhaps, point to an optimum alternation rate of between 4Hz and 8Hz (see Figure 1.19), and the applied research would suggest a broad span of optimum alternation rates from 0.5Hz to 10Hz (Cook et al., 2000; Huang et al., 2011; Sanders & McCormick, 1992; Spalek et al., 2009). It should be noted, however, that much of the theoretical research and applied studies are, in general, referring to the detection of flicker, rather than the detection of a flickering signal in an alternating display. This may be a subtle distinction but is, nevertheless, an important one. A radiologist is not aiming to detect whether the display is flickering, rather, they would be using the alternation of the two images of a pair to identify differences between the images. Having identified the differences, the radiologist then needs to differentiate between differences that represent normal variations between the images and differences that represent an abnormal variation, such as the appearance or growth of a tumour. Whilst this experiment has shown only small to moderate differences in performance for the detection of a Gaussian blob across a range of frequencies from 1-10Hz, the use of a Gaussian blob and synthetic images may increase the difficulty of discriminating the signal from non-signal differences and, thus, limit the performance of the observer. The use of real abnormalities and trained radiologists may increase the ability to differentiate between non-signal (normal differences) and signal (abnormal differences) resulting in improved performance and, perhaps, indicate a clear optimum alternation frequency. Further research using trained radiologists with real mammograms and real tumours is necessary to determine this.

Notwithstanding the possibility that real tumours could elicit a discrete optimum alternation rate, the results from this experiment, with a small to medium effect seen in both Gaussian white noise and power law noise, when varying the alternation rate from 1 Hz to 10 Hz, indicates that using 2 Hz as the alternation rate may offer marginal improvements in detection performance. Despite this, the small to medium effect sizes suggest that the use of any other alternation rate, within the range of 1-10Hz, would not constitute any great advantage or disadvantage for signal detection over other frequencies within that range. It would be, therefore, worth considering other advantages and disadvantages of each alternation rate, such as observer comfort, when selecting an alternation rate, or, indeed, providing a variable alternation rate, between 1 Hz and 10 Hz, such that the observer can adjust the rate to their own personal preference for comfort and effectiveness.

Chapter 7

Experiment 4: Comparison of Animated Presentation against Traditional Mirror Symmetric Presentation for the Detection of a Synthetic Tumour in Real Mammogram Backgrounds

7.1   Introduction

Experiment 1, using Gaussian white noise and power law noise backgrounds, has shown that improvements in symmetry between an image pair presented side-by-side appear to confer little or no benefit to an observer who is trying to detect the presence of a signal in one of the images. This may be an important finding for breast radiology, where it is common practice to hang image pairs in a mirror symmetric configuration with the rationale that violation of the symmetric pattern will assist with tumour detection (Andolina & Lillé, 2010; Bun, 2002; Harvey & March, 2013; Kopans, 2007; Sickles, 2007). Intuitively, it would be expected that such violations would be readily detected and that as the level of symmetry improved this process of detection would

become easier and performance would improve. The result of experiments using Gaussian white noise and power law noise images show that improvements in the level of symmetry, as measured by the inter-image correlation, are not matched by improvements in observer performance.

In contrast to this, Experiment 2 tested an alternative display protocol: an animated display format, where the two images from an image pair were displayed sequentially in the same location in a continuous movie loop. In this format, increases in correlation did lead to an improvement in observer performance suggesting that the animated presentation enabled real observers to decorrelate the image pair and perform with a strategy similar to that of the ideal observer. This finding was established using noise backgrounds that initially consisted of Gaussian white noise, to establish a theoretical basis for the experiment, followed by a background consisting of power law noise. As discussed in section 1.6.4, power law noise was chosen for its statistical similarity to real mammograms (Burgess et al., 2001) and the results of Experiment 2 intimated at the potential of an animated movie presentation technique as an aid to mammogram readers and radiologists for the detection of potential tumours in paired mammograms.

Whilst power law noise backgrounds are broadly considered to be good models for real mammograms (Burgess et al., 2001; Reiser et al., 2013), and, therefore, suitable for theoretical research, a number of studies, nevertheless, suggest that real mammograms may not behave entirely as predicted by power law noise models (Bochud et al., 1999, 2004; Burgess et al., 2001) (see sections 1.6.4 and 1.6.5 for a fuller discussion of this). Thus, having established a theoretical basis for an animated display protocol using Gaussian white noise and power law noise, to validate its usefulness in a clinical scenario, it is necessary that the protocol be tested using the stimuli encountered in the clinical scenario. This study, therefore, took the next step towards the end goal of a full clinical trial by using real mammogram images but with a synthetic tumour and naive (non-radiologists) participants.

Thus, the aims of this study were to determine whether:

1. An animated movie presentation protocol would enable the real observer to utilise the correlation present between the two images of real mammogram pairs to improve the observer's performance for the detection of a synthetic tumour in one image of the pair.

2. An animated movie presentation protocol would confer an advantage over static presentation, as measured by lower contrast thresholds, for the detection of a synthetic tumour in paired mammograms.

7.2 Theory

The theoretical basis remains the same as detailed in sections 4.2.1, which shows that, for an ideal observer, performance will improve as the correlation between the two images of a pair increases. For an image pair with power law noise backgrounds, the ideal observer will first pre-whiten each noise field (Burgess, 2010, pp. 26-46; Bochud, 2013, pp. 153-164) prior to decorrelating the image pair and cross-correlating with the signal template. For real mammogram backgrounds the theory is the same as for power law noise backgrounds. Once again, the ideal observer will pre-whiten each mammogram image prior to decorrelating the mammogram pair and then cross-correlating each image with the signal template.

7.3 Method

7.3.1 Apparatus

The apparatus used in Experiment 4 was the same as for Experiment 1a with the exception that, when viewing from a distance of 52 cm from the monitor screen, the dimensions of each image display were 21 deg wide and 32.6 deg high.

7.3.2 Stimuli

The stimuli were real mammogram images with a superimposed Gaussian blob signal. The mammogram images were supplied by Plymouth National Health Service Trust Derriford Hospital's Primrose Unit and prepared for use in the experiment as described below.

7.3.2.1 Image preparation

Images were supplied from an image bank held by Derriford Hospital's Primrose Unit and all images were anonymised by Primrose Unit staff prior to release. All image preparation was carried out using Irfan View for Windows 7, version 4.36, except where otherwise stated. One hundred case folders were supplied, each case containing eight images, four each from two separate screenings, referred to as prior and post screenings. Each screening contained the left mediolateral oblique (MLO) image, the right MLO image, the left cranial-caudal (CC) image and the right CC image.

The images were relabeled to indicate the case number, the image view and the screening occasion, e.g. 0001LMLO Prior was case 0001, left mediolateral oblique view and the prior screening. 0001RCC Post was case 0001, right cranial-caudal view and the post screening. When supplied, the majority of images were 2082 x 2800 pixels with the remainder 2800 x 3518 pixels. The latter

were cropped by removing extraneous black surround to give an X:Y axes aspect ratio of 0.74 (to approximately match the X:Y ratio of the 2082 x 2800 pixel images), and then resized to 2082 x 2800 pixels.

Three cases were removed as unsuitable owing to the inclusion of large artefacts such as breast implants or pacemakers and two more removed where the prior and post were clearly from different women. A further resizing was carried out to reduce the amount of black border and maximise the actual breast image area in the display. As would be expected from a random sample of women, the range of breast size varied over a wide range and, therefore, a compromise size of 1800 x 2800 pixels was chosen, giving good breast image display size whilst losing a small number of images that had an actual breast image X dimension greater than 1800 pixels.

A clean-up of the images was then carried out using Irfan View image editing tools to remove all extraneous information, such as measurement scales and image view markings, and any non-breast tissue, such as torso below the breast. The images were also converted to greyscale. As a result of the prior and post image sets having been taken, typically, three years apart using different scanners, there was usually a difference in mean grey level. To maximise the effectiveness of the animated method of displaying the images and prevent the whole image from flashing the image grey levels were equalised using a bespoke normalisation programme written using the 'R' environment software. Normalisation, in this context, refers to the process of matching the mean and variance of each image. In the 'R' environment image pixel levels are represented by values from zero to one, with zero being pure black and one being pure white, with shades of grey in between. Normalisation was accomplished by firstly standardising these original image pixel values of the breast image (not including the black background) and then dividing these standardised values by a common range denominator (to set the required standard deviation) plus the required mean, truncating the resulting values at zero and one. Applying this process to each image ensured that the images had equal mean grey levels and an equal variance.

Prior and post image pairs were registered using the BUnwarpJ plugin available as part of the Fiji image processing package (Abramoff et al., 2004). BUnwarpJ is a 2D Image registration method based on elastic deformations represented by B-splines. Identifiable landmarks, such as the nipple and skin boundary, were utilised to improve the registration wherever possible. The elastic deformation used to align the two images occasionally resulted in the appearance of black segments on the image, corresponding to an unscanned area of that breast. These black segments were removed by cropping the image pair at the pixel value corresponding to the limit of black segment. The X:Y axes ratio was maintained throughout and the image resized back to 1800 x 2800 pixels on completion. As a measure of the level of correspondence between the two images

following registration, the correlation between the prior and post breast images was measured using a bespoke programme written using the 'R' environment software. It is important to note that the correlation measure was between non-zero grey levels in the two actual breast images and did not include the black surround (which would falsely inflate the correlation level). The post-registration correlation values are recorded in Table 7.1.

The final stage of image preparation was to convert the images to bitmap format and resize them to 494 x 768 pixels for use in the experiment programme. A bank of 183 CC image pairs were used in the experiment with the same image bank used for both static and animated movie display modes.

The images were displayed using 256 grey levels and in all of the experiments, to allow for the circumstance that the blob may be randomly placed in a location that is close to 255 (i.e. pure white), thus inhibiting an increase in grey levels resulting from the addition of blob, the actual mean grey level was shifted down by a factor of 0.1 to create "headroom" for the blob. Thus the mean grey level was 102 rather than 127.5. The mean grey level was used to calculate the grey level for the blob as shown in equation 7.1.

$$Blob\ grey\ level = mean\ grey\ level \times blob\ contrast \tag{7.1}$$

The mean grey level in the vicinity of the blob (Imgrey) was measured by taking the mean grey level of the pixels surrounding the blob and this was used to calculate the local contrast of the blob as shown in equation 7.2.

$$Local\ contrast = \frac{Blob\ grey\ level - Imgrey}{Imgrey} \tag{7.2}$$

The local contrast was then used to determine contrast thresholds for each correlation bin (correlation bins are discussed in the results section on page 180) which were calculated from the 50% correct point of the pooled observers' psychometric function of proportion of yes responses versus the local contrast of the blob, fitted using probit regression.

7.3.2.2   The signal

The signal was a Gaussian blob truncated at $\pm$ 3.5 standard deviations with a spatial SD of 2.34mm (0.26 deg). The signal was added to one image of the pair in a random sequence of image plus signal and image only and was randomly placed at any position within the breast region of the image (signal known statistically (SKS)). The contrast of the signal, when present, had two levels 0.1 and 0.2 and the contrast level was randomly selected.

Image Pair Number	Correlation Between Images	Image Pair Number	Correlation Between Images	Image Pair Number	Correlation Between Images	Image Pair Number	Correlation Between Images
1	0.84	47	0.84	93	0.73	139	0.75
2	0.79	48	0.84	94	0.78	140	0.74
3	0.81	49	0.84	95	0.67	141	0.93
4	0.85	50	0.83	96	0.71	142	0.92
5	0.87	51	0.85	97	0.77	143	0.81
6	0.78	52	0.80	98	0.69	144	0.79
7	0.85	53	0.79	99	0.82	145	0.89
8	0.89	54	0.84	100	0.97	146	0.90
9	0.89	55	0.87	101	0.84	147	0.84
10	0.83	56	0.91	102	0.79	148	0.84
11	0.81	57	0.85	103	0.81	149	0.84
12	0.84	58	0.82	104	0.85	150	0.83
13	0.91	59	0.74	105	0.87	151	0.85
14	0.86	60	0.78	106	0.78	152	0.80
15	0.89	61	0.81	107	0.85	153	0.79
16	0.90	62	0.89	108	0.89	154	0.84
17	0.74	63	0.79	109	0.89	155	0.87
18	0.80	64	0.81	110	0.83	156	0.91
19	0.85	65	0.92	111	0.81	157	0.85
20	0.90	66	0.89	112	0.84	158	0.82
21	0.85	67	0.88	113	0.91	159	0.74
22	0.85	68	0.89	114	0.86	160	0.78
23	0.76	69	0.87	115	0.89	161	0.81
24	0.82	70	0.84	116	0.90	162	0.89
25	0.84	71	0.81	117	0.74	163	0.79
26	0.91	72	0.91	118	0.80	164	0.81
27	0.88	73	0.84	119	0.85	165	0.92
28	0.88	74	0.82	120	0.90	166	0.89
29	0.81	75	0.65	121	0.85	167	0.88
30	0.84	76	0.72	122	0.85	168	0.89
31	0.81	77	0.85	123	0.76	169	0.87
32	0.82	78	0.78	124	0.82	170	0.84
33	0.76	79	0.80	125	0.84	171	0.81
34	0.82	80	0.80	126	0.91	172	0.91
35	0.86	81	0.87	127	0.88	173	0.84
36	0.86	82	0.84	128	0.88	174	0.82
37	0.85	83	0.82	129	0.81	175	0.65
38	0.85	84	0.82	130	0.84	176	0.72
39	0.75	85	0.87	131	0.81	177	0.85
40	0.74	86	0.85	132	0.82	178	0.78
41	0.93	87	0.84	133	0.76	179	0.80
42	0.92	88	0.84	134	0.82	180	0.80
43	0.81	89	0.85	135	0.86	181	0.87
44	0.79	90	0.89	136	0.86	182	0.84
45	0.89	91	0.84	137	0.85	183	0.82
46	0.90	92	0.90	138	0.85		

Table 7.1: Post-registration correlation levels for the 183 image pairs used in Experiment 4.
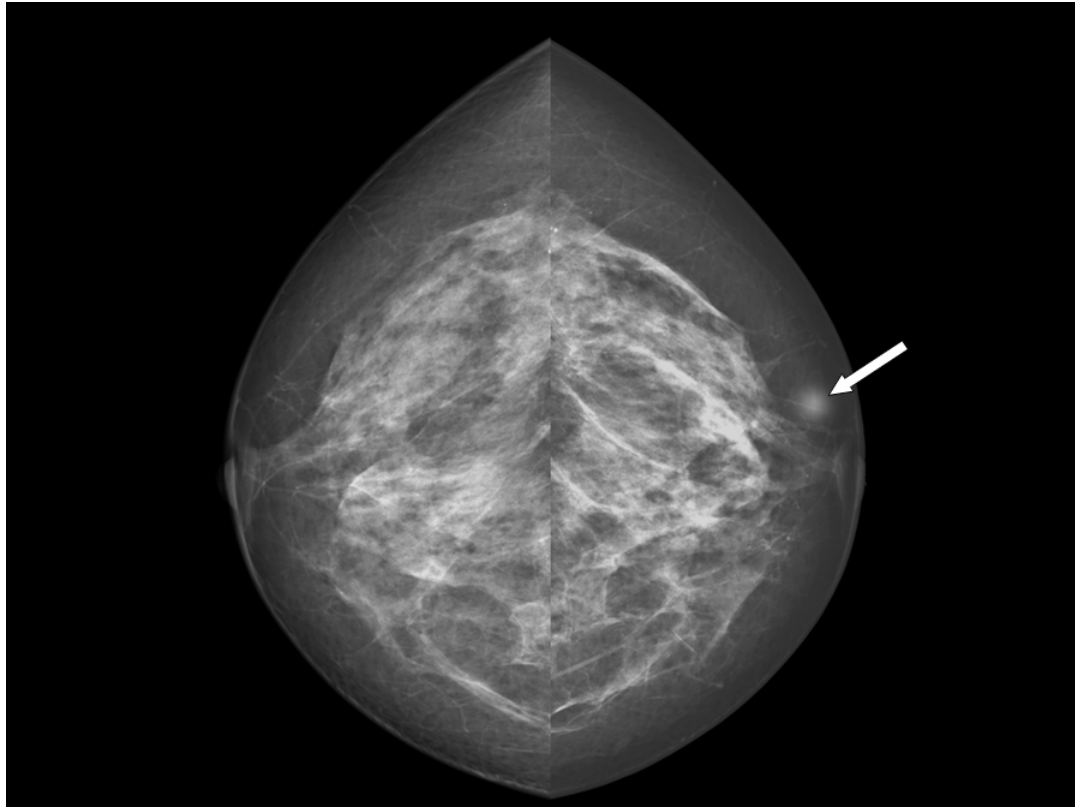
Figure 7.1: Example static display showing the prior and post CC mammogram views displayed in a mirror symmetric format. A Gaussian blob signal is shown in the right hand image (arrowed).

7.3.3 Procedures

The images were displayed as prior and post pairs in either a static or movie display mode. For the static display the image pair was displayed side-by-side in a mirror symmetric format and, if present, the Gaussian blob signal was added to the right hand image (corresponding to the post or later image) of the pair. The two images were separated by 0.22 deg. For the movie display the images were presented as same orientation images in the same location, displayed sequentially with a frame duration of 250 ms. and the Gaussian blob signal, if present, was added to the second image (corresponding to the post or later image) of the pair. For both display modes each image was displayed in a patch of 494 x 768 pixels subtending 21 deg on the X axis and 32.6 deg on the Y axis. The image was surrounded by a black region.

The experiment used a rating scale paradigm and on each trial the observer was presented with the stimuli, as shown in Figure 7.1 (static) and Figure 7.2 (movie). Each presentation mode (static or movie) was presented in one 60 trial block and within one session the presentation modes were run in random order. At the start of each session, the observer was presented with an image of the target signal as a reminder or to familiarise the observer with the signal and they were briefed on the response scale.

Frame duration = 250ms
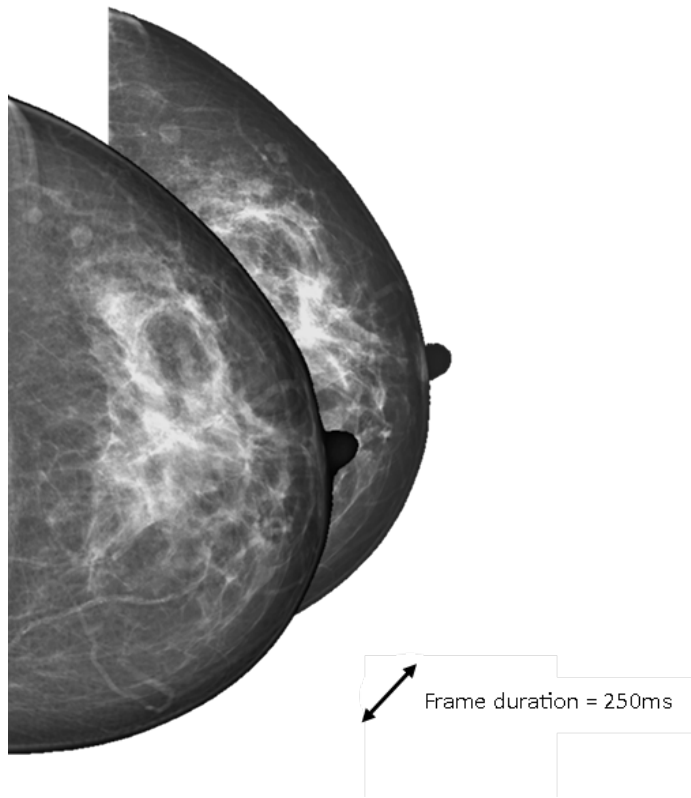
Figure 7.2: Example two frame animated movie sequence showing prior and post displays displayed sequentially in the same location.

The response scale was rated from 1-4 indicating the observer's confidence that the signal was present or not and a key to the scale was shown on the keypad. A response of 1 indicated a low confidence that the lesion was present and 4 indicated a high confidence that the lesion was present with responses of 2 and 3 representing intermediate levels of confidence. The observer was given a number of practice runs to become familiar with the procedure and the signal characteristics to facilitate a SKS protocol. On each trial, the observer was presented with the display, and the observer chose the response scale indicating their confidence that the signal was present or not. The display remained on the screen until a response was given. Feedback for wrong responses was indicated by a pulsed central red square. Each observer conducted a minimum of four sessions, where each session included blocks of trials using both display modes in random order and took approximately 30 minutes to complete.

7.3.4   Observers

Thirty two observers participated in the experiment, however, only the data from 24 observers was used as data from eight observers was discarded. This was because these observers had only used the extreme values in the response rating scale. All except one were inexperienced observers but received training prior to commencing the study. GR was the author and an experienced psychophysical observer. No observer had any background in radiology or medical physics. All observers had normal or corrected-to-normal vision.

7.4   Results and Discussion

This study had two aims:

1. To determine whether animated movie presentation would enable the real observer to decorrelate real mammogram pairs and, therefore, use the correlation present between the images.

2. To determine whether animated movie presentation would confer an advantage over mirror symmetric presentation, as measured by lower contrast thresholds, for the detection of a synthetic tumour in paired mammograms.

I will now examine the evidence for these in turn:

			Static	Movie
Bin number	Mean correlation	Correlation range	Number of trials	Number of trials
1	0.722	0.647-0.762	825	819
2	0.784	0.763-0.794	827	835
3	0.804	0.796-0.810	1023	990
4	0.819	0.813-0.823	650	748
5	0.837	0.830-0.841	742	772
6	0.844	0.8411-0.849	823	919
7	0.855	0.850-0.862	830	807
8	0.875	0.863-0.887	840	833
9	0.895	0.889-0.903	800	791
10	0.918	0.904-0.972	800	766

Table 7.2: Correlation bins showing mean correlation, correlation range and number of trials within each bin.

7.4.1 Does animated presentation enable the real observer to decorrelate real mammogram pairs and, therefore, use the correlation present between the images?

For the first aim of the study contrast thresholds were measured as a function of the correlation between the two mammograms of a pair presented to the observer. The correlation was a measure of the level of correspondence between the two mammograms of the pair, with higher correspondence having higher correlation levels. As each mammogram pair had a correlation unique to that pair, the correlation levels were partitioned into 10 bins with approximately equal numbers of trials in each bin. The mean correlation, correlation range and number of trials in each bin is shown in Table 7.2.

The responses were segregated into yes (ratings 3 and 4) or no (ratings 1 and 2) and the probit regression used these binary values. Each threshold represents a fit to at least 742 trials, with the number of trials per bin shown in Table 7.2.

If the observers are able to use the correlation between the mammograms, as used by the ideal observer, then we would expect, as shown in section 4.2.1, to see the contrast threshold decline as correlation increases in proportion to $\sqrt{1-\rho}$ where ρ is the correlation between the two mammograms of the pair. Figure 7.3 shows this not to be the case with almost flat functions for both static and movie presentation modes.

For each presentation mode, we can determine the rate of change of contrast threshold as correlation varies and hence, measure how effectively each mode enables the real observer to use the correlation between the two noise fields, by fitting the model:

$$Contrast\ Threshold = constant + b\sqrt{1-\rho} \qquad (7.3)$$
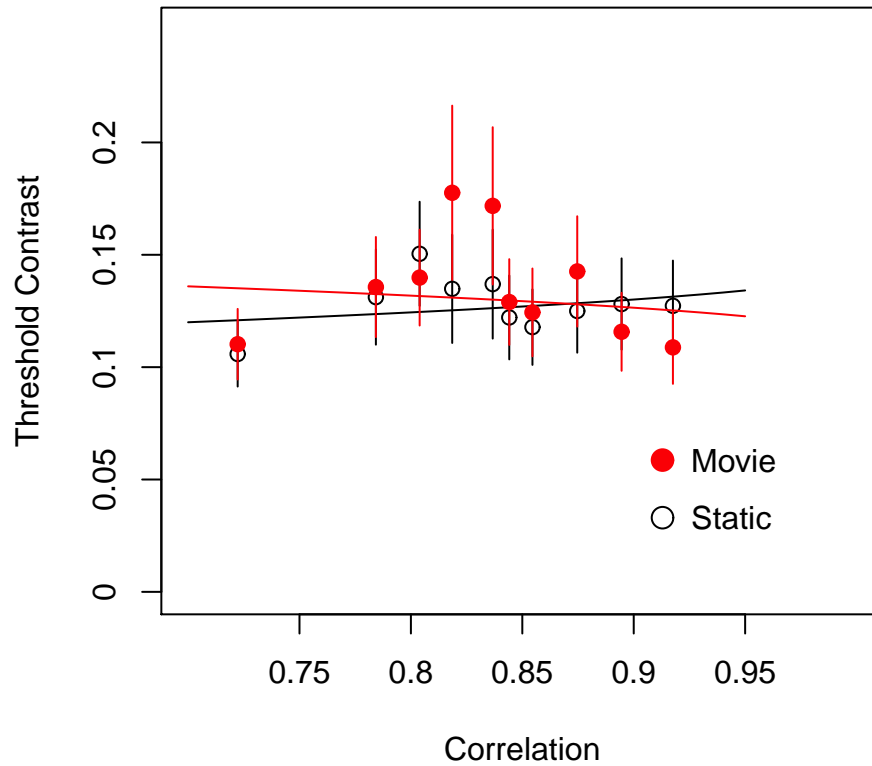
Figure 7.3: Contrast threshold as a function of correlation for the static(open circles) and movie (filled circles) conditions for the detection of a signal in mammogram pairs using the pooled data from 24 participants and 10 correlation bins. Error bars show 95% confidence intervals.

The coefficient b measures the degree to which an observer's performance is improved by the correlation between the two noise fields and fitting the model shown in Equation 7.3 using the data from the static and movie presentation modes confirms the conclusion drawn from Figure 7.3. Both functions are flat and show no significant change across the range of correlations presented (static: $b = -0.04$, $p = 0.47$ and movie: $b = 0.04$, $p = 0.68$). The results, therefore, show that neither static or movie presentation enables the observer to make use of the correlation between the image pair, however, note that a restricted range of correlations was used (0.7 - 0.9) and the results may differ with a larger range.

		Static			Movie			t-test	
Bin	Mn corr	LCL	C_t	UCL	LCL	C_t	UCL	t	p
1	0.722	0.096	0.108	0.125	0.098	0.112	0.130	0.34	0.73
2	0.784	0.115	0.133	0.158	0.117	0.135	0.161	0.13	0.90
3	0.804	0.132	0.152	0.179	0.122	0.140	0.164	0.73	0.47
4	0.819	0.115	0.135	0.163	0.149	0.181	0.229	1.93	0.06
5	0.837	0.118	0.138	0.167	0.144	0.172	0.214	1.56	0.12
6	0.844	0.106	0.122	0.144	0.113	0.129	0.151	0.48	0.63
7	0.855	0.104	0.118	0.137	0.108	0.125	0.148	0.53	0.60
8	0.875	0.111	0.127	0.148	0.123	0.144	0.173	1.06	0.29
9	0.895	0.110	0.127	0.151	0.101	0.116	0.136	0.85	0.40
10	0.918	0.110	0.127	0.150	0.095	0.109	0.128	1.39	0.17

Table 7.3: Contrast thresholds (C_t) with 95% confidence limits for each correlation bin for the detection of a blob in mirror symmetric static and animated movie displays. The t-test columns show the t statistic and p value for the difference between contrast thresholds for static and movie displays for each bin.

7.4.2 Does animated presentation confer an advantage for over mirror symmetric presentation, as measured by lower contrast thresholds, for the detection of a synthetic tumour in paired mammograms?

Reference to Figure 7.3 and the results of section 7.4.1 clearly show that animated movie presentation does not confer an advantage, in terms of lower threshold contrast levels, over mirror symmetric static presentation for the detection of a synthetic blob in paired mammogram displays. Table 7.3 confirms this, showing the contrast thresholds with 95% confidence limits for each correlation bin for the detection of a blob in static and movie displays and the level of significance for the difference between the contrast thresholds for static and movie displays for each bin. Reference to Table 7.3 show that there is no significant difference between the contrast thresholds for static and movie displays across the full range of mean correlations.

Further analysis was conducted using the receiver operating characteristic (ROC) methodology. ROCs are a widely used method for comparing the performance of two imaging modalities and will be used here to compare the performance of the static displays and the movie displays. Each image was rated for the observer's confidence of the lesion being present with the response scale rated from 1-4. A response of 1 indicated a low confidence that the lesion was present and 4 indicated a high confidence that the lesion was present with responses of 2 and 3 representing intermediate levels of confidence.

The study design uses multiple readers viewing multiple cases, usually referred to as a multiple-reader multiple-case (MRMC) study design, and this has been suggested to be the best practice methodology for assessing competing viewing modes (Wagner et al., 2002). The ROC curve is the plot of the true positive rate (the proportion of correctly classified positive observations or sensi-
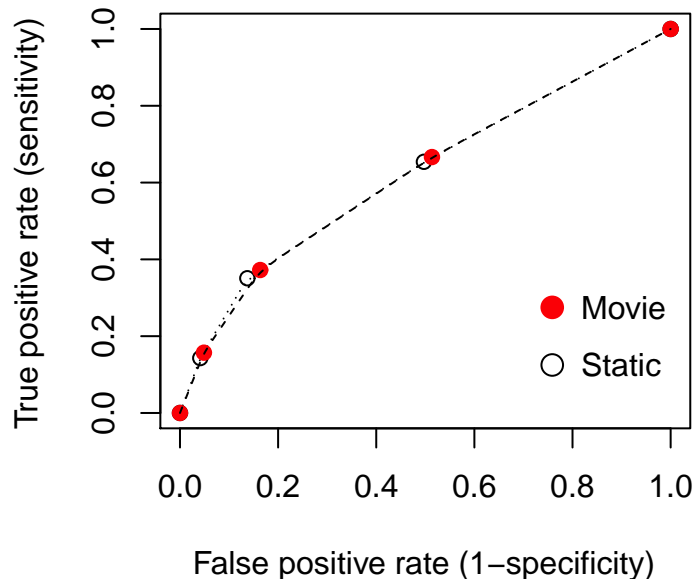
Figure 7.4: Comparison of detection performance for the static display (open circles) and the movie display (filled circles). Note that both curves are essentially the same and superimposed upon each other.

tivity) versus the false positive rate (the proportion of incorrectly classified negative observations or one minus the specificity) plotted for each confidence level. The area under the ROC curve (AUC) is used as the measure of performance and an AUC $=1$ indicates perfect performance and an AUC $= 0.5$ indicates performance at chance level.

Figure 7.4 shows the detection performance for the static display and for the movie display. The AUC, calculated in "R" using the "flux" package and "auc" command, for the static display was 0.62 (standard error $= 0.0037$, 95% CI [0.62, 0.63]) and for the movie display 0.62 (standard error $= 0.0034$, 95% CI [0.62, 0.63]) indicating no significant difference between the observers' performance using static and movie displays for the detection of a Gaussian blob in a real mammogram background.

7.5 General Discussion

This study aimed to investigate whether the real observers using animated movie presentation were able to decorrelate real mammogram pairs and, therefore, use the correlation present between the images and, whether this would confer an advantage over mirror symmetric static presentation, as measured by lower contrast thresholds, for the detection of a synthetic tumour in paired mam-

mograms. Experiments 2a and 2b, using either Gaussian white noise backgrounds or synthetic mammograms with power law noise backgrounds, demonstrated that observers using animated movie displays were able to make use of the correlation between image pairs to facilitate lower detection thresholds for a blob in one image of the pair. Observers using mirror symmetric static displays were shown to be unable to benefit in the same way to improvements in correlation between the two images of the pair. The results of the current experiment clearly show that, with real mammogram backgrounds, the observers using animated movie displays are unable to use the correlation between the images and, consequently, movie presentation does not confer any advantage over static presentation for the detection of a synthetic tumour in paired mammograms.

This raises the question of why animation was effective in Experiments 2a and 2b but not in the current experiment. The most obvious difference between the experiments is the use of real mammogram backgrounds and it is possible that the loss of effect for animation is as a result of the different properties of mammograms when compared to Gaussian white noise and power law noise. A reduction in efficiency for the detection of a signal in mammogram backgrounds in comparison to power law noise backgrounds was expected due to the impact on search effectiveness as well as increases in both false positives and negatives resulting from the anatomic structure of real mammograms (as discussed in section 1.6.5). However, it was not anticipated that this would have the qualitative effect of negating the observer's ability to use the correlation between the two images as well as a quantitative effect on their detection efficiency.

Notwithstanding the differences highlighted between synthetic and real mammograms, there were also a number of differences in the experimental set up between this experiment and previous experiments and it is possible that the effect of animation was lost as a result of experimental changes rather than because of the properties of real mammograms. The study protocol used in previous experiments was a 2AFC task and the current experiment used a yes/no rating scale task. Previous experiments also used 256 x 256 pixel image squares solely consisting of the background noise with no boundaries or discontinuities. The real mammogram image size was determined by the shape of the breast and the limitations of the laboratory display resulting in an image size of 494 x 768 pixels. The real image contained a whole breast, of varying dimensions, with the remainder of the image being black background, thereby introducing boundaries into the image. The use of square regions is not unusual in mammography research (examples are: Burgess et al. (2001); Burgess & Judy (2007); Myers et al. (1985); Reiser et al. (2013)), however these images usually represent one tissue type (or synthetic representation of one tissue type) within each image and across a set of images used in an experiment. A whole breast image, however, can contain a range of tissue types including fatty tissue and fibroglandular tissue as well as other anatomic

features such as the pectoral muscle in MLO views and the make up of the breast structure will vary from image to image.

Another difference between the experiments is the size of the signal used. A Gaussian blob has been used as the signal throughout all of the experiments, however, in previous experiments the blob was truncated at ± 3 standard deviations with a spatial standard deviation of 0.43 degrees. This produced a signal subtending 2.57 degrees and measuring 23 mm. In the current experiment the blob size was chosen to reflect a realistic tumour size and was truncated at ± 3.5 standard deviations with a spatial standard deviation of 0.26 degrees producing a signal that subtended 1.13 degrees and measured 10.3 mm.

The changes in experimental protocol were carefully considered to move towards increased realism and to more closely reflect the clinical scenario and there is no evidence to suggest that any of the changes would result in a loss of effect for the animated movie displays. As with the change from power law noise backgrounds to mammogram backgrounds, a reduction in efficiency as a result of the experimental changes could have been anticipated, however, it is difficult to see how these changes could lead to the loss of effect for animation. Despite this, before drawing conclusions about the incompatibility of real mammograms for use in animated displays, it would be sensible to rule out all other possible causes for the loss of effect, however unlikely they may appear. Further research is, therefore, required to investigate the effectiveness of animation in power law noise and real mammogram backgrounds ensuring that the experimental set up is identical for both background types to rule out the potential confounds discussed above and this will be the aim of the next experiment of this thesis.

Chapter 8

Experiment 5: An Investigation into the use of Animated Presentation for the Detection of a Synthetic Tumour in Real Mammograms and Power Law Noise Backgrounds

8.1 Introduction

Experiment 2a and Experiment 2b, using either Gaussian white noise backgrounds or synthetic mammograms with power law noise backgrounds, demonstrated that observers using animated displays were able to make use of the correlation between image pairs to facilitate lower detection thresholds for a blob in one image of the pair. Experiment 4 investigated the use of animated displays with real mammogram backgrounds, however, the results of Experiment 4 showed that, with real mammogram backgrounds, the observers were unable to use the correlation between the images and did not, consequently, show any advantage in performance, as measured by lower threshold contrasts, over static presentation for the detection of a synthetic tumour in paired mammograms.

It is possible that the properties of real mammograms are not compatible with the use of an animated display technique, as discussed in section 7.5, however, section 7.5 also discussed a number

of potential confounds related to the experimental methodology of Experiment 4. These included the study protocol, which in the earlier experiments was a 2AFC task whereas in Experiment 4, a yes/no rating scale task was used. The size and appearance of the images was also changed with earlier experiments using solid square images of 256 x 256 pixels but Experiment 4 used real mammogram images with an image size of 494 x 768 pixels, determined by the shape of the breast. The real images also contained a whole breast, of varying dimensions, with the remainder of the image being black background. The size of the signal was also changed, primarily to produce a more realistic size for the real images used in Experiment 4.

Whilst there is no evidence to suggest that any of the changes discussed above would result in a loss of effect for the animated displays, this experiment will aim to rule out these potential confounds by comparing the effect of using an animated display with power law noise images and real mammogram backgrounds in an experimental set up that ensures that the same conditions are used for both backgrounds. Thus, this experiment will use a common image size of 500 x 500 pixels, a common blob type and size and a common experimental protocol such that a direct comparison of observer performance with power law noise backgrounds and real image backgrounds can be made. The aim of this experiment will, therefore, be to compare observer performance for the detection of a Gaussian blob in paired images with either a power law noise background or a real mammogram background using either mirror symmetric displays, which will be referred to as static displays or animated displays, which we will refer to as movie displays, over a range of correlation values to determine whether the observers are able to utilise the correlation present between the images.

8.2 Method

8.2.1 Apparatus

The apparatus used in Experiment 5 was the same as for Experiment 1a with the exception that, when viewing from a distance of 52 cm from the monitor screen, the dimensions of each image display were 13.3 deg square.

8.2.2 Images

Two types of image were used; real mammogram image sections extracted from the images supplied by Derriford Hospital's Primrose Unit image bank and power law noise images with low pass $1/f^3$ power spectrum. Both types of image were prepared to ensure that, apart from the background type (real mammogram or power law noise), they were, as far as could be controlled, identical.

Nominal correlation	Correlation range
0.5	0.45 to 0.549
0.6	0.55 to 0.649
0.7	0.65 to 0.749
0.8	0.75 to 0.849
0.9	0.85 to 0.949

Table 8.1: Nominal correlation value of image pairs and range of correlation values within each band as used in Experiment 5.

8.2.2.1 Image preparation

Real mammogram images   The real mammogram images were supplied from an image bank held by Derriford Hospital's Primrose Unit and all images were anonymised by Primrose Unit staff prior to release. Image preparation, including registration of image pairs, was carried out as described in section 7.3.2.1, except the measurement of correlation between the images and the conversion to 494 x 768 pixel bitmap images was not carried out for this experiment. Thus, at this juncture, a large bank of whole breast image pairs, cleaned up and registered were available for further preparation.

This experiment used square sections of the paired images and, to produce these, 1000 x 1000 pixel regions were cropped from random areas of the prepared images with the corresponding 1000 x 1000 regions cropped from the other image of the pair. Care was taken to ensure that the image consisted wholly of breast tissue with no breast boundary or background included. Inclusion of the black background would falsely inflate the correlation between the images and inclusion of skin boundaries would provide potential confounds when comparing against power law noise images with no boundaries. Once resized to 1000 x 1000 pixels, the Pearson correlation between the image pairs was measured and recorded and the images were grouped into five correlation bands as shown in Table 8.1.

Sixty images from each correlation band were selected for use, giving a total of 300 image pairs. The final stage of image preparation was to convert the images to bitmap format and resize them to 500 x 500 pixels for use in the experiment programme. Thus, a bank of 300 real mammogram image pairs were used in the experiment with the same image bank used for both static and movie display modes.

Power law noise images   The power law noise images were generated by frequency domain filtering of Gaussian white noise to give a low pass, $1/f^3$ power law noise image. The image pairs were generated as 500 x 500 pixel images and, once again, 300 image pairs were selected with 5 inter-image correlation levels. The nominal correlation levels were 0.5, 0.6, 0.7, 0.8 and 0.9 with actual inter-image correlation within ± 0.01 of the nominal value. For example, the 0.5 image pairs

contained images with inter-image correlation between the range of 0.49 to 0.51.

8.2.3   The signal

The signal was a Gaussian blob truncated at ±3.0 standard deviations with a spatial SD of 8.75mm (0.96 deg). The signal was added to one image of the pair in a random sequence of image plus signal and image only and was randomly placed centrally within the image (signal known exactly (SKE)). The contrast of the signal, when present, was fixed within each run at 0.01, 0.015, 0.02, or 0.025, as discussed in section 8.2.4. The same signal was used for all display modes and image types.

8.2.4   Procedures

The real mammogram or power law noise image pairs were displayed in either a static or movie display mode, thus giving a set of four presentation modes:

1. Power law noise static.

2. Power law noise movie.

3. Mammogram static.

4. Mammogram movie.

For the static displays the image pair was displayed side-by-side in a mirror symmetric format and, if present, the Gaussian blob signal was added to the right hand image of the pair. The two images were separated by 0.22 deg. For the movie displays the images were presented as the same orientation images in the same location, displayed sequentially with a frame duration of 250 ms. and the Gaussian blob signal, if present, was added to the second image of the pair. For both display modes each image was displayed in a patch of 500 x 500 pixels subtending 13.3 deg square. The image was surrounded by a black region.

The experiment used a yes/no paradigm and on each trial the observer was presented with the stimuli, as shown in Figure 8.1 (power law noise static), Figure 8.2 (power law noise movie), Figure 8.3 (mammogram static) and Figure 8.4 (mammogram movie). Each one of the four presentation modes was presented in a 60 trial block and within one session the presentation modes were run in random order. At the start of each session, the observer was presented with an image of the target signal as a reminder or to familiarise the observer with the signal and they were briefed on the task.
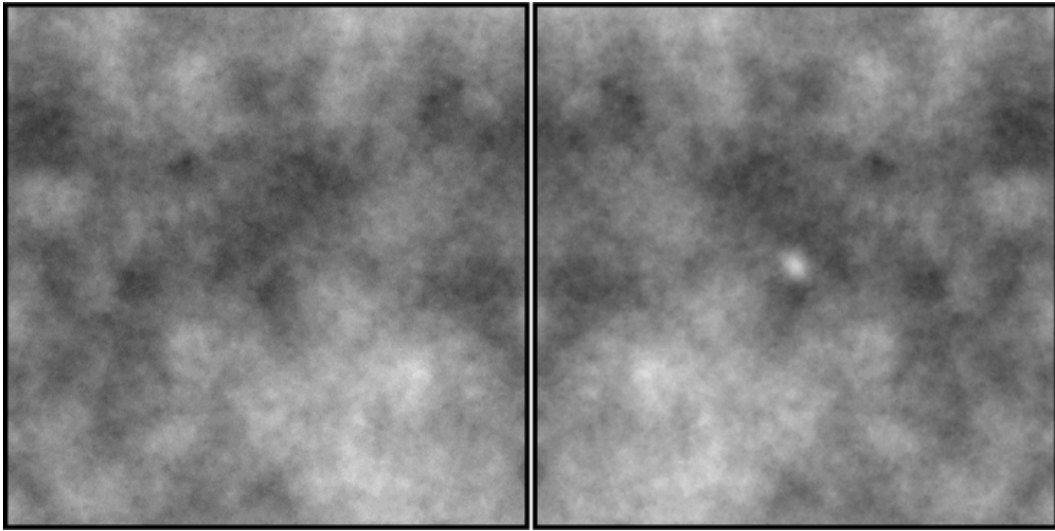
Figure 8.1: Example static display with synthetic mammogram image pair with low-pass $1/f^3$ power spectrum noise. The inter-image correlation shown is 0.9. A Gaussian blob having contrast well above threshold is shown on the right.

Frame duration = 250ms
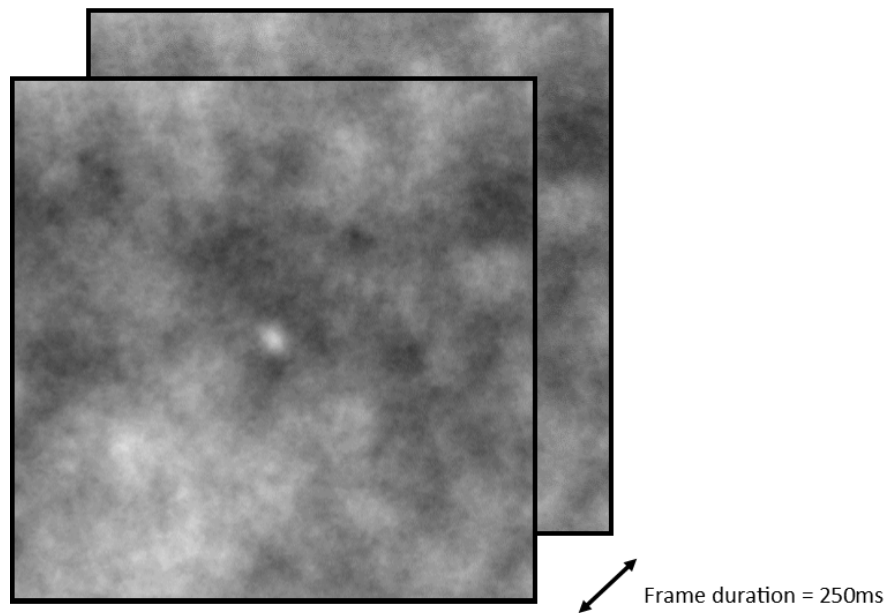
Figure 8.2: Example movie display with synthetic mammogram image pair with low-pass $1/f^3$ power spectrum noise. The inter-image correlation shown is 0.9. A Gaussian blob having contrast well above threshold is shown on the top image for clarity.
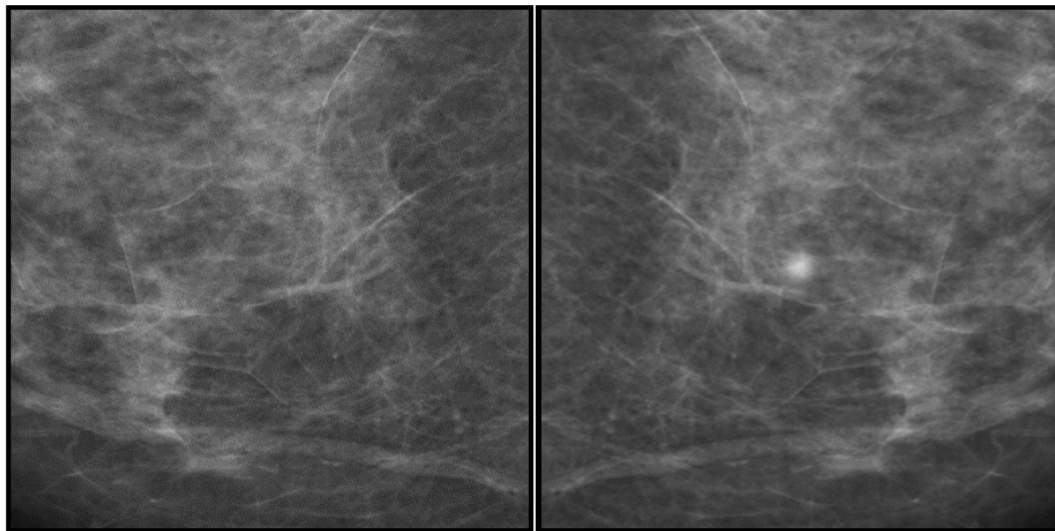
Figure 8.3: Example static display with real mammogram image pair. The inter-image correlation shown is 0.7. A Gaussian blob, having contrast well above threshold, is shown on the right.

Frame duration 250ms
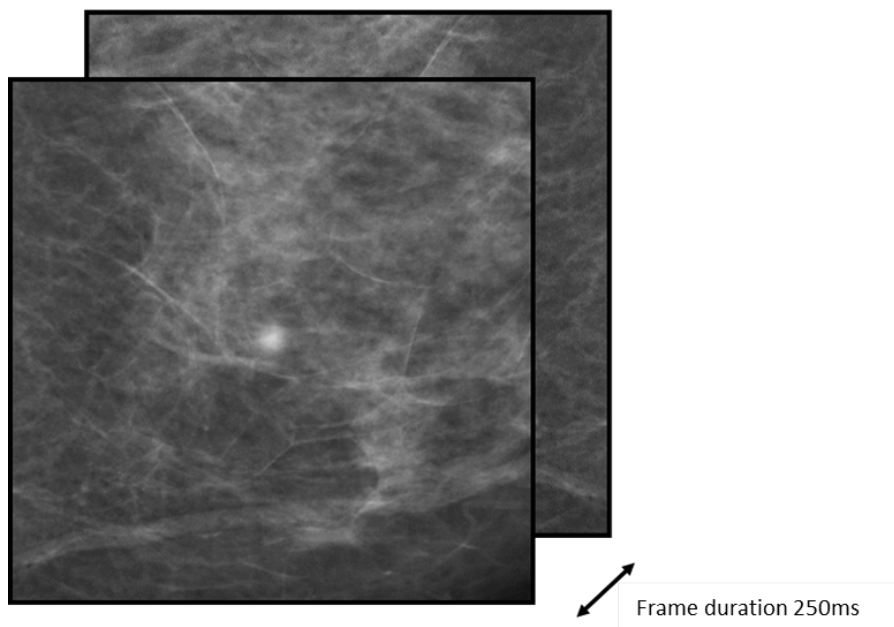
Figure 8.4: Example movie display with real mammogram image pair. A Gaussian blob having contrast well above threshold is shown on the top image for clarity.

A number of pilot trials were carried out to enable the participants to become familiar with the task and to establish the signal contrast to be used for the trials. Contrasts of 0.025, 0.02, 0.015 and 0.01 were piloted to establish the contrast that enabled the observer to achieve a detectability index of d' approximately equal to 1 with real mammogram images at an inter-image correlation of 0.5. Once established, the observer used the same contrast level throughout all of the trials and a contrast of 0.02 was found to be appropriate for 12 observers with one observer (GR - the thesis author and experienced psychophysical observer) using a contrast of 0.01.

To re-familiarise the participants with the task, on the first trial of each run, the signal was always present and the observers were instructed to identify it and select "yes" with a right mouse click. Throughout the testing, no time limit was specified for the observers to make their decisions, however, they were instructed not to dwell too long and make a guess if unsure. The display remained on the screen until a response was given and, apart from the first trial, the signal was randomly present or absent. The observer's task was to click the right mouse button for "yes - signal present" and the left mouse button for "no - signal absent". Incorrect responses were indicated by a pulsed enlargement of the central red square. The observers were instructed to always complete full sets of trials (i.e. power law noise static, power law noise movie, mammogram static, mammogram movie), which were presented in random order. Each participant completed a minimum of 20 full sets over an average of 10 visits.

8.2.5 Observers

Thirteen observers participated in the experiment. All except one were inexperienced observers but received training prior to commencing the study. GR was the author and an experienced psychophysical observer. No observer had any background in radiology or medical physics. All observers had normal or corrected-to-normal vision.

8.3 Results and Discussion

This study aimed to compare observer performance for the detection of a Gaussian blob in paired images with either a power law noise background or a real mammogram background using either mirror symmetric static displays or animated movie displays over a range of correlation values to determine whether the observers are able to utilise the correlation present between the images. Performance was measured using the detectability index d' calculated as follows:

$$d' = z(Hit\ rate) - z(False\ alarm\ rate) \tag{8.1}$$

(Macmillan & Creelman, 2005, p. 9)

where:

$d' =$ detectability index

$z =$ z score

And:

$$Hit\ rate = \frac{H}{N_H} \qquad\qquad (8.2)$$

$$False\ alarm\ rate = \frac{F}{N_F} \qquad\qquad (8.3)$$

where:

$H =$ number of hits

$N_{H=}$ number of signal trials

$F =$ number of false alarms

$N_F =$ number of noise trials

The variance for d' can be calculated as follows

$$var\left(d'\right) = \frac{H\left(1-H\right)}{N_H\left(\varPhi H\right)^2} + \frac{F\left(1-F\right)}{N_F\left(\varPhi F\right)^2} \qquad\qquad (8.4)$$

(Macmillan & Creelman, 2005, p. 271)

where;

$\varPhi =$ height of normal density function

The data from the 13 users was pooled and d' and its standard error for each correlation band were calculated using equations 8.1 and 8.4. The results are plotted in Figure 8.5.

Figure 8.5 shows a clear difference in the performance of real observers for the detection of a Gaussian blob signal with power law noise backgrounds in comparison to real mammogram backgrounds. Two features are apparent. Firstly, there is a qualitative difference in performance for the observers using movie displays with a power law noise background when compared to an observer using movie displays with a mammogram background. With the power law noise background, it is clear that the observer is able to use the correlation between the images to improve performance, whereas, with real mammogram backgrounds, the observer is not and, as a result, shows no improvement in performance as correlation increases. The second feature is a quantitative difference between the performance levels achieved with power law noise backgrounds, which are significantly higher than those achieved with real mammogram backgrounds.

Dealing with the qualitative issue first. If the observers are able to make use of the correlation
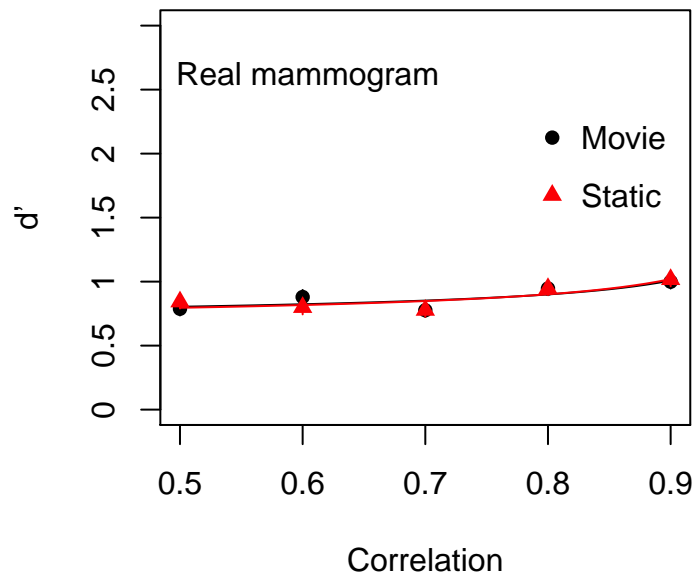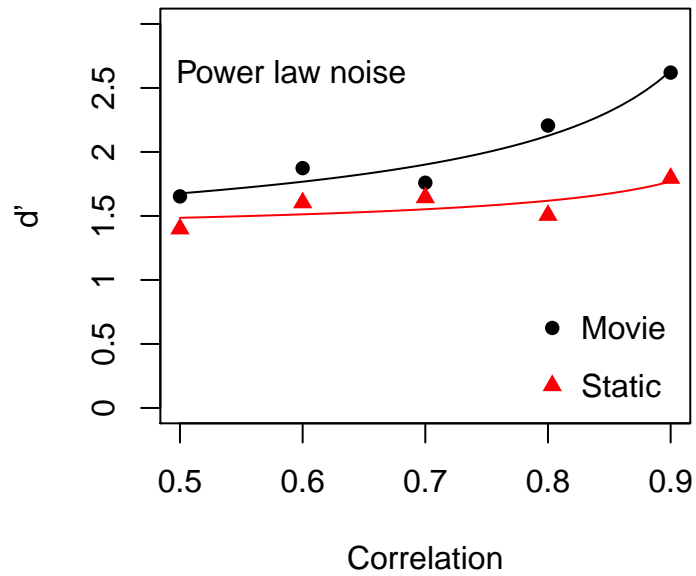
Figure 8.5: d' plotted as a function of inter-image correlation for the detection of a Gaussian blob signal in paired noise images for animated movie displays (circles) and mirrored static displays (triangles) using the pooled data from 13 participants. The top image shows the data plotted for power law noise backgrounds and the bottom image shows the data plotted for real mammogram backgrounds. Error bars showing 95% confidence intervals have been plotted but are not visible as they are smaller than the symbols. Fits of Equation 8.6 are shown.

between the image pairs (as per the ideal observer), then their performance, as measured by d', will increase as the correlation between the images increases. Reference to the section 1.7.1 shows that the performance of the ideal decorrelating observer will be given by the equation:

$$d' = \sqrt{\frac{2E}{(1-\rho)\sigma^2}} \qquad (8.5)$$

(Kay, 1998, p. 112)

Where:

d' = Signal detection theory measure of performance.

E = Signal energy.

ρ = Inter image correlation between the two images of a pair.

σ^2 = Noise variance

The signal energy and noise variance are held constant throughout the experiment and, therefore, for a real observer:

$$d' = a + \frac{k}{\sqrt{1-\rho}} \qquad (8.6)$$

where:

d' = detectability index

ρ = inter-image correlation

a = constant

k = constant

The constant a sets the performance of the observer for an inter-image correlation equal to zero and the constant k determines the observers use of the inter-image correlation. A positive value of k would indicate that the observer is able to make use of the correlation, whereas a value of k equal to zero would show that they were unable to use the correlation. The ideal observer would have a value of k equal to 1. For the pooled observer's data, using movie displays with power law noise, there was a significant positive slope ($k = 0.54$, $p = .006$) indicating that the observer was able to use the correlation between the power law noise pairs. By contrast, for the pooled observer's data, using movie displays with real mammogram backgrounds, the slope was not significantly different from zero ($k = 0.13$, $p = .06$) indicating that the observer was not able to use the correlation between real mammogram pairs. For the pooled observer's data using static displays, the slope was not significantly different from zero with power law noise backgrounds ($k = 0.14$, $p = .16$) or with real mammogram backgrounds ($k = 0.12$, $p = .09$) indicating that the

observers were not able to use the correlation when using static displays with either power law noise pairs or real mammogram pairs.

The second issue related to the quantitative difference between the pooled observer's performance with a power law noise background in comparison to a real mammogram background. The d' values and their standard errors for each condition at correlation of 0.5 are summarised in Table 8.2. A correlation of 0.5 was chosen as this reflects the correlation at which the performance achieved using each mode is closest. Reference to Table 8.2 shows that performance is significantly lower with real mammogram backgrounds than with power law noise backgrounds for movie displays ($t = 12.91$, $p < .001$) and for static displays ($t = 7.84$, $p < .001$) at a correlation of 0.5.

Display type	Background type	d'	SE	t	p
Movie	Power law noise	1.53	0.048	12.91	$<.001$
	Real mammogram	0.70	0.043		
Static	Power law noise	1.31	0.045	7.84	$<.001$
	Real mammogram	0.82	0.044		

Table 8.2: Pooled observer's performance values for animated movie and mirror symmetric static displays with power law noise and mammogram backgrounds at a correlation of 0.5. The final two columns show the t values and significance levels for the difference between the d' values in power law noise and real mammogram backgrounds for the movie and static displays.

8.4   Discussion

The aim of this experiment was to compare observer performance for the detection of a Gaussian blob in paired images with either a power law noise background or a real mammogram background using either mirror symmetric static displays or animated movie displays over a range of correlation values to determine whether the observers are able to utilise the correlation present between the images. It is clear from the results that the observers using movie displays with power law noise backgrounds are able to make use of the correlation between the images to aid in the detection of a Gaussian blob signal in one of the images. However, it is equally clear that when using movie displays with real mammogram backgrounds, they are not able to make the same use of the correlation. Having ensured that the potentially confounding factors relating to the experimental set up, as discussed in Experiment 4, had been negated by using the same conditions for both power law noise backgrounds and for real mammogram backgrounds, the results would suggest that the properties of real mammogram images, in relation to the use of correlation differ from those of power law noise images. These differences will be explored further in section 8.4.1.

8.4.1 Investigating the relationship between the local correlation and global correlation of an image

When faced with the problem of detecting a signal in paired noise fields, the ideal observer would first decorrelate (or pre-whiten) the two images of the pair to remove redundant information. (Note that this stage of decorrelation is taking place between the images, not within each image. The theoretical approach of the ideal observer, outlined in sections 4.2.1 and 4.3.2, suggests that for power law noise or real mammogram images, the ideal observer will first pre-whiten within each image to reduce each image to white noise, before decorrelating between the two images).

Redundant information, in the context of signal detection in paired synthetic or real mammogram images, relates to features of the images that are common to both images and, as a result, the values of those corresponding pixels between the two images will be highly correlated. Decorrelating the image pair will, therefore, remove the redundant information. Greater similarity between the two images will result in a higher correlation between the images and the removal of more redundant information, making the signal easier to detect. Referring to the theory in section 4.2.1, we would expect the ideal observer's performance, as measured by $d^{'}$, to improve as correlation increases in proportion to $\frac{1}{\sqrt{1-\rho}}$, where ρ is the correlation between the two images of the pair. The ideal observer would have a positive slope equal to 1 and the results show that for real observers, when detecting a signal in paired power law noise backgrounds, performance using movie displays also shows a significant positive slope ($slope = 0.54$, $p = .006$), indicating that these observers were able to use the correlation between the images, albeit with a lower efficiency than the ideal observer.

This is not the case for real observers when detecting a signal in real mammogram backgrounds using movie displays, with no significant increase in performance as correlation increased, exhibiting a slope that was not significantly different from zero ($slope = 0.13$, $p = .06$), indicating that these observers were not able to use the correlation between real mammogram pairs.

The question raised by this was; why do observers detecting signals in power law noise appear to use the inter-image correlation, whereas, observers detecting signals in real mammograms do not? This led to the question of what correlation does the visual system see? The correlation reported is the global inter-image correlation for the two 500 x 500 pixel images, however, when detecting a signal in a localised region it was hypothesised that, because of the limited visual angle of the fovea, the visual system would use a more limited area and would utilise this local correlation. The fovea subtends about 1.7 degrees of visual angle, thus at the viewing distance of 520mm the fovea would subtend roughly an area of 60 pixels in diameter.

It had been assumed that the inter-image correlation would be scale invariant; that is, it would

Step number	Image size in pixels
1	500 x 500
2	294 x 294
3	173 x 173
4	102 x 102
5	60 x 60
6	35 x 35
7	21 x 21

Table 8.3: Pixel x pixel dimensions for each image size step as used in section 8.4.1.

remain constant across different scales and, therefore, the correlation between two corresponding localised regions would be equal to the global correlation of the image pair. If the correlation of real mammograms behaved in a non-scale invariant manner this may explain the failure of observers to utilise the correlation and show any improvement in performance as correlation increased. This investigation, therefore, aimed to measure the inter-image correlation for power law noise image pairs and real mammogram pairs across a range of decreasingly small regions of the original images to determine whether the inter-image correlation is scale invariant.

8.4.1.1 Method

For the images used in this investigation, an area of approximately 60 pixels in diameter would fall on the fovea and the signal used in this experiment was a Gaussian blob truncated at ± 3.0 standard deviations with a spatial SD of 8.75mm (0.96 deg), representing a blob diameter of 36 pixels, with only a proportion of this visible at low contrast levels. The range of image region sizes chosen, therefore, ranged from the maximum image size of 500 x 500 pixels down to images which were a 21 x 21 pixel region of the original image, reducing in equal steps by a ratio of 1 : 1.7 to give a reasonable number of region sizes. The region sizes are shown in Table 8.3.

The images used for this investigation were the same 300 real mammogram image pairs used in the main part of this experiment (see section 8.2.2.1 for details of preparation) and 300 power law noise images, generated as 500 x 500 pixel images as described in section 8.2.2.1. The mean correlation for the real mammogram images was 0.70 and, therefore, the starting correlation for the power law noise pairs was chosen also as 0.70.

8.4.1.2 Results

The Pearson correlation was measured between each image pair at each region size and the mean correlation recorded and plotted against the region size with the region size plotted on a logarithmic scale for clarity, as shown in Figure 8.6.

Figure 8.6 shows how the inter-image correlation remains virtually constant for power law
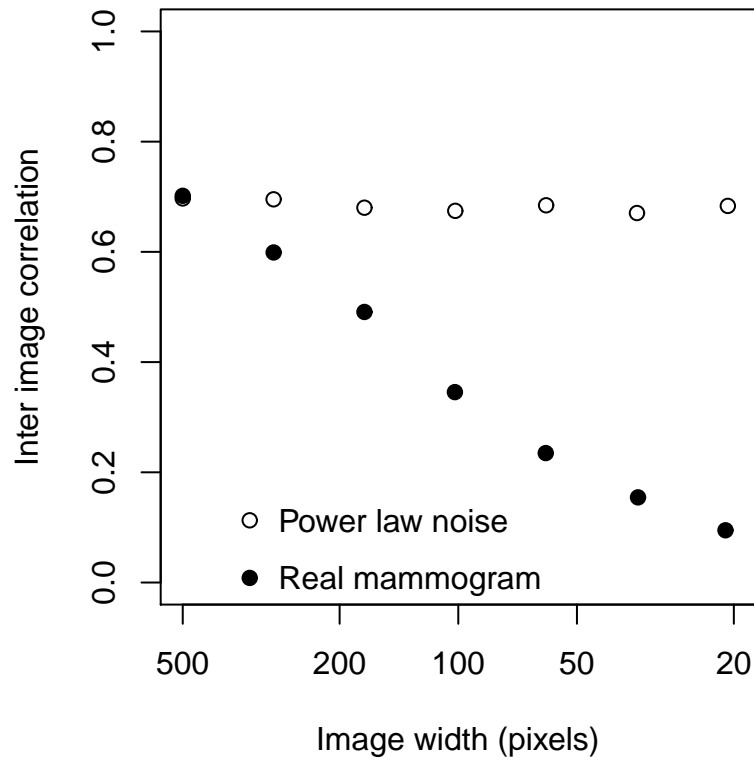
Figure 8.6: Plot of the inter-image correlation against the size of the image region (measured in pixels) for power law noise regions (open circles) and for real mammogram regions (filled circles).

noise image pairs as the size of the region reduces, with a 21 x 21 pixel region displaying the same correlation as the full size image from which it has been cut. This is not the case for real mammogram pairs, where the correlation is seen to fall off linearly as the logarithm of the region size is reduced.

8.4.1.3   Conclusion

This investigation aimed to measure the inter-image correlation for power law noise image pairs and real mammogram pairs across a range of decreasingly small regions of the original images, to determine whether the inter-image correlation is scale invariant. The results are clear; for inter-image correlation, power law noise backgrounds behave in a scale invariant manner, whereas real mammograms do not.

It is suggested that the observer will use only the local background immediately around the signal to be detected and, thus, it is the localised inter-image correlation that is utilised in the decorrelation (or pre-whitening) process. Referring to Figure 8.6, for real mammogram image pairs the correlation between the localised regions surrounding the signal is likely to be approximately 0.2, a level of correlation shown in Experiment 2 to be only marginally more beneficial to observers using movie displays over those using static displays. The results, therefore, may explain the failure

200

to exhibit improved performance using movie displays to assist in the detection of a signal in real mammogram backgrounds, as seen in Experiment 4 and in this experiment.

Before discussing the use of correlation in the two types of images, however, the significant quantitative difference between the pooled observer's performance with power law noise backgrounds in comparison to their performance with real mammogram backgrounds will be briefly discussed. A number of factors serve to impair an observer's performance in real mammograms in comparison to synthetic images (see section 1.6.5). Anatomical noise, as discussed in section 1.6.5, tends to be the most limiting feature in real mammograms, not only affecting the way observers search the image for abnormalities but also mimicking abnormalities, causing false positive responses and masking or camouflaging abnormalities, leading to false negative responses (Samei et al., 2000, pp. 660-678). A combination of these effects would serve to weaken an observer's ability to correctly identify a tumour and would explain the quantitative difference in performance seen between observers using power law noise backgrounds and those using real mammogram backgrounds.

8.5   General Discussion

The aim of this experiment was to compare observer performance for the detection of a Gaussian blob in paired images with either a power law noise background or a real mammogram background using either mirror symmetric static displays or animated movie displays over a range of correlation values to determine whether the observers are able to utilise the correlation present between the images. It was found that static displays did not facilitate the use of the correlation between the two images of a pair for either background type, with no evidence of improvements in performance as correlation increased. For observers using movie displays, however, it was found that with power law noise images they were able to utilise the correlation between the images of a pair and show improved performance as correlation increased, whereas, with real mammograms, they were not. Thus, summarising, observers using movie displays with paired power law noise backgrounds were able to decorrelate the two noise patches of the pair, whereas with movie displays and real mammogram noise backgrounds and with static displays and both types of noise background they were not.

A subsequent investigation was carried out into why, with movie displays, the observers were able to decorrelate the background noise patches when they consisted of power law noise but not able to do so when the background noise patches were real mammograms. It was hypothesised that the observers would decorrelate a local region around the blob, being limited to the visual angle subtended by the fovea. This subsequent investigation, therefore, aimed to measure the inter-

image correlation for power law noise image pairs and real mammogram pairs across a range of decreasingly small regions of the original images to determine whether the inter-image correlation is constant regardless of image size.

The real mammogram images were found not to be scale invariant, with the correlation decreasing as the size of the region taken from the image reduced. The observers using movie displays with real mammograms would, therefore, be effectively attempting to decorrelate the localised areas where the correlation would, referring to Figure 8.6, be expected to be approximately 0.2. This could, therefore, explain the failure of the observers to decorrelate the image pairs.

The non-scale invariant nature of real mammograms suggests that the global correlation of the whole image is not a good indicator of whether a movie display technique would be effective. It should not be concluded from this, however, that a movie display is not viable for use with real mammograms, rather, that a different measure, possibly based on local correlation, is necessary. Although mammograms are widely modelled as $1/f^3$ noise (Burgess & Judy, 2007; Burgess et al., 2001; Reiser et al., 2013), it is clear that they are not truly $1/f^3$, because such noise is scale invariant whereas mammograms are not.

For this thesis, the Pearson correlation between the whole breast area of each image (the global correlation) has been used as the measure of the level of correspondence between the two images of a pair. The Pearson correlation is a widely used measure (Bozek et al., 2011b; Celaya-Padilaa et al., 2013; Chiou et al., 2007; Guo et al., 2006) and, in a comparison of 12 different similarity measures, Filev et al. (2005) found that, along with the cosine coefficient and Goodman and Kruskal's Gamma coefficient, Pearson correlation gave the best indication of the alignment of corresponding masses on paired mammograms. Indeed, the results seen in this thesis, with power law noise, provide further support for the use of Pearson correlation as a similarity measure. If the contrast threshold for the detection of a blob is used as a measure of effectiveness of a display technique, with lower contrast thresholds indicative of better performance, then the relationship between correlation and the contrast threshold for detection seen in Experiment 2a and Experiment 2b shows that the inter image correlation is a good predictor of performance.

Notwithstanding this, the global nature of the Pearson correlation used in this thesis does not reflect the localised correlations that we now believe to be important for the effective use of an animated display technique. Indeed, the registration method used to maximise the Pearson correlation between the images of a pair, as used in this thesis, may not be suitable for aligning localised regions of the breast to maximise these local correlations. The registration algorithm that has been used was BUnwarpJ, a 2D image registration method based on elastic deformations using identifiable landmarks, such as the nipple and skin boundary, which effectively aligns the

major features of the image but may be less effective at bringing the corresponding areas of the parenchyma into closer correlation. An alternative registration method, perhaps based on localised regions of the breast, may improve local correlation between corresponding areas and, thus, enable the effective use of an animated movie display technique with real mammograms.

A number of regional registration techniques have been tested, such as exemplified by Timp et al. (2005), who identified localised search regions and used three registration measures, Pearson correlation, mass likelihood and a distance measure, to improve the registration quality. They verified the effectiveness of their registration technique by demonstrating an improved detection rate for masses when compared to other methods (Timp et al., 2005). Similarly, Hadjiiski et al. (2001) included a local alignment step, registering localised regions and matching to the closest region in the other image, once again, improving lesion detection rates and Sanjay-Gopal et al. (1998) have developed an automated regional registration technique to improve lesion detection on mammograms. The studies cited here have measured the effectiveness of registration by conducting detection trials, but do not report the localised correlation measures. To further the research developed in this thesis, however, the Pearson correlation values of corresponding local regions will be required to investigate the hypothesis that the human visual system uses these local correlation values to decorrelate the corresponding localised regions of image pairs to assist in the detection of the blob signal. To achieve this, further research into registration techniques, identifying the method that best improves local correlation values is required.

Experiment 2a and Experiment 2b clearly demonstrate the effectiveness of animation, as used in the movie display mode, as a tool to enable the observer to utilise the correlation between the images and, as a result, facilitate lower detection thresholds. The failure of animated displays to replicate this with real mammograms does not negate this but indicates that the global inter-image correlation of a real mammogram pair does not reflect the local inter-image correlation of the regions subtended by the fovea. Whilst currently available registration techniques can increase global correlation levels, these techniques do not enhance the correlation between localised regions of the images and, therefore, do not improve the effectiveness of an animated display. For animated displays to be of benefit with real mammograms, it will be necessary to develop a registration technique biased towards regional or local registration, such that the correlation between the corresponding localised areas of the breast parenchyma is maximised. If this can be achieved, it is believed that animated displays could prove to be an effective display mode with real mammogram images that will support lower detection thresholds and earlier detection of tumours.

Chapter 9

General Discussion and Conclusions

The aim of this thesis was to investigate whether an animated display technique could be an effective methodology for the detection of a tumour in a paired mammogram display. A psychophysical approach was used to address this goal, first establishing a theoretical benchmark that could be tested in progressively more realistic scenarios. The theoretical benchmark was determined through the use of the concept of the ideal observer, a theoretical device that achieves optimal performance for a designated task (Geisler, 2011; Green & Swets, 1966, p. 151), thus establishing the optimum strategy against which the subsequent human observers' results could be compared. To minimise the possibility that the results may be confounded by extraneous variables, such as may be introduced by the use of real mammogram images or unknown signals, the research was first conducted using Gaussian white noise backgrounds with a Gaussian blob signal in a signal known exactly protocol; referred to by Kay (1998, p. 94) as the "...simple versus simple hypothesis...", before progressing to power law noise backgrounds and mammogram backgrounds and variable signal positions.

The ideal observer strategy for the detection of a signal known exactly in paired Gaussian white noise backgrounds was calculated and checked through simulations. The ideal strategy was found to be the decorrelation (or pre-whitening) of the paired noise fields, thus enabling the observer to discount redundant information (Hyvärinen et al., 2009, p. 126), before cross-correlating the stimulus with a template of the signal (Burgess & Ghandeharian, 1984a). Redundant information relates to pixel values of the images that are common to, or related in, both images and, as a result, the values of those corresponding pixels between the two images will be highly correlated. Decorrelating the image pair will, therefore, remove the redundant information. Greater similarity between the two images will result in higher correlation between the images and the removal of more redundant information, making the signal easier to detect; thus as the level of correlation

increases, the contrast threshold for the detection of the blob signal will decrease. With a power law or real mammogram background, the ideal strategy was shown to be to first pre-whiten within each image before decorrelating between the images and cross-correlating the resulting stimuli with a template of the signal.

Whilst the ultimate goal was the investigation of an animated display mode, Experiment 1 first tested the effectiveness of the traditional side-by-side, mirror symmetric displays, currently the most widely used reading protocol adopted by breast radiologists (Haygood & Dogan, 2013; Kopans, 2007, p. 367). This provided a comparison against which the performance of animated displays could be compared and, also, investigated the widely held, but seemingly anecdotal, belief that the use of mirror symmetry would assist the radiologist in the detection of abnormalities. The results of Experiment 1 showed that human observers using side-by-side, mirror symmetric displays were not able to use the optimum strategy of decorrelating the paired noise fields and cross correlating the resulting stimuli with, at best, only moderate reductions in contrast threshold as the correlation increased. Most observers showed no improvement as correlation increased and mirror symmetric displays were found to not elicit significantly different contrast thresholds across the range of correlation from repeat symmetrical displays. This was not an unexpected finding and a similar result to Pomplun (1998) who found no benefit of mirror symmetric presentation over repeat presentation for the detection of the target in one noise field of the symmetric pair. A number of other studies relating to the ability of the human visual system to detect violations of mirror symmetry also raised doubts about how effective mirror symmetric displays would be for the detection of a signal in one half of the symmetric pattern. These included Bruce & Morgan (1975), who showed that violations were weakly detected when they were distant from the mid-line, a finding echoed by Barlow & Reeves (1979) and Jenkins (1982). Symmetry detection by the human visual system has also been shown to be weaker in more complex images (Huang & Pashler, 2002; Rainville & Kingdom, 2002; Tapiovaara, 1990) and, additionally, to be relatively insensitive to minor violations of the symmetric pattern (Tjan & Liu, 2005), both conditions reflecting the images and the synthetic tumour used in the experiments in this thesis.

The research cited here may explain the failure of the human observers to benefit from the symmetric presentation of paired images seen in Experiment 1, however, this failure may be underpinned by the inability to make use of the correlation between the image pairs when they are presented side by side, in either mirror symmetric format or repeat symmetric format. The results of Experiment 1 certainly suggest an inability to decorrelate the image pairs and this is supported by the results of studies by Ahumada & Beard (1997) and Watson et al. (1997), neither of whom found any benefit for the detection of a signal in image pairs with a correlation of one between

the images over that achieved with image pairs with a correlation of zero. Although the results in the study by Burgess & Colborne (1988) possibly do lend support to the notion that their observers were able to decorrelate the image pairs, reporting a difference in performance between paired images, displayed side-by-side, with no correlation and perfect correlation (i.e. no symmetry and perfect symmetry), no significance levels or confidence limits were given and it is, therefore, difficult to draw firm conclusions from this.

Numerous clinical studies have shown the benefit of using paired images in mammography, particularly paired images taken some time interval apart (Callaway et al., 1997; Frankel et al., 1995; Roelofs et al., 2007; Sickles, 2007; Sumkin et al., 2003; Thurfjell et al., 2000; White et al., 1994)), however, the conclusion drawn from Experiment 1 is that displaying paired images side-by-side, in either a mirror symmetric or repeat symmetric format, does not enable the human observer to benefit from the correlation between the images and does not facilitate the optimum use of the information available within the images. As a result, we, therefore, suggest that mirror symmetric displays do not represent the optimum viewing modality for paired image displays such as those frequently encountered in mammography; they are no better, or worse, than repeat symmetric displays. An optimal display mode would be one that enables the human observer to operate with a strategy similar to that of the ideal observer, to allow the human observer to utilise the correlation between the images and, hence, improve detection performance.

One option is the use of an animated display mode and the aim of Experiment 2 was to investigate whether the use of animated presentation would enable human observers to use the correlation present between the images and decorrelate the paired noise fields in the same way as the ideal observer. The results of Experiment 2 showed that animated displays did provide a display mode that enabled human observers to use the correlation between the images of a pair to improve their performance, operating with a similar strategy to that of the ideal observer. The results illustrate the sensitivity of the human visual system to moving and flickering stimuli, as highlighted by Adelson & Bergen (1986); Franconeri et al. (2005); Spalek et al. (2009); Watson (1986). Mammogram pairs can achieve a high level of correlation that, in certain circumstances can be increased by image registration, thus making a decorrelation strategy an ideal approach for removing redundant information and facilitating easier detection of any potential tumour. Thus, animated displays make a viable and attractive proposition for use in mammography and the ability to utilise the correlation between paired images could be the mechanism underpinning the successful use of image toggling techniques (Hasegawa et al., 2008; Honda et al., 2014; van Engeland et al., 2003). The need for high levels of correlation is illustrated by Hasegawa et al. (2008) and Honda et al. (2014) who emphasised the requirement for accurate registration to enable

toggling between images to be an effective method for finding abnormalities in mammograms. It is interesting to note that the toggling techniques employed by Hasegawa et al. (2008); Honda et al. (2014); van Engeland et al. (2003) did not specify an optimum rate of toggling and this was under the control of the observer.

The results of Experiment 3 are, therefore, pertinent, showing that varying the rate of alternation had a weak effect with Gaussian white noise and power law noise backgrounds, with an optimum alternation rate at approximately 2Hz. Thus, the use of a variable alternation rate, under the control of the radiologist, should not compromise the improved performance achievable through the use of animated displays but would enable the radiologist to alternate the images at a rate that they find comfortable and also permit the freezing of the images to allow the radiologist to focus in on features of particular interest. This feature would offer a valuable additional tool in the radiologist's armoury, using an animated movie display mode to facilitate improved detection of differences between paired mammograms, that may represent a tumour, and freezing the display to examine the differences identified in closer detail, thus facilitating improved detection and discrimination. Of course, at this juncture, this claim is unsubstantiated and would require further research with real mammograms and radiologists to investigate this.

It is clear that for a decorrelation strategy the inter image correlation is the primary factor in determining the performance of an observer when attempting to detect a signal in paired noise fields, such as encountered in mammography, and this is clearly seen from the results of Experiment 2. Thus, Experiment 4 used registered mammogram pairs and aimed to demonstrate the effectiveness of an animated display technique for the detection of a synthetic tumour in real mammogram backgrounds. With registration improving the global correlation levels to a range between 0.65 and 0.97, with a mean correlation value of 0.84, it was expected that animated displays would facilitate lower detection thresholds than those achieved using the mirror symmetric displays. The results of Experiment 4, however, showed this not to be the case with no difference found in performance levels for the two display modes. It was surmised that this may be an indication that the properties of real mammograms differed from those of power law noise images, in some way, as yet unknown, that rendered an animated display mode unsuitable. This, however, was considered unlikely, particularly considering the success seen with toggled displays (Hasegawa et al., 2008; Honda et al., 2014; van Engeland et al., 2003), and a number of issues were also raised against the experimental set up that may have contributed to the failure to see an effect with animated displays. Further research was, therefore, carried out in Experiment 5, with all the known confounding variables eradicated enabling a direct comparison of observer performance for the detection of a synthetic signal embedded in real mammogram backgrounds and power law noise backgrounds.

The aims of Experiment 5 were, therefore, to compare observer performance for the detection of a Gaussian blob in paired images with either a power law noise background or a real mammogram background using either mirror symmetric static displays or animated movie displays over a range of correlation values to determine whether the observers are able to utilise the correlation present between the images. The results were clear; as expected, observers were not able to use the correlation between the images to improve their performance with the images presented side-by-side, in a mirror symmetric static display with either power law noise backgrounds or real mammogram backgrounds. For both background types, with static displays, varying the correlation had no effect on observer performance.

The picture was very different, however, when power law noise images were presented as animated movie displays. Increasing the correlation between the images of a pair now enabled the observers to significantly improve their performance, replicating the results seen in Experiment 2. The observers were utilising the correlation present between the images, using a similar strategy to the ideal observer of decorrelation to reduce redundant information within the image pair and enable easier detection of the signal. The introduction of real mammogram backgrounds, however, was seen to negate the effect of animated movie displays and observers using movie displays with real mammogram backgrounds were not able to use the correlation between the images to improve their performance. As with the static displays, varying the correlation had no effect on observer performance.

This was an intriguing finding, prompting a search for the property associated with real mammograms that caused the effect of animation to disappear and, with an awareness of the importance of the inter-image correlation for a decorrelation or differencing strategy, concentrating on this property. It was theorised that the important correlation value was the correlation between corresponding localised areas in the two mammograms of the pair, the size of which possibly reflecting the visual angle subtended by the fovea. This investigation revealed that, when measuring the local correlation between corresponding regions of the image pairs for decreasingly small windows of the original images, power law noise images exhibited scale invariance, with a constant correlation level as the window size decreased, whereas real mammogram pairs did not, exhibiting a decreasing correlation level as the window size decreased. For real mammograms, using the mean values for the set of images, a correlation value of 0.84 for the whole image (the global correlation) decreased to approximately 0.2 when the window size was reduced to a size reflecting the visual angle subtended by the fovea. If the premise that the ability of human observers to decorrelate between two images of a pair is limited by the visual range of the fovea, this would explain the failure of observers to use the correlation between images and the failure to observe an improvement

in performance as correlation increased.

The results of Experiment 5 indicate that, for animated movie displays to be a viable alternative to mirror symmetric static displays in clinical mammography, it is necessary to identify, or develop, a registration algorithm that emphasises the local correlation values and this should be the focus of future research. A number of possible techniques that focus on regional or localised registration have been researched (Hadjiiski et al., 2001; Sanjay-Gopal et al., 1998; Timp et al., 2005), however, none of these techniques have been aimed towards the use of animated display technology. Future research should, therefore, concentrate on the investigation and development of registration algorithms that enhance the correlation levels of corresponding localised regions in paired mammograms. It is unclear, at this stage, what measure of registration efficacy would be used; clearly, global Pearson correlation is not a suitable candidate, however, a number of other possibilities exist such as mutual information, Euclidean distance measures or visual inspection (Guo et al., 2006). Certainly, one measure of the effectiveness of a registration technique, could reflect the experiments used in this thesis, measuring the performance of human observers using an animated display for the detection of a signal in paired noise backgrounds. Higher performance levels would indicate better registration and other candidate measures could be compared to the performance scores of the human observers to determine an appropriate measure. For some images, however, it may not be possible to resolve the problem through better registration. It may be the case that even perfectly registered images will exhibit low correlation levels, both globally and locally, due to differences resulting from circumstances such as breast tissue changing over time, differing levels of compression within the scanner and variation in the angle of x-ray delivery.

9.0.1   Conclusion

It is clear that animated displays do enable the human observers to utilise the correlation between images and, as a result, improve their performance for the detection of signals in paired noise backgrounds and it is equally clear that mirror symmetric displays, as traditionally used in mammography, do not. Although this finding is tempered by the subsequent finding that, when using real mammograms, the registration methods used in this thesis did not provide the localised correlation levels necessary to exploit this powerful tool, this finding, nevertheless, represents an important opportunity to develop imaging techniques that could lead to improved cancer detection and diagnosis and to improved outcomes for those affected by this disease.

References

Abbey, C. K. & Barrett, H. H. (2001). Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 18(3), 473–488.

Abbey, C. K. & Boone, J. M. (2008). An ideal observer for a model of x-ray imaging in breast parenchymal tissue. In E. A. Krupinski (Ed.), *Digital Mammography*, volume 9th International Workshop of *Lecture Notes in Computer Science 5116* (pp. 393–400). Tucson, AZ, USA: IWDM Springer.

Abbey, C. K. & Eckstein, M. P. (2000). *Estimates of human-observer templates for a simple detection task in correlated noise*, (pp. 70–77). SPIE - International Society for Optical Engineering.

Abbey, C. K. & Eckstein, M. P. (2002). Classification image analysis: estimation and statistical inference for two-alternative forced-choice experiments. *Journal of Vision*, 2(1), 66–78.

Abbey, C. K. & Eckstein, M. P. (2007). Classification images for simple detection and discrimination tasks in correlated noise. *Journal of the Optical Society of America A*, 24(12), B110–B124.

Abbey, C. K. & Eckstein, M. P. (2010). *The Handbook of Medical Image Perception and Techniques*, chapter Observer models as a surrogate to perception experiments, (pp. 240–250). Cambridge University Press: Cambridge, United Kingdom.

Abbey, C. K., Zemp, R. J., Liu, J., Lindfors, K. K., & Insana, M. F. (2006). Observer efficiency in discrimination tasks simulating malignant and benign breast lesions with ultrasound. *IEEE Transactions on Medical Imaging*, 25(2), 198–209.

Abramoff, M. D., Magalhaes, P. J., & Ram, S. J. (2004). Image processing with imagej. *Biophotonics International*, 11(7), 36–42.

Adelson, E. H. & Bergen, J. R. (1986). The extraction of spatio-temporal energy in human and machine vision. *Proceedings from the Workshop on Motion: Representation and Analysis*, (pp. 151–155).

Advisory Committee on Breast Cancer Screening (2006). *Screening for breast cancer in England: Past and future [NHSBSP Publication No 61]. Sheffield: NHS Cancer Screening Programmes. Retrieved from:* *http://www.cancerscreening.nhs.uk/breastscreen/publications/nhsbsp61.pdf.* Technical Report NHSBSP Publication No 61, NHS Cancer Screening Programmes, Sheffield.

Ahumada, A. J. & Beard, B. L. (1997). Image discrimination models: Detection in fixed and random noise. *Human Vision, Visual Processing, and Digital Display VIII*, 3016, 34–43. In B. Rogowitz & T. N. Pappas (Eds.).

Andolina, V. F. & Lillé, S. L. (2010). *Mammographic imaging: A practical guide.* Philadelphia: Lippincott Williams & Wilkins, 3rd edition.

Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing? *Network: Computation in Neural Systems*, 3, 213–251.

Atkins, M. S., Moise, A., & Rohling, R. (2006). An application of eyegaze tracking for designing radiologists' workstations: Insights for comparative visual search tasks. *ACM Transactions on Applied Perception*, 3(2), 136–151.

Barlow, H. (2001). Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12, 241–253.

Barlow, H. B. (1956). Retinal noise and absolute threshold. *Journal of the Optical Society of America*, 46(8), 634–639.

Barlow, H. B. (1957). Increment thresholds at low intensities considered as signal/noise discriminations. *Journal of Physiology*, 136, 469–488.

Barlow, H. B. (1961). *Current Problems in Animal Behaviour*, chapter The coding of sensory messages, (pp. 331–360). Cambridge University Press: Cambridge.

Barlow, H. B. (1977). *Photoreception in Vertebrates*, chapter Retinal and central factors in human vision limited by noise. Academic Press: London.

Barlow, H. B. (1978). The efficiency of detecting changes of density in random dot patterns. *Vision Research*, 18(6), 637–650.

Barlow, H. B. & Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research*, 19, 783–793.

Baylis, G. C. & Driver, J. (1994). Parallel computation of symmetry but not repetition in single visual shapes. *Visual Cognition*, 1, 377–400.

Baylis, G. C. & Driver, J. (2001). Perception of symmetry and repetition within and across visual shapes: Part-descriptions and object based attention. *Visual Cognition*, 8(2), 163–196.

Baylor, D. A., Matthews, G., & Yau, K. W. (1980). Two components of electrical dark noise in toad retinal rod outer segments. *Journal of Physiology*, 309, 591–621.

Blackwell, H. R. (1946). Contrast thresholds of the human eye. *Journal of the Optical Society of America*, 25, 624–643.

Blakemore, C. & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology*, 203, 237–260.

Bochud, F. (2013). *Physics of mammographic imaging*, chapter Observer models for breast imaging, (pp. 153–164). CRC Press/Taylor & Francis: Boca Raton.

Bochud, F. O., Abbey, C. K., & Eckstein, M. P. (2004). Search for lesions in mammograms: statistical characterization of observer responses. *Medical Physics*, 31, 24–36.

Bochud, F. O., Valley, J. F., Verdun, F. R., Hessler, C., & Schnyder, P. (1999). Estimation of the noisy component of anatomical backgrounds. *Medical Physics*, 26, 1365–1370.

Bontrager, K. L. & Lampignano, J. (2013). *Textbook of Radiographic Positioning and Related Anatomy*. St Louis Missouri: Elsevier Health Sciences, 8th edition.

Bozek, J., Grgic, M., & Schnabel, J. A. (2011a). Validation of rigid registration of mammographic images. Number 53 in International Symposium ELMAR-2011 Zadar, Croatia.

Bozek, J., Grgic, M., & Schnabel, J. A. (2011b). Validation of rigid registration of medical images. *Proceedings 53rd International Symposium ELMAR-2011*, (pp. 11–16).

Bruce, V. G. & Morgan, M. J. (1975). Violations of symmetry and repetition in visual patterns. *Perception*, 4(3), 239–249.

Building Services Index (2015). *Audible fire alarms and the new BS EN 54-23 standard*. Technical report.

Bun, P. A. M. (2002). *The practice of mammography: Pathology, technique, interpretation, adjunct modalities*, chapter Strategy for viewing the mammogram, (pp. 184–186). Thieme Publishers: Stuttgart.

Burgess, A. (1985). Visual signal detection. III. On bayesian use of prior knowledge and cross correlation. *Journal of the Optical Society of America*, 2(9), 1498–1507.

Burgess, A. E. (1999). Visual signal detection with two-component noise: Low-pass spectrum effects. *Journal of the Optical Society of America A, Optics and Image Science*, 16, 694–704.

Burgess, A. E. (2010). *Signal detection theory - a brief history*, (pp. 26–46). Cambridge University Press: Cambridge, United Kingdom.

Burgess, A. E. & Colborne, B. (1988). Visual signal detection. IV. Observer inconsistency. *Journal of the Optical Society of America*, 5(4), 617–627.

Burgess, A. E. & Ghandeharian, H. (1984a). Visual signal detection. I. Ability to use phase information. *Journal of the Optical Society of America A*, 1(8), 900–905.

Burgess, A. E. & Ghandeharian, H. (1984b). Visual signal detection. II. Signal-location identification. *Journal of the Optical Society of America*, 1(8), 906–910.

Burgess, A. E., Jacobson, F. L., & Judy, P. F. (2001). Human observer detection experiments with mammograms and power-law noise. *Medical Physics*, 28, 419–437.

Burgess, A. E., Jennings, R. J., & Wagner, R. F. (1982). Statistical efficiency: A measure of human visual signal-detection performance. *Journal of Applied Photographic Engineering*, 8, 76–78.

Burgess, A. E. & Judy, P. F. (2007). Signal detection in power-law noise: Effect of spectrum exponents. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 24(12), B52–B60.

Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science*, 214(4516), 93–94.

Bushberg, J. T., Ed. (2002). *The essential physics of medical imaging*. Lippincott Williams & Wilkins.

Callaway, M. P., Boggis, C. R. M., Astley, S. A., & Hutt, I. (1997). The influence of previous films on screening mammographic interpretation and detection of breast carcinoma. *Clinical Radiology*, 52(7), 527–529.

Campbell, F. W. & Green, D. G. (1965). Optical and retinal factors affecting visual resolution. *Journal of Physiology (London)*, 181, 576–593.

Campbell, F. W. & Gubish, R. W. (1966). Optical quality of human eye. *Journal of Physiology (London)*, 186, 558–578.

Campbell, F. W. & Robson, J. G. (1968). Application of fourier analysis to the visibility of gratings. *Journal of Physiology*, 197(3), 551–566.

Cancer Research UK (2014a). Breast cancer. Retrieved from: Cancer Research UK: $http :$ $//publications.cancerresearchuk.org/downloads/Product/CS_K F_B REAST.pdf$.

Cancer Research UK (2014b). Breast cancer mortality statistics. Retrieved from: Breast Cancer UK: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/breast/mortality/uk-breast-cancer-mortality-statistics.

Cancer Research UK / Wikimedia Commons (2014). The breasts and lymphatic system. Retrieved from: Breast Cancer UK: http://www.cancerresearchuk.org/about-cancer/type/breast-cancer/about/the-breasts-and-lymphatic-system.

Carlbom, I. (1994). Computer- assisted registration, segmentation, and 3d reconstruction from images of neuronal tissue sections. *IEEE Transactions on Medical Imaging*, 13(2), 351–362.

Castella, C., Eckstein, M. P., Abbey, C. K., Kinkel, K., Verdun, F. R., Saunders, R. S., Samei, E., & Bochud, F. O. (2009). Mass detection on mammograms: Influence of signal shape uncertainty on human and model observers. *Journal of the Optical Society of America A, Opt Image Sci Vis.*, 26(2), 425–436.

Castella, C., Kinkel, K., Verdun, F. R., Eckstein, M. P., Abbey, C. K., & Bochud, F. O. (2007). Mass detection on real and synthetic mammograms: Human observer templates and local statistics. *Proc. SPIE 6515, Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment*, 6515.

Celaya-Padilaa, J. M., Rodriguez-Rojas, J., Trevino, V., & Tamez-Pena, J. G. (2013). Local image registration a comparison for bilateral registration mammography. *Proc. SPIE 8922, IX International Seminar on Medical Information Processing and Analysis*, 8922.

Chakraborty, D. P. (2006). An alternate method for using a visual discrimination model (VDM) to optimize soft-copy display image quality. *Journal of the Society for Information Display*, 14(10), 921–926.

Chiou, Y., Lin, C., & Lin, C. (2007). Hybrid registration of corresponding mammogram images for automatic detection of breast cancer. *Biomedical Engineering: Applications, Basis and Communications*, 19(6), 359–374.

Cohn, T. E., Ed. (1993). *Visual detection: Collected works in optics*, volume 3. Washington, DC: Optical Society of America.

Conrey, B. & Gold, J. M. (2009). Pattern recognition in correlated and uncorrelated noise. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 26(11), B94–B109.

Cook, S., Quigley, C., & Clift, L. (2000). *Motor Vehicle and Pedal Cycle Conspicuity - Part 1: Vehicle Mounted Warning Beacons, Summary Report.* Technical Report Final Report: 9/33/13, The Department of the Environment, Transport and the Regions.

Cornsweet, T. N. (1970). *Visual perception.* London: Academic Press.

Dakin, S. C. & Herbert, A. M. (1998). The spatial region of integration for visual symmetry detection. *Proceedings. Biological Sciences / The Royal Society*, 265, 659–664.

Darvas, G. (2007). *Symmetry.* Basel/Berlin/Boston: Birkhäuser.

de Lange, H. (1958). Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light. I. Attention characteristics with white and colored light. *Journal of the Optical Society of America*, 48(11), 777–783.

Destexhe, A. & Rudolph-Lilith, M. (2012). *Neuronal noise*, volume 8 of *Springer Series in Computational Neuroscience.* Springer Science & Business Media.

Dogan, B. E. (2012). *Digital mammography: A practical approach*, chapter Preparing digital mammography images for interpretation, (pp. 22–26). Cambridge University Press: Cambridge.

Ellemberg, D., Wilkinson, F., Wilson, H. R., & Arsenault, A. S. (1998). Apparent contrast and spatial frequency of local texture elements. *Journal of the Optical Society of America A, Optics and Image Science*, 15(7), 1733–1739.

Eltonsy, N. H., Elmaghraby, A. S., & Tourassi, G. D., Eds. (2007). *Bilateral breast volume asymmetry in screening mammograms as a potential marker of breast cancer: Preliminary experience*, volume 5 of *IEEE International Conference on Image Processing.* IEEE.

Erickson, B. J., Wood, C. P., Kaufmann, T. J., Patriarche, J. W., & Mandrekar, J. (2011). Optimal presentation modes for detecting brain tumor progression. *American Journal of Neuroradiology*, 32, 1652–1657.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, Optics and Image Science*, 4(12), 2379–2394.

Field, D. J. (1999). Wavelets, vision and the statistics of natural scenes. *Philosophical Transactions of the Royal Society A*, 357(1760), 2527–2542.

Field, D. J. & Brady, N. (1997). Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research*, 37(23), 3367–3383.

216

Filev, P., Hadjiiski, L., Sahiner, B., Chan, H. P., & Helvie, M. A. (2005). Comparison of similarity measures for the task of template matching of masses on serial mammograms. *Medical Physics*, 32(2), 515–529.

Fleury, S. & Jamet, E. (2014). Facilitating the comparison of multiple visual items on screen: The example of electronic architectural plan correction. *Applied Ergonomics*, 45, 601–607.

Fraknoi, A. (2009). The discovery of Pluto: Generally unknown aspects of the story.

Franconeri, S. L., Hollingworth, A., & Simons, D. J. (2005). Do new objects capture attention? *Psychological Science*, 16(4), 275–281.

Frankel, S. D., Sickles, E. A., Curpen, B. N., Sollitto, R. A., Ominsky, S. H., & Galvin, H. B. (1995). Initial versus subsequent screening mammography: Comparison of findings and their prognostic significance. *American Journal of Roentgenology*, 164(5), 1107–1109.

Gajdos, C., Tartter, P. I., Bleiweiss, I. J., Hermann, G., de Csepel, J., Estabrook, A., & Rademaker, A. W. (2002). Mammographic appearance of nonpalpable breast cancer reflects pathologic characteristics. *Annals of Surgery*, 235(2), 246–251.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research*, 51(7), 771–781.

Graham, N. & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: A comparison of single-channel and multiple-channels models. *Vision Research*, 11(3), 251–259.

Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. London: John Wiley and Sons Inc.

Guo, Y., Sivaramakrishna, R., Lu, C.-C., Suri, J. S., & Laxminarayan, S. (2006). Breast image registration techniques: A survey. *Medical & Biological Engineering & Computing*, 44(1-2), 15–26.

Hadjiiski, L., Chan, H. P., Sahiner, B., Petrick, N., & Helvie, M. A. (2001). Automated registration of breast lesions in temporal pairs of mammograms for interval change analysis–local affine transformation for improved localization. *Medical Physics*, 28(6), 1070–1079.

Hakama, M., Coleman, M. P., Alexe, D.-M., & Auvinen, A. (2008). Cancer screening: Evidence and practice in Europe. *European Journal of Cancer*, 44(10), 1404–1413.

Harvey, J. & March, D. E. (2013). *Making the Diagnosis: A Practical Guide to Breast Imaging.* Philadelphia: Elsevier Saunders.

Harvey, J. A., Nicholson, B. T., & Cohen, M. A. (2008). Finding early invasive breast cancers: A practical approach. *Radiology*, 248(1), 61–76.

Hasegawa, A., Neemuchwala, H., Tsunoda-Shimizu, H., Honda, S., Shimura, K., Sato, M., Koyama, T., Kikuchi, M., & Hiramatsu, S. (2008). *Digital Mammography: 9th International Workshop, IWDM 2008 Tucson, AZ, USA, July 20-23, 2008, Proceedings*, chapter A tool for temporal comparison of mammograms: Image toggling and dense-tissue-preserving registration, (pp. 447–454). Springer-Verlag: Berlin.

Haygood, T. M. & Dogan, B. E. (2013). *Digital mammography: A practical approach*, chapter Interpretation of digital screening mammography, (pp. 46–61). Cambridge University Press: Cambridge. reference to hanging protocol on p.50.

He, X. & Park, S. (2013). Model observers in medical imaging research. *Theranostics*, 3(10), 774–786.

Health and Social Care Information Centre (2014). *Breast Screening Programme, England. Statistics for 2012-13 (Version 1.0)*. Technical report.

Hecht, S. (1945). *Science in Progress*, chapter Energy and Vision, (pp. 75–97). iv. Yale University Press: New Haven.

Hecht, S., Shlaer, S., & Pirenne, M. H. (1942). Energy, quanta and vision. *The Journal of General Physiology*, 25, 819–840.

Heeger, D. J. (1992). Normalisation of cell responses in cat striate cortex. *Visual Neuroscience*, 9, 181–197.

Heine, J. J. & Malhotra, P. (2002). Mammographic tissue, breast cancer risk, serial image analysis, and digital mammography: Part 1. Tissue and related risk factors. *Academic Radiology*, 9, 298–316.

Hess, R. F. (2011). *Adler's physiology of the eye*, chapter Early processing of spatial form, (pp. 613–626). Elsevier: London, 11th edition.

Hess, R. F. & Snowden, R. J. (1992). Temporal properties of human visual filters: Number, shapes and spatial covariation. *Vision Research*, 32(1), 47–59.

Honda, S., Tsunoda, H., Fukuda, W., & Saida, Y. (2014). A novel image toggle tool for comparison of serial mammograms: Automatic density normalization and alignment-development of the tool and initial experience. *Japanese Journal of Radiology*, 32(12), 725–731.

Huang, K.-C., Lin, R.-T., & Wu, C.-F. (2011). Effects of flicker rate, complexity, and color combinations of chinese characters and backgrounds on visual search performance with varying flicker types. *Perceptual and Motor Skills*, 113(1), 201–214.

Huang, L. & Pashler, H. (2002). Symmetry detection and visual attention: A "binary-map" hypothesis. *Vision Research*, 42(11), 1421–1430.

Hubel, D. H. (1995). *Eye, brain, and vision*. New York: Scientific American Library.

Hyvärinen, A., Hurri, J., & Hoyer, P. O. (2009). *Natural image statistics: A probabilistic approach to early computational vision*. New York: Springer-Verlag.

Irwin, D. E. (1991). Information integration across saccadic eye-movements. *Cognitive Psychology*, 23, 420–456.

Jenkins, B. (1982). Redundancy in the perception of bilateral symmetry in dot textures. *Perception & Psychophysics*, 32, 171–177.

Johnson, J. P., Lubin, J., Nafziger, J. S., & Chakraborty, D. P. (2002). Visual discrimination modeling of lesion detectability. *Proc. SPIE 4686, Medical Imaging 2002: Image Perception, Observer Performance, and Technology Assessment*, 248.

Judy, P. F. (1996). Detection of clusters of simulated calcifications in lumpy noise backgrounds. *Proc. SPIE 2712, Medical Imaging*, 39, 39–46.

Judy, P. F., Swensson, R. G., & Szulc, M. (1981). Lesion detection and signal-to-noise ratio in ct images. *Medical Physics*, 8(1), 13–23.

Julesz, B. (1971). *Foundations of Cyclopean perception*. Chicago: University of Chicago Press.

Kay, S. M. (1998). *Fundamentals of statistical signal processing, volume II: Detection theory*. New Jersey: Prentice Hall.

Kelly, D. (1972). Adaptation effects on spatio-temporal sine-wave thresholds. *Vision Research*, 12(1), 89–101.

Kelly, D. H. (1979). Motion and vision. II. stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America*, 69(10).

Kelly, D. H. (1984). Retinal inhomogeneity. I. Spatiotemporal contrast sensitivity. *Journal of the Optical Society of America A, Optics and Image Science*, 1(1), 107–113.

Kersten, D. (1987). Statistical efficiency for the detection of visual noise. *Vision Research*, 27, 1029–1040.

Kersten, D. & Mamassian, P. (2009). *Encyclopedia of Neuroscience*, volume 5, chapter Ideal Observer Theory, (pp. 89–95). Academic Press: Oxford.

Kierkels, J. J. M., Veldkamp, W. J. H., Bouwman, R. W., & van Engen, R. E. (2012). Power-law, beta, and (slight) chaos in automated mammography breast structure characterization. In A. A. Maidment, P. R. Bakic, & S. Gavenonis (Eds.), *Breast imaging. 11th international workshop, IWDM 2012* (pp. 537–544). Berlin: Springer.

Kingdom, F. A. A. & Prins, N. (2010). *Psychophysics: A practical introduction.* London: Elsevier, 1st edition.

Koning, A. & Wagemans, J. (2009). Detection of symmetry and repetition in one and two objects: Structures versus strategies. *Experimental Psychology*, 56(1), 5–17.

Kopans, D. B. (2007). *Breast Imaging.* Philadelphia: Lippincott Williams & Wilkins., 3rd edition.

Kulikowski, J. J. & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *Journal of Physiology*, 232(1), 149–162.

Kupinski, M. A., Hoppin, J. W., Clarkson, E., & Barrett, H. H. (2003). Ideal-observer computation in medical imaging with use of Markov-chain Monte Carlo techniques. *Journal of the Optical Society of America A Optical Image Science Vision*, 20(3), 430–438.

Lau, T. & Bischof, W. F. (1990). Automated detection of breast tumors using the asymmetry approach. *Computers and Biomedical Research*, 24, 273–295.

Lazerson, H. E. (1984). Blink comparator. *Archives of Ophthalmology*, 102(4), 635.

Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America*, 4(2), 391–404.

Levin, L. A., Nilsson, S. F. E., Hoeve, J. V., & Wu, S. M. (2011). *Adler's physiology of the eye.* London: Elsevier Saunders, 11 edition.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49, 467–477.

Libbrecht, K. (2006). *Ken Libbrecht's field guide to snowflakes.* St Paul, Minnesota: Voyageur Press.

Livshits, G. & Kobyliansky, E. (1991). Fluctuating asymmetry as a possible measure of developmental homeostasis in humans: A review. *Human Biology*, 63, 441–466.

Locher, J. & Wagemans, J. (1993). Effects of element type and spatial grouping on symmetry detection. *Perception*, 22, 565–587.

Lu, K. H. (1965). Harmonic analysis of the human face. *Biometrics*, 21, 491–505.

Lu, Z.-L. & Dosher, B. (2008). Characterizing observers using external noise and observer monoise assessing internal representations with external noise. *Psychological Review*, 115(1), 44–82.

Lu, Z.-L. & Dosher, B. (2014). *Visual psychophysics: From laboratory to theory.* Cambridge, MA: The MIT Press.

Ludwig, C. J. H., Ranson, A., & Gilchrist, I. A. (2008). Oculomotor capture by transient events: A comparison of abrupt onsets, offsets, motion, and flicker. *Journal of Vision*, 8(14), 1–16.

Macmillan, N. A. & Creelman, C. D. (2005). *Detection Theory: A User's Guide.* Lawrence Erlbaum Associates Inc., 2nd edition.

Manahilov, V., Calvert, J., & Simpson, W. A. (2005). Why is second-order vision less efficient than first-order vision? *Vision Research*, 45, 2759–2772.

Mandler, M. B. & Makous, W. (1984). A three channel model of temporal frequency perception. *Vision Research*, 24(12), 1881–1887.

Mayer, B. & Phillips, C. A. (1983). Video blink comparator: U.S. Patent No. US4404590 A. Washington DC: U.S. Patent and Trademark Office.

McKenna, R. J. (1994). The abnormal mammogram radiographic findings, diagnostic options, pathology, and stage of cancer diagnosis. *Cancer*, 74:, 244–255.

Michaelson, J. S., Silverstein, M., Wyatt, J., Moore, G. W. R., Halpern, E., Kopans, D. B., & Hughes, K. (2002). Predicting the survival of patients with breast carcinoma using tumor size. *Cancer*, 95, 713–723.

Møller, A. P. & Thornhill, R. (1998). Bilateral symmetry and sexual selection: A meta-analysis. *The American Naturalist*, 151, 174–192.

Mousa, D. S. A., Brennan, P. C., Ryan, E. A., Lee, W. B., Tan, J., & Mello-Thoms, C. (2014). How mammographic breast density affects radiologists' visual search patterns. *Academic Radiology*, 21(11), 1386–1393.

Myers, K. J. (1985). *Visual perception in correlated noise (models)*. Dissertation-reproduction (electronic), The University of Arizona, Tucson, AZ.

Myers, K. J., Barrett, H. H., Borgstrom, M. C., Patton, D. D., & Seeley, G. W. (1985). Effect of noise correlation on detectability of disk signals in medical imaging. *Journal of the Optical Society of America A*, 2(10), 1752–1759.

Nagaraja, N. S. (1964). Effect of luminance noise on contrast thresholds. *Journal of the Optical Society of America*, 54, 950–955.

National Cancer Institute / Wikimedia Commons (2006). Woman undergoing a mammogram of the right breast. Retrieved from: Breast Cancer UK: https://commons.wikimedia.org/wiki/File:Mammogram.jpg.

Pelli, D. G. (1981). *Effects of visual noise.* PhD thesis, Cambridge University.

Pelli, D. G. (1990). *Vision: Coding and Efficiency*, chapter The quantum efficiency of vision, (pp. 3–24). Cambridge University Press: Cambridge.

Perry, N., Broeders, M., de Wolf, C., Törnberg, S., Holland, R., von Karsa, L., & Puthaar, E., Eds. (2006). *European Guidelines for quality assurance in breast cancer screening and diagnosis.* Luxembourg: European Communities, 4th edition.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *Proceedings of the IRE Professional Group on Information Theory*, 4, 171–212.

Plett, J., Guarini, M., & Irarrazaval, P. (2007). Enhancement of visual perception through dynamic cues: An application to mammograms. *Proceedings of the IEEE International Conference on Image Processing*, II, 441–444,.

Pomplun, M. (1998). *Analysis and models of eye movements in comparative visual search.* Phd thesis, Göttingen: Cuvillier.

Pratt, W. K. (1978). *Digital image processing.* New York: Wiley.

Rainville, S. J. M. & Kingdom, F. A. A. (2002). Scale invariance is driven by stimulus density. *Vision Research*, 42, 351–367.

Reiser, I., Edwards, A., & Nishikawa, R. (2013). Validation of a power-law noise model for simulation of small-scale breast tissue. *Physics in Medicine and Biology*, 58, 6011–6027.

Reiser, I. & Nishikawa, R. M. (2006). Identification of simulated microcalcifications in white noise and mammographic backgrounds. *Medical Physics*, 33, 2905–2911.

Reiser, I. & Nishikawa, R. M. (2010). Task-based assessment of breast tomosynthesis: Effect of acquisition parameters and quantum noise. *Medical Physics*, 37(4), 1591–1600.

Robson, J. G. (1966). Spatial and temporal contrast-sensitivity functions of the visual system. *Journal of the Optical Society of America*, 56(8), 1141–1142.

Roelofs, A. A. J., Karssemeijer, N., Wedekind, N., Beck, C., van Woudenberg, S., Snoeren, P. R., Hendriks, J. H. C. L., del Turco, M. R., Bjurstam, N., Junkermann, H., Beijerinck, D., Séradour, B., & Evertsz, C. J. G. (2007). Importance of comparison of current and prior mammograms in breast cancer screening. *Radiology*, 242(1), 70–77.

Rolland, J. P. & Barrett, H. H. (1992). Effect of random background inhomogeneity on observer detection performance. *J Opt Soc Am A*, 9(5), 649–658.

Sakitt, B. (1972). Counting every quantum. *Journal of Physiology*, 223, 131–150.

Samei, E., Eyler, W., & Baron, L. (2000). *Handbook of medical imaging: Physics and psychophysics*, volume 2, chapter Effects of anatomical structure on signal detection, (pp. 655–682). SPIE Press: Washington.

Samei, E., Flynn, M. J., Eyler, W. R., & Peterson, E. (1998). Effect of local background anatomical patterns on the detection of subtle lung nodules in chest radiographs. *Proc. SPIE 3340, Medical Imaging*, 3340, 44–54.

Sanders, M. S. & McCormick, E. J. (1992). *Human factors in engineering and design*. London: McGraw-Hill Inc, 7th edition.

Sanjay-Gopal, S., Chan, H.-P., Petrick, N., Wilson, T. E., Sahiner, B., Helvie, M. A., & Goodsitt, M. M. (1998). Regional mammogram registration technique for automated analysis of interval changes of breast lesions. *Proceedings SPIE, 3338, Medical Imaging 1998: Image Processing*, 3338, 118.

Schade, O. H. (1956). Optical and photoelectric analog of the eye. *Journal of the Optical Society of America*, 46(9), 721–739.

Scheib, J. E., Gangestad, S. W., & Thornhill, R. (1999). Facial attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society Biological Sciences*, 266, 1913–1917.

Scutt, D., Lancaster, G. A., & Manning, J. T. (2006). Breast asymmetry and predisposition to breast cancer. *Breast Cancer Research*, 8(2), 1–7.

Shaw De Paredes, E. (2007). *Atlas of Mammography*. Philadelphia: Lippincott Williams & Wilkins, 3rd edition.

Sickles, E. A. (2007). The spectrum of breast asymmetries: Imaging features, work-up, management. *Radiologic Clinics of North America*, 45, 765–771.

Sickles, E. A., D'Orsi, C. J., & Bassett, L. W. (2013). *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*, chapter ACR BI-RADS® Mammography. American College of Radiology: Reston, VA.

Simoncelli, E. P. & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24, 1193–1216.

Simpson, W. A., Falkenberg, H. K., & Manahilov, V. (2003). Sampling efficiency and internal noise for motion detection, discrimination, and summation. *Vision Research*, 43, 2125–2132.

Spalek, T. M., Kawahara, J., & Lollo, V. D. (2009). Flicker is a primitive visual attribute in visual search. *Canadian Journal of Experimental Psychology*, 63(4), 319–322.

Stork, D. G. & Falk, D. S. (1987). Temporal impulse responses from flicker sensitivities. *Journal of the Optical Society of America A, Optics and Image Science*, 4, 1130–1135.

Sumkin, J. H., Holbert, B. L., Herrmann, J. S., Hakim, C. A., Ganott, M. A., Poller, W. R., Shah, R., Hardesty, L. A., & Gur, D. (2003). Optimal reference mammography: A comparison of mammograms obtained 1 and 2 years before the present examination. *American Journal of Roentgenology*, 180(2), 343–346.

Suryanarayanan, S., Karellas, A., Vedantham, S., Waldrop, S. M., & D'Orsi, C. J. (2005). Detection of simulated lesions on data-compressed digital mammograms. *Radiology*, 236(1), 31–36.

Swets, J. A. (1961). Is there a sensory threshold? *Science*, 134, 168–177.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, 68(5), 301–340.

Tabár, L., Duffy, S. W., Vitak, B., Chen, H. H., & Prevost, T. C. (1999). The natural history of breast carcinoma: What have we learned from screening? *Cancer*, 86(3), 449–462.

Tahmoush, D. & Samet, H. (2006). Using image similarity and asymmetry to detect breast cancer. In J. M. Reinhardt & J. P. W. Pluim (Eds.), *Proceedings. of SPIE, Medical Imaging 2006: Image Processing*, volume 6144 (pp. 61441S1–61441S7).: SPIE.

Tanner, W. P. & Birdsall, T. G. (1958). Definitions of d' and eta as psychophysical measures. *Journal of the Acoustical Society of America*, 30, 922–928.

Tanner, W. P. & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61(6), 401–409.

Tapiovaara, M. (1990). Ideal observer and absolute efficiency of detecting mirror symmetry in random images. *Journal of the Optical Society of America A*, 7, 2245–2253.

Thurfjell, M. G., Vitak, B., Azavedo, E., Svane, G., & Thurfjell, E. (2000). Effect on sensitivity and specificity of mammography screening with or without comparison of old mammograms. *Acta Radiologica*, 41(1), 52–56.

Timp, S., van Engeland, S., & Karssemeijer, N. (2005). A regional registration method to find corresponding mass lesions in temporal mammogram pairs. *Medical Physics*, 32(8), 2629–2638.

Tjan, B. S. & Liu, Z. (2005). Symmetry impedes symmetry discrimination. *Journal of Vision*, 5(10), 888–900.

Tolhurst, D. J. (1973). Separate channels for the analysis of the shape and the movement of a moving visual stimulus. *Journal of Physiology*, 231(3), 385–402.

Tolhurst, D. J. & Tadmor, Y. (2000). Discrimination of spectrally blended natural images: Optimisation of the human visual system for encoding natural images. *Perception*, 29(9), 1087–1100.

Tomkins, J. L. & Kotiaho, J. S. (2002). Fluctuating asymmetry. In: eLS.. John Wiley & Sons Ltd.

Tovée, M. J., Tasker, K., & Benson, P. J. (2000). Is symmetry a visual cue to attractiveness in the human female body? *Evolution and Human Behavior*, 21, 191–200.

Treder, M. S. (2010). Behind the looking-glass: A review on human symmetry perception. *Symmetry*, 2(3), 1510–1543.

Treisman, A. & Patterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception & Performance*, 10, 12–31.

Tyler, C. W. (2002). *Human symmetry perception and its computational analysis*. London: Taylor & Francis Ebooks.

van Engeland, S., Snoeren, P. R., Karssemeijer, N., & Hendriks, J. H. C. L. (2003). Optimized perception of lesion growth in mammograms using digital display. *Proceedings of the SPIE, Medical Imaging: Image Perception, Observer Performance and Technology Assessment*, 5304, 25–30.

Va'rkonyi, P. I. & Domokos, G. (2006). Symmetry, optima and bifurcations in structural design. *Nonlinear Dynamics*, 43, 47?58.

Veldkamp, W. J. H., Thijssen, M. A. O., & Karssemeijer, N. (2003). The value of scatter removal by a grid in full field digital mammography. *Medical Physics*, 30(7), 1712–1718.

Venkatesan, A., Chu, P., Kerlikowske, K., Sickles, E. A., & Smith-Bindman, R. (2009). Positive predictive value of specific mammographic findings according to reader and patient variables. *Radiology*, 250(3), 648–657.

Wade, T. J. (2010). The relationships between symmetry and attractiveness and mating relevant decisions and behavior: A review. *Symmetry*, 2, 1081–1098.

Wagemans, J. (1995). Detection of visual symmetries. *Spatial Vision*, 9, 9–32.

Wagner, R. F., Beiden, S. V., Campbell, G., Metz, C. E., & Sacks, W. M. (2002). Assessment of medical imaging and computer-assist systems: Lessons from recent experience. *Academic Radiology*, 9(11), 1264–1277.

Wandell, B. A. (1995). *Foundations of vision*. Sunderland, MA: Sinauer Associates.

Wang, X., Lederman, D., Tan, J., Wang, X. H., & Zheng, B. (2011). Computerized prediction of risk for developing breast cancer based on bilateral mammographic breast tissue asymmetry. *Medical Engineering & Physics*, 33(8), 934–942.

Watson, A. B. (1986). *Temporal sensitivity*, volume Handbook of perception and human performance, (pp. 6–1–6–43). Wiley: New York.

Watson, A. B., Borthwick, R., & Taylor, M. (1997). Image quality and entropy masking. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human Vision and Electronic Imaging II*, volume 3016 of *SPIE Proceedings Perceptual Models for Image Processing* (pp. 358–371).

Watson, A. B. & Nachmias, J. (1977). Patterns of temporal interaction in the detection of gratings. *Vision Research*, 17(8), 893–902.

Watson, A. B. & Robson, J. G. (1981). Discrimination at threshold: Labelled detectors in human vision. *Vision Research*, 21(7), 1115–1122.

Wenderoth, P. (1996). The effects of dot pattern parameters and constraints on the relative salience of vertical bilateral symmetry. *Vision Research*, 36(15), 2311–2320.

White, K., Berbaum, K., & Smith, W. L. (1994). The role of previous radiographs and reports in the interpretation of current radiographs. *Investigative Radiology*, 29(3), 263–265.

Whittle, P. (1994). *Lightness, brightness, and transparency*, chapter The psychophysics of contrast brightness, (pp. 35–110). Lawrence Erlbaum Associates: Hillsdale, NJ.

Wickens, T. D. (2001). *Elementary signal detection theory*. Oxford: Oxford University Press.

Wilkinson, F., Wilson, H. R., & Ellemberg, D. (1997). Lateral interactions in peripherally viewed texture arrays. *Journal of the Optical Society of America A, Optics and Image Science*, 14(9), 2057–2068.

Wilson, H. R. & Wilkinson, F. (2004). *The visual neurosciences: volume 2*, chapter Spatial channels in vision and spatial pooling, (pp. 1060–1068). The MIT Press: London.

World Health Organisation (WHO) (2002). *National cancer control programmes: Policies and managerial guidelines*. Geneva: World Health Organisation, 2nd edition.

Wyawahare, M. V., Patil, P. M., & Abhyankar, H. K. (2009). Image registration techniques: An overview. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 2, 11–26.

Youlden, D. R., Cramb, S. M., Dunn, N. A., Muller, J. M., Pyke, C. M., & Baade, P. D. (2012). The descriptive epidemiology of female breast cancer: An international comparison of screening, incidence, survival and mortality. *Cancer Epidemiology*, 36, 237–248.

Zanca, F., Ongeval, C. V., Jacobs, J., Poyry, P., Marchal, G., & Bosmans, H. (2007). Evaluation of the global effect of anatomical background on microcalcifications detection. *Proc. SPIE 6515, Medical Imaging*, 65150H, 651556–1–651556–12.

Zheng, B., Sumkin, J. H., Zuley, M. L., Wang, X., Klym, A. H., & Gur, D. (2012). Bilateral mammographic density asymmetry and breast cancer risk: A preliminary assessment. *European Journal of Radiology*, 81, 3222–3228.

Zheng, B., Tan, M., Ramalingam, P., & Gur, D. (2014). Association between computed tissue density asymmetry in bilateral mammograms and near-term breast cancer risk. *The Breast Journal*, 20, 249–257.

Zonderland, H. & Smithuis, R. (2013). Bi-rads for mammography and ultrasound 2013 updated version.