Classification: Social Sciences

A half second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast.

Karla K. Evans, Ph.D. *, Tamara Miner Haygood, Ph.D., M. D. **, Julie Cooper, M.D. ****, Anne-Marie Culpan, Ph.D., D.C.R.(R) *****, Jeremy M. Wolfe, Ph.D. ***

* Department of Psychology, University of York, Heslington, York, YO10 5DD, UK

** Department of Diagnostic Radiology, U. T. MD Anderson Cancer Center, Houston, TX, 77030, U.S.A.

*** Harvard Medical School & Department of Diagnostic Radiology, Brigham and Women's Hospital, Boston, MA, 02139-4170, U.S.A.

**** Department of Radiology, York Teaching Hospital, York, YO31 8HE, UK

***** Division of Biomedical Imaging, University of Leeds, Leeds, LS2 9JT, UK

Corresponding author:

Karla K. Evans
Department of Psychology
The University of York
York, YO10 5DD, North Yorkshire
UK
Email: karla.evans@york.ac.uk
Tel. +44-01904-32-4601
Fax. +44-01904-32-3190

Keywords: gist processing, medical image perception, attention, mammography

Abstract

Humans are very adept at extracting the "gist" of a scene in a fraction of a second. We have found that radiologists can discriminate normal from abnormal mammograms at above chance levels after a half second viewing (d’~1) but are at chance in localizing the abnormality. This pattern of results suggests that they are detecting a global signal of abnormality. What are the stimulus properties that might support this ability?

We investigated the nature of the “gist” signal in four experiments by asking radiologists to make detection and localization responses about briefly presented mammograms in which the spatial frequency, symmetry and/or size of the images was manipulated. We show that the signal is stronger in the higher spatial frequencies. Performance does not depend on detection of breaks in the normal symmetry of left and right breasts. Moreover, above chance classification is possible using images from the normal breast of a patient with overt signs of cancer only in the other breast. Some signal is present in the portions of the parenchyma (breast tissue) that do not contain a lesion or that are in the contralateral breast. This signal does not appear to be a simple assessment of breast density but rather the detection of the abnormal gist may be based on a widely-distributed image statistic, learned by experts. The finding that a global signal, related to disease, can be detected in parenchyma that does not contain a lesion has implications for improving breast cancer detection.

Significance Statement

Discovering characteristics of a signal that indicates to medical experts the presence of cancer in a non-invasive screening technique in a blink of an eye has implications for improving cancer detection. Here we report two surprising facts about this signal. First, it is much stronger in the high spatial frequencies (fine detail) than in the low frequencies. Second, it is widely distributed with signal being present well away from the actual visible locus of disease even in the breast

contralateral to visible signs of disease. Though this signal is not, in itself, definitive, it has the potential to be used in automated aids to medical screening and incorporated into training protocols for medical experts, speeding up and improving cancer detection.

Introduction

Rapid extraction of scene "gist" (1-4) is a very useful aspect of routine visual perception that allows us to allocate our time and attention intelligently when confronted with new visual information (Can I find food here? Is there danger here?). The signals that we extract upon our first glimpse of a scene are imperfect but not random. Experts often anecdotally report gist-like experiences with complex images in their domain of expertise. For instance, we have shown that radiologists can distinguish normal from abnormal mammograms at above chance levels in as little as a quarter of a second while non-experts cannot (5). The gist of abnormality appears to be a global signal. Radiologists can detect it but cannot even crudely localize the abnormality under these conditions.

Detecting the gist of breast cancer might be more than a curiosity, if that signal could be used to improve performance in breast cancer screening. Screening mammography can reduce mortality through early diagnosis of disease (6). Breast cancer is the most prevalent cancer in women and is the second leading cause of cancer deaths in women (7). In North America, screening mammography has a false negative rate of 20-30% (8,9) and a recall rate of about 10% (10). With a disease prevalence of about 0.3% (11), the vast majority of those recalled will not have cancer. Thus, there is significant room for improvement.

It has been argued for many years that an initial, global processing step is an important component in expert medical image perception that might constrain or filter subsequent search (12-15) with the two most prominent models (16,17) each placing great emphasis on experts' ability to process and evaluate information from large regions of an image (18). These models are broadly consistent with two-stage models of visual search (19,20), developed in the basic vision literature that propose that there is a limited set of features that can be used to guide attention and subsequent serial stage that allows for 'binding' of features to permit identification of objects. Global processing of scene gist is a component of a recent modification of this class of model (21). This formulation proposes there is a *selective* pathway that can be used to recognize one (or a very few) objects at a time. Access to this limited-capacity process is controlled by attention and the deployment of attention is guided by the basic features, mentioned above. There is also a *non-selective* pathway, capable of rapid extraction of "global image statistics" like the average orientation of a set of line segments or the average size of objects (22-24). Perhaps more interestingly, the distribution of basic features, the "spatial envelope" (25,26), contains information that allows for semantic categorization of scenes (e.g. natural vs. urban) without the need to recognize specific objects in the scene.

It is important not to oversell the capabilities of the non-selective pathway. It is engaged in global processing and cannot reliably recognize specific objects. Moreover, the discriminations made on the basis of a first glimpse, while not random, are typically far from perfect. Returning to mammography, Evans et al. (5) found that, while experts could classify mammograms as normal or abnormal at above chance levels, they were at chance in their ability to localize abnormalities. Nevertheless, mammograms appear to contain a signal indicating abnormality. This profile of image statistics or global properties might guide attention or, at least, might alert the radiologist to the possible presence of an abnormality in a mammogram.

In this paper, we investigate the nature of this global signal in the hope that the signal could be better exploited by radiologists or used by designers of computer-aided detection systems to improve breast cancer screening. Our results show that the signal is concentrated in the high spatial frequencies of the image. It is not based on symmetry between two breasts or density of the breasts. Finally, the signal is detectable in breast tissue away from the location of the actual abnormality, including in the contralateral breast. In each of four experiments, we presented experienced radiologists with unilateral or bilateral mammograms (craniocaudal (CC) or mediolateral oblique (MLO) views of both breasts) or sections of mammograms for 500 msec (allowing for, perhaps, 2 volitional fixations). The stimuli were followed by a mask (a white outline of the breasts). Observers rated each stimulus on a scale from 0 (certainly recall this patient) to 100 (certainly normal) (**Figure 1**). If the stimulus was a full breast or pair of breast images, observers were asked to localize the abnormality on an outline of that breast image. We also obtained density ratings from other radiologists for the mammogram stimulus set used in the experiments (Full methods are presented following the Results and Discussion sections).

Results

Experiment 1 asked if the abnormality signal was based on a disruption in the usual bilateral symmetry of the breasts. Studies have noted that asymmetry can be a strong indicator for developing breast cancer (27, 28). Indeed, research has suggested that bilateral mammographic density asymmetry could be a significantly stronger risk factor for breast cancer development in the near-term than either woman's age or mean mammographic density (29).

We measured observers' ratings of abnormality to three types of images: 1) Baseline - both breasts from the same woman, 2) Asymmetry 1 – breast images from two different women. On positive/abnormal trials, one breast image was abnormal while the other was a normal image from another woman. 3) Asymmetry 2 - breasts are from two different women. On positive trials, one breast image was abnormal with a lesion while the other image came from the breast contralateral to a lesion in another woman (**Figure 2**). D', the signal detection measure of performance, is calculated by comparing ratings of the abnormal condition to the ratings of the otherwise equivalent normal condition. When both breasts came from the same woman, expert radiologists could reliably exceed chance performance (avg. d' = 1.14, t(13) = 8.69, p< 0.0001). When the two breast images came from two different women, radiologists could still perform the task (avg. d'=.66, t(13) = 6.28, p< 0.0001), though their performance was significantly worse than when both breasts were from the same woman (planned comparison, t(13) = 7.03, p=0.018). When the abnormal case consists of one breast with an abnormality and the other breast was the breast contralateral to the lesion from a different woman, again, radiologists could do the task (avg. d'=.40, t(13) = 3.02, p< .00097) but their performance was weaker than the performance in the condition where both breasts were from the same woman (p=0.054). Performance did not differ significantly between the two asymmetric conditions (p>0.05). We can conclude from these results that symmetry may be part of what allows an expert to distinguish a normal from abnormal case in a glance, but it is not required since there is above chance performance in the artificial, asymmetric conditions.

Though participants could detect the presence of abnormality, they could not localize that abnormality when it was present (see Figure S1). Localization performance was not significantly different than chance. Localization was best for the baseline condition (21%), but still not above chance performance (t(13)=1.38, p=0.196). In addition, as shown in Evans et al. (5), localization performance did not improve as the confidence rating increased.

Is the signal of abnormality simply breast density, with dense breasts rated as more abnormal? In the baseline condition, d' was significantly better (t(13)=6.93, p<0.0001) than the d' derived from density estimates made by other radiologists. In the Asymmetry conditions, the observed d' was not significantly better than d' derived from density rating (t(13)=1.84, p=0.089; t(13)=0.48, p=0.647). However, if observers were basing their abnormality ratings on an assessment of density, one would expect that the gist and density ratings would be correlated, which they are not (r=0.02). One might also expect a difference in density ratings between normal and abnormal images. However, there is no reliable difference in this image set. A one-way ANOVA on density rating revealed no effect of image type (F(4, 115) = 1.55, p = 0.19) while a one-way ANOVA revealed a large effect of image type on abnormality rating (F(4, 115) = 18.5, p< 0.0001). Thus while the magnitude of the effect in the asymmetrical cases is similar to what could be obtained from a quick assessment of density, there is no evidence that density is the signal that was being used by our observers. Absence of evidence is not proof and it might be that a more statistically powerful experiment might show a relationship of perceived density and the 'gist' of abnormality (e.g. an experiment with density and abnormality ratings made by the same observers). A different, perhaps simpler, way to test the symmetry question and to revisit the density question is to present radiologists with only brief presentation of a *single* breast image at one time, rather than with a paired viewing of the left and right breasts. That is the purpose of Experiment 2.

Experiment 2: Participants rated the appearance of single breast images. In addition to determining if observers can discriminate between normal and abnormal images in the absence of any possible symmetry signal, testing on single breast mammograms made it possible to assess whether the breast contralateral to an abnormal breast could be discriminated from breasts from negative cases. The left panel of Figure 3 shows that observers were able to distinguish between images of single normal and abnormal breasts (d'=1.16; (t(14)=8.35 p<0.0001). What is more, as shown in the right panel of Figure 3, their performance remained above chance when distinguishing normal from an image contralateral to the breast with a lesion (d'= 0.59; (t(14)=8.35 p<0.0001) though performance in that condition is significantly worse than performance with abnormal images (paired t(14)=5.8, p=0.00004). As in Experiment 1, the weaker performance, obtained with images contralateral to the lesion, was of a magnitude similar to what would be obtained if observers based their ratings on breast density. However, as in Experiment 1, there is no evidence that the radiologists were using that density signal. As before, the relationship of density ratings to abnormality ratings was weak or non-existent (r= 0.06 of ratings and density across images and r=-0.02 for the contralateral images alone). Further, there was no effect of the objective type of image (normal vs. abnormal) on density ratings (F(4, 115) = 0.71, p= 0.49) but there was a large effect of image type on abnormality ratings (F (4, 115) = 46.06, p< 0.0001). As in Experiment 1, the average localization performance of observers for images with the abnormality in a single breast was not significantly above chance level (t(14)= .91, p=0.378).

Experiment 3: Any texture can be decomposed into a set of sinusoidal gratings of different spatial frequencies, amplitudes, orientations, and phases. Experiment 3 examined the spatial frequency composition of the signal of abnormality. Radiologists viewed normal and abnormal, bilateral mammograms in each of three counterbalanced conditions: unfiltered full images equivalent to the baseline condition of Experiment 1, high-pass filtered images and low-pass filtered images shown as in Figure 4a. There was a significant difference between conditions (F(2,16)=52.35, p<0.0001). Specifically, the signal for interpreting mammography in 500 msec resides far more strongly in the high spatial frequencies, suggesting that the information is present in some aspect of the finer detail of the parenchymal texture (**Figure 4b**). High-pass performance was reliably greater than chance (d'=0.97; t(8) = 8.05, p <0.0001) and better than performance on low-pass images (low-pass d'=0.26, paired t-test t(8) = 5.30, p=0.002). High-pass performance did not differ from performance with unfiltered images (d': 0.97 vs. 1.06, t(8)=0.61, p = 0.56).

Again, the rated density of the images cannot explain radiologists' performance in any of the three conditions. The derived d' from the average density rating was d'=0.09, and that is significantly lower than the performance for unfiltered images (d'= 1.06, t(8) = 8.81, p <0.0001), high-pass images (d'=.97, t(8) = 7.43, p <0.0001) or low-pass images (d'= 0.26 (t(8) = 6.00, p <0.0003). None of the correlations of image density and image abnormality rating were significant (all F(1,53) < 2.2, all p > 0.14).

These findings are interesting for at least two reasons. First, if radiologists were simply using density as the signal, one might expect better performance from low spatial frequencies. Second, outside of radiology, the more typical finding in the appreciation of scene gist is that it is the low spatial frequency content that can be appreciated first in a brief flash; not the higher frequencies, though 500 msec would be long enough to appreciate both low and high frequencies in a typical scene gist experiment (30). Since localization performance remained poor across all conditions (best for high-pass filtered images but still not above chance, t(8)=0.86, p=0.414, **Figure S2**), we conclude that it is not a specific detail of the lesion that is supporting the decision but, rather, abnormality is judged based on some aspect of the overall texture that is best visualized in the higher spatial frequencies. Perhaps the signal is related to processes that create indications of disease like spicules that might be enhanced in a high-pass view, but a larger data set would be needed to test such a hypothesis.

Experiment 4:If the signal of abnormality is present throughout the parenchyma as would be predicted if that signal is truly a global signal, then it follows that a signal should be found in isolated regions of the breast that deliberately exclude the lesion. Alternatively, even though radiologists cannot explicitly localize abnormalities after a 500 msec flash, the signal might still arise exclusively from some small portion of the breast rather than being distributed widely. To test that hypothesis, in Experiment 4, we presented 256 x 256 pixel patches of mammograms and asked radiologists to distinguish between normal and three types of potentially abnormal patches: patches containing the lesion, lesion-free patches from the abnormal breast, and lesion-free patches from the breast contralateral to the lesion. Observer's performance differed significantly between the three types of samples (F(2, 20)=109.14, p<0.0001). However, all three types of patches from abnormal cases could be distinguished from normal at above chance levels. This can be seen by noting that virtually all of the individual observer data lies above the main diagonal, chance line in Figure 5. Performance on sections with the lesions was significantly better than patches without the lesion from either the ipsilateral (p<0.0001) or contralateral breast (p<0.0001). Performance on ipsilateral and contralateral patches without a visible lesion did not differ (p=0.473). The density estimates, made by other radiologists for these small patches, produce areas under the ROC curve (AUC) between 0.47 and 0.49, essentially at the 0.5, chance, level. Apparently, there is no signal in the density ratings for these small patches of breast parenchyma. There was no difference between the average density ratings for the different types of sections (F(3, 196) = 0.09, p= 0.97) and the density ratings were not significantly correlated with the abnormality ratings (all r^2 < 0.05, all p > 0.12).

These results provide interesting insight into the signal supporting radiologists' performance in these tasks. Unsurprisingly, when the section includes the lesion, attention will be directed to the lesion and performance is better than if the radiologist is looking at the entire breast with the lesion in an unknown location. Of more interest, there is some signal in sections of parenchyma ipsilateral and contralateral to the lesion. The signal is weak (**Figure 5** conditions B&C), corresponding to d' values of only 0.33-0.40 in the sections that did not include the lesion. However, note that the patches show only about 1/8th of a single breast. If we model the whole breast as consisting of 8 independent samples with d' 0.33 to 0.40, performance for a presentation of the whole breast would yield d' between 0.9 and 1.2. This is comparable to or somewhat higher than the d' for whole breasts in Experiments 1-3. If results from the whole breast are actually worse than would be predicted from small patches, that suggests that the signals, combined across the whole breast are not entirely independent. In any case, the local signal is in principle, strong enough to support the results obtained with whole breasts, when combined across the whole breast.

Discussion

Radiologists report anecdotally that some images seem to be 'bad' when they first appear, before any specific pathology is localized. No one would suggest that diagnosis should be based on these first glimpses. However, there is now a body of research, including the work reported here, that indicates that this sense of the gist of a medical image can be based on a measurable signal (5,12,15). Our goal, in the present paper, has been to investigate the nature of the signal that allows expert observers to classify mammograms as normal or abnormal at above chance levels after a brief exposure. Experiments 1 and 2 undermined the hypothesis that observers were responding to a break in the normal rough symmetry between left and right breasts. In Experiment 1 the symmetry was disrupted and in Experiment 2, observers only viewed a single breast image. In both cases comparing normal and abnormal images, it remains possible to perform the classification task with a d' a bit better than 1.0. While radiologists may use symmetry between two breasts as an important sign in normal mammography, it is not the signal that allows for classification of mammograms after a half second of exposure.

Localization performance was consistently poor, suggesting that classification is based on a global signal, spread across the breast. The first novel finding in this paper is the evidence in Experiment 2 that this signal is present in the breast contralateral to the breast containing the abnormality. Experiment 4 found evidence for the signal in sections of parenchyma that did not contain an abnormality, regardless of whether they came from the ipsilateral or contralateral breast. Performance with these small sections is about what one would expect if the signal were being pooled across the entire image when the entire image is present. This finding may have clinical significance in the light of recent evidence that women with false positive screening mammograms were at an increased risk of developing breast cancer compared to those with true negatives (31). Perhaps, even if localized signs of cancer were not unambiguously visible at the initial screening, radiologists still may have been influenced by the global signal of abnormality that we are studying here.

Experiment 3 provides another interesting finding; that the signal for abnormality is far stronger in a high-pass filtered mammogram than in a low-pass filtered image. Given prior results on recognition of briefly presented images (e.g. the global-local effect: 32, 33), one might have expected some sort of advantage for the coarser information in the low-pass image. Instead, we found the information about abnormality resides in the higher frequencies.

It is worth noting that the ability to detect abnormality at above chance levels is a learned skill of expert radiologists. In previous work (5), we had non-experts attempt the task. They performed at chance levels. It would be interesting to know if general radiologists who read fewer mammograms are able to detect this global signal of abnormality.

A distributed global signal of abnormality in breast cancer might be a useful component in a Computer-Aided Detection (CAD) system (34). The normal goal of a CAD system is to direct the radiologist's attention to specific, suspicious locations. Though these systems perform at a level comparable to that of an expert radiologist, they have not been hugely successful in clinical practice (35), in part because the positive predictive value of any given CAD mark is very low in a mammography screening situation where the prevalence of disease is low. As a result, radiologists tend to dismiss the correct CAD marks when they occur (36). It is possible that the signal that supports classification in the experiments reported here, could be used as an additional piece of information for a CAD system. A CAD mark in the presence of a global abnormality signal might be a more suspicious mark than one in the absence of the signal. The presence of the signal in the breast contralateral to the abnormality also raises an interesting clinical possibility. It may be that the signal is present before the actual lesion appears. If so, it could be used as a warning sign, suggesting greater vigilance much as breast density is used as risk factor today (37). In thinking about any of these possibilities, it is critical to remember that radiologists' ability to detect abnormality in half a second is probabilistic. They perform above chance but far from perfect and far from their performance under normal conditions of reading mammograms. The gist signal might be useful but, by itself, it is nowhere near definitive. In conclusion, there is a global signal that can be measured by asking radiologists to classify mammograms in a fraction of a second. That signal is probably the basis of the initial "holistic" impression of an image that is thought to guide radiologists when they view images in a normal, clinical setting (12, 38). If properly quantified, it could also be a component of automated aids to mammography.

Methods

Participants

All study participants were attending radiologists specializing in breast imaging. Across the four studies we tested 49 radiologists: Experiment 1 - fourteen radiologists (11 female, 3 male; average age 53), average 19 years in practice (range 4 to 34 years) reading, on average, 7,650 cases in the last year (range 6,000 to 10,000). Experiment 2 - Fifteen radiologists (12 female, 3 male; average age 49), average 19 years in practice (range 10 to 35 years), reading on average 7280 cases in the last year (range 3,000 to 15,000). Experiment 3 - nine radiologists (5 female, 4 male, average age 50) average 15 years in practice (range 7 to 39), reading on average 7100 cases in the last year (range 4,000 to 10,000). In Experiment 4 – eleven radiologists (10 female, 1 male; average age 52), average 20 years in practice (range 4 to 34 years) reading on average 7,800 cases in the last year (range 6,000 to 10,000). The radiologists who participated in Experiment 1, 2 & 4 were recruited from five NHS hospital Trusts in Yorkshire and Cambria, United Kingdom. In Experiment 3 radiologists were recruited from U. T. MD Anderson Cancer Center, Houston, Texas, USA. All the participants had normal or corrected-to-normal vision and gave informed consent. The experiments had institutional review board approval from University of York, U. T. MD Anderson Cancer Center and the NHS Hospital Trusts.

Stimuli and Materials

The stimuli used in the four experiments were derived from 120 bilateral full-field digital mammograms. The starting resolution of the two mammograms side by side was 1,980 x 2,294 pixels, These were then downsized to fit on a monitor with a resolution of 1,920 x 1,080. Mammograms were acquired from anonymized cases from Brigham and Women's Hospital, Boston, United States. All the cases included at least 4 images (left and right breast mediolateral oblique (MLO) views and craniocaudal (CC) views. Half of the cases showed cancerous abnormalities while the rest were normal. Abnormal cases were either screen-detected cancers, histologically verified, or mammograms that had been done 1 to 2 years prior to a screen-detected cancer and that had been interpreted as negative but later retroactively determined by a study radiologist to have contained visible abnormalities. The abnormalities demonstrated on mammograms were "subtle" masses and architectural distortions. Lesion subtlety was determined by the study radiologists based on their experience. We did not include calcifications or more obvious cancers as it is of less interest to show that a stimulus like a bright white spot can be detected in less than a second. The average size of the lesions in the test set mammograms was 18 millimeters (range 10 - 48 mm).

Experiments 1 & 3 used all of the 120 bilateral mammograms. For Experiment 3, these original images were Fourier-transformed and two types of filtered images were computed. A low-pass image was created by removing all the information above 2 cycles per degree (at a 57 cm viewing distance) leaving only the low spatial frequencies of the original image. A high-pass imaged was created by removing information that was below 6 cycles per degree leaving only the high spatial frequency information of the original images. This resulted in three sets of images: original intact images, the same images but with only low spatial frequency information present and images with only the high spatial frequency information present.

In Experiment 2, we used 120 unilateral breasts, taken from the bilateral full-field digital mammograms used in Experiment 1. A third of the single mammograms had a confirmed yet subtle abnormality (e.g. mass or architectural distortion), another third were taken from completely normal cases, and the last third were mammograms of breasts that contained no abnormality but that were the breast contralateral to a breast containing an abnormality.

The stimuli used in Experiment 4 consisted of 200 sections taken from the original full-field digital mammograms (including both CC and MLO views). Mammogram sections were cropped to 256 x 256 pixels using Photoshop CS6. A quarter of the sections included a lesion, centered in the patch. There were three types of no lesion sections: a) section taken from the image of an abnormal breast but not containing the lesion, b) section taken from the breast contralateral to a breast containing a lesion and c) section taken from a completely normal case.

Two of the authors (TMH, JC) who are practicing radiologists provided density ratings for each left and right mammographic image for all the images in the stimulus set on a four-point scale (1-fatty, 2-scattered fibroglandular, 3-heterogeneously dense, 4-extremely dense). The density ratings of the two radiologists were significantly correlated for both breasts (Left breast r=0.56, p< .00001; Right breast r= 0.43, p< .00001). Rated density of abnormal cases was slightly higher than for normal (2.80 vs. 2.65, but not significantly, one-tailed t-test t(188)=1.64, p=0.101). If classification of normal vs. abnormal was based on the average density ratings given by the two radiologists, the predicted d' would be 0.26. The density ratings of the two radiologists for the single breast subset of stimuli used in Experiment 2 were also significantly correlated (r=0.36, p<0.0001). The density raters also gave a density rating for the four types of section we used in Experiment 4. A one-way independent ANOVA on the average density rating revealed no significant main effect of type of section (F(1,199)=0.86, p=0.968). Thus there was no significant difference in the density rating of the four types of small section.

The Experiments 1, 2 & 4 were conducted on a Macintosh, MacBook Pro using MATLAB R2012b. All observers viewed the experiment on a 27.5 inch, liquid-crystal color screen with a 1920 x 1080 resolution, a usable intensity range of 2–260 candelas per square meter, a contrast ratio of 188:1 and refresh rate of 144Hz at a viewing distance of 57 cm. Experiment 3 was conducted on a Dell Precision™ M6500 laptop using MATLAB R2012b. The experiment was displayed on a 17" screen at a viewing distance of 57 cm. The display monitor had a resolution of 1920 x 1200 (Dell, Round Rock, Texas.) and a refresh rate of 85Hz.

Procedure

Across the study all four experiments used the same experimental paradigm of brief stimuli presentation. All observers in each experiment viewed the same images with the order randomized across trials. After 3-6 practice trials, depending on the experiment, each trial consisted of the following sequence of events. First, a fixation cross appeared in the center of the screen for 500 msec. This was followed by the brief, 500 msec presentation of a pair of mammograms (Experiment 1 & 3), side by side, a single mammogram (Experiment 2) or a single section (Experiment 4). After the brief presentation, observers saw a white outline of the previously presented breasts (Experiment 1-3) or a white noise mask for another 500 msec (Experiment 4). In Experiments 1-3, even if they did not think the case was abnormal, radiologists were asked to indicate the most likely location of an abnormality with a mouse click on the display screen. Following this, observers were asked to provide a 0-100 rating (where 0 stands for clearly abnormal) scale how likely it was that there was an abnormality. Feedback was provided only for the initial practice trials. All the observers were alone when performing the experiment.

In Experiment 1 participants completed 120 trials across five possible trial types: a) The two breasts were from the same woman: one breast with an abnormality, one normal (20 trials), b) The two breasts were from two different women: one breast with an abnormality, the other normal and from a completely normal case (20 trials), c) The two breasts were from different women; one with an abnormality, the other normal but from an abnormal case (20 trials), d) The two breasts were from the same woman: both breasts normal (30 trials), e) Finally, the two breasts were from different women: both breasts and both cases completely normal (30 trials). Thus, overall, half of the cases were normal, half abnormal. These five types of presentation were used to create the three comparisons described in the results. A comparison of conditions (a) and (d) replicates the previous work on detection of the gist of abnormality (Baseline). Comparisons of conditions (b) and

(c) with condition (e) (Asymmetry 1 & 2) test for the presence of a non-selective, gist signal when a symmetry cue cannot be used.

In Experiment 2, participants completed 120 trials evenly divided between images of three types of breast: Normal, Abnormal, and Contralateral (being the normal breast contralateral to an abnormal breast).

In Experiment 3, participants completed 3 blocks of 120 trials, for a total of 360 experimental trials in which they viewed CC or MLO views of mammograms. In each block, the observers saw only one set of images: the original intact image set, the low spatial frequency image set or the high spatial frequency image set. The viewing order of the blocks was counterbalanced across observers.

In Experiment 4, observers completed 2 blocks of 100 experimental trials each in which they viewed sections of mammograms evenly divided between the four types described above.

Data Analysis

Assessing Detection Performance. The observers in all four experiments gave confidence ratings on a scale from 0 (clearly abnormal) to 100 (clearly normal). For a given rating threshold, scores above that rating can be considered "true negatives", if the stimulus is normal and "miss" or "false negative" errors, if the stimulus is abnormal. Scores bellow the level are deemed "hits" or "true positives", if the case is abnormal and "false alarm" or "false positive" errors if the case is normal. Categorizing responses in this manner for a range of values sweeps out a receiver operating characteristic (ROC) curve. Thus, signal detection measures of d', criterion, and area under the ROC curve (AUC) can be derived. For purposes of calculating d', we used a rating threshold of 50.

The analysis of Experiment 1 is somewhat complicated because there are three types of abnormal case (Trial Types a, b & c) and two types of normal case (Trial Types d & e). For each of the three critical comparisons of normal and abnormal, the hit rate is derived from one of the three abnormal conditions and the false alarm rate is derived from one of the two normal conditions. When the abnormal cases are those in which left and right images were from the same woman (Trial Type a), the false alarm rate is derived from the normal cases in which left and right images are also from one woman (Trial Type d). When the abnormal cases are those in which the left and right images are taken from the mammograms of different women (Trial Types b & c), the normal cases are, likewise, taken from cases in which the left and right images come from different women (Trial Type e).

For Experiment 2 and 4, the average d' and ROC curves were created by taking the false alarm rates from the single breast or sections taken from normal breasts and pairing them with the hit rate from each of the two potentially abnormal conditions in Experiment 2 or three potentially abnormal conditions in Experiment 4.

Calculating Density d'. A value of d' can also be calculated from the average density ratings using the same method as described above for the abnormality rating. Breast density was rated on a 4 point scale from 1= fatty and 4 = extremely dense. For normal images, using a threshold of 2.5, if the density rating was above threshold, that rating would be categorized as a "false positive". If it was below, it was deemed to be a "true negative". For abnormal mammograms, if the density rating was above the 2.5 cut off then it was categorized as a "hit"; if below, it was a "miss". We used values above threshold as the analog of target present (abnormal) response because previous research has found that increased density is associated with higher likelihood of cancer (29).

Assessing Localization Performance. To assess localization performance the observers were asked to click on an outline mask of the breast to indicate where they thought an abnormality was most likely to have been located. Localization performance was measured by determining what percentage of observers' clicks fell into the predetermined regions of interest (ROI) centered on abnormalities. We then calculated the percentage of correctly localized abnormalities in respect to the overall number of abnormalities. Chance levels for localization performance were determined as average percentage of the breast encompassed by the ROI (abnormal region). Different abnormal cases would have larger or smaller ROIs. Averaged across cases with lesions, the ROI area was 18% in Experiment 1; Experiment 2= 6%; Experiment 3=16%. These values, then, represent the chance of hitting an ROI by placing a random mark on the breast outline.

Acknowledgement

This research was supported in part by a grant from the John S. Dunn, Sr., Distinguished Chair in Diagnostic Radiology of U. T. MD Anderson Cancer Center. JMW was supported by NIH, EY017001.

References

1. Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W. (1974). On the information extracted from a glance at a scene. J Exp Psychol, 103(3), 597.
2. Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. Nature, 253(5491), 437-438.
3. Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. Proc Natl Acad Sci USA, 99(14), 9596-9601.
4. Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. Vision Res, 46(11), 1762-1776.
5. Evans, K. K., Georgian-Smith, D., Tambouret, R., Birdwell, R. L., & Wolfe, J. M. (2013). The gist of the abnormal: Above-chance medical decision making in the blink of an eye. Psychon Bull Rev, 20(6), 1170-1175.
6. Smith, R. A., Cokkinides, V., & Eyre, H. J. (2004). American Cancer Society guidelines for the early detection of cancer, 2004. CA, Cancer J Clin, 54(1), 41-52.
7. Siegel, R., Naishadham, D., & Jemal, A. (2013). Cancer statistics, 2013. CA, Cancer J Clin, 63(1), 11-30.
8. Bird, R. E., Wallace, T. W., & Yankaskas, B. C. (1992). Analysis of cancers missed at screening mammography. Radiology, 184(3), 613-617.
9. Majid, A. S., de Paredes, E. S., Doherty, R. D., Sharma, N. R., & Salvador, X. (2003). Missed Breast Carcinoma: Pitfalls and Pearls 1. Radiographics, 23(4), 881-895.
10. Gur, D. et al. (2004). Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. J Natl Cancer Inst, 96(3), 185-190.
11. Lee, C. S., Bhargavan-Chatfield, M., Burnside, E. S., Nagy, P., & Sickles, E. A. (2016). The National Mammography Database: Preliminary Data. AJR Am J Roentgenol, 206(4), 883-890. doi: 10.2214/AJR.15.14312

12. Kundel, H. L., & Nodine, C. F. (1975). Interpreting Chest Radiographs without Visual Search 1. Radiology, 116(3), 527-532.

13. Carmody, D. P., Nodine, C. F., & Kundel, H. L. (1981). Finding lung nodules with and without comparative visual scanning. Percept Psychophys, 29(6), 594-598.

14. Oestmann, J. W. et al. (1988). Lung lesions: correlation between viewing time and detection. Radiology, 166(2), 451-453.

15. Mugglestone, M. D., Gale, A. G., Cowley, H. C., & Wilson, A. R. M. (1995, April). Diagnostic performance on briefly presented mammographic images. In Medical Imaging 1995 (pp. 106-115). International Society for Optics and Photonics.

16. Nodine, C. F., & Kundel, H. L. (1987). The cognitive side of visual search in radiology. Eye Movements: From Psychology to Cognition. North Holland, Elsevier Science, 572-58

17. Swensson, R. G. (1980). A two-stage detection model applied to skilled visual search by radiologists. Percept Psychophys, 27(1), 11-16.

18. Reingold, E. M., & Sheridan, H. (2011). Eye movements and visual expertise in chess and medicine. Oxford handbook on eye movements, 528-550.

19. Treisman, A. (2006). How the deployment of attention determines what we see. Vis cogn, 14(4-8), 411-443.

20. Wolfe, J. M. (2007). Guided Search 4.0: Current Progress with a model of visual search. In W. Gray (Ed.), Integrated Models of Cognitive Systems (pp. 99-119). New York: Oxford.

21. Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. Trends Cogn Sci, 15(2), 77-84.

22. Ariely, D. (2001). Seeing sets: Representation by statistical properties. Psychol Sci, 12(2), 157-162.

23. Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. Vision Res, 43(4), 393-404.

24. Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. Journal of Vision, 11(12), 18-18.

25. Oliva, A. (2005). Gist of the scene. Neurobiology of attention, 696(64), 251-258.

26. Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition Prog Brain Res, 155, 23-36.

27. Scutt, D., Lancaster, G. A., & Manning, J. T. (2006). Breast asymmetry and predisposition to breast cancer. Breast Cancer Res, 8(2), R14.

28. Sun, W. at al. (2014). Prediction of near-term risk of developing breast cancer using computerized features from bilateral mammograms. Computerized Medical Imaging and Graphics, 38(5), 348-357.

29. Zheng, B. et al. (2012). Bilateral mammographic density asymmetry and breast cancer risk: A preliminary assessment. Eur J Radiol, 81(11), 3222-3228.

30. Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. Psychol Sci, 5(4), 195-200.

31. Henderson, L. M., Hubbard, R. A., Sprague, B. L., Zhu, W., & Kerlikowske, K. (2015). Increased Risk of Developing Breast Cancer after a False-Positive Screening Mammogram. Cancer Epidemiol Biomarkers Prev, 24(12), 1882-1889

32. Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. Cognitive psychology, 9(3), 353-383.

33. Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: a critical review. Psychol Bull, 112(1), 24.

34. Gierach, G. L., Li, H., Loud, J. T., Greene, M. H., Chow, C. K., Lan, L., ... & Mai, P. L. (2014). Relationships between computer-extracted mammographic texture pattern features and BRCA1/2 mutation status: a cross-sectional study. Breast Cancer Research, 16(4), 1.

35. Cole, E. B. et al. (2014). Impact of computer-aided detection systems on radiologist accuracy with digital mammography. AJR Am J Roentgenol, 203(4), 909.

36. Nishikawa, R. M., Giger, M. L., Jiang, Y., & Metz, C. E. (2012). Re: Effectiveness of computer-aided detection in community mammography practice. J Natl Cancer Inst, 104(1), 77-77.

37. McCormack, V. A., & dos Santos Silva, I. (2006). Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev, 15(6), 1159-1169.

38. Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: gaze-tracking study 1. Radiology, 242(2), 396-402.

Figure Legends

Figure 1
Experimental procedure for Experiments 1-4. Experiment 2 just showed a unilateral breast image and Experiment 4 used only a piece of the breast image.

Figure 2

ROCs for the three conditions of Experiment 1. Solid colored line – average ROC, light dotted lines – individual observers. Dark dotted line, hypothetical ROC if judgments were based on density ratings. A) Images of two breasts from the same woman. One breast abnormal on the positive trials. d'= 1.14 compared to d'=0.18 derived from the density ratings B) Images from two different women, one image abnormal on the positive trials; the other, always drawn from a negative case. d'= 0.66 compared to d'= 0.47 derived from the density ratings. C) Images from two different women, one image abnormal on the positive trials; the other image was the breast contralateral to a lesion in another woman. d'= 0.40 compared to d'=0.34 derived from the density ratings.

Figure 3

ROC curves for single breast image data. Light dashed lines are individual observer data. The solid line shows the average data and the dark dashed line shows the ROC that can be derived from the density data.

Figure 4

a) Example images used in Experiment 3: A) High-pass filtered, B) low-pass filtered, and C) unfiltered views of a breast stimulus.

b) ROCs for the three conditions of Experiment 3. Solid colored lines – average ROCs, dashed lines – individual observers. A) Baseline, unfiltered/intact images, B) Low pass filtered images, C) – High pass filtered images.

Figure 5

ROCs for the three conditions of Experiment 4. Solid colored lines – average ROCs, dashed lines – individual observers. A) Abnormal section contains lesion, d'=1.47 (99.9% CI 1.20 - 2.12) B) Abnormal section ipsilateral to lesion, d'= 0.33 (99.9% CI 0.17 - 0.49). C) Abnormal section contralateral to lesion, d'=0.40 (99.9% CI 0.21 - 0.58). In all cases, the hit rate is derived from sections taken from a normal case.

Figure 1.

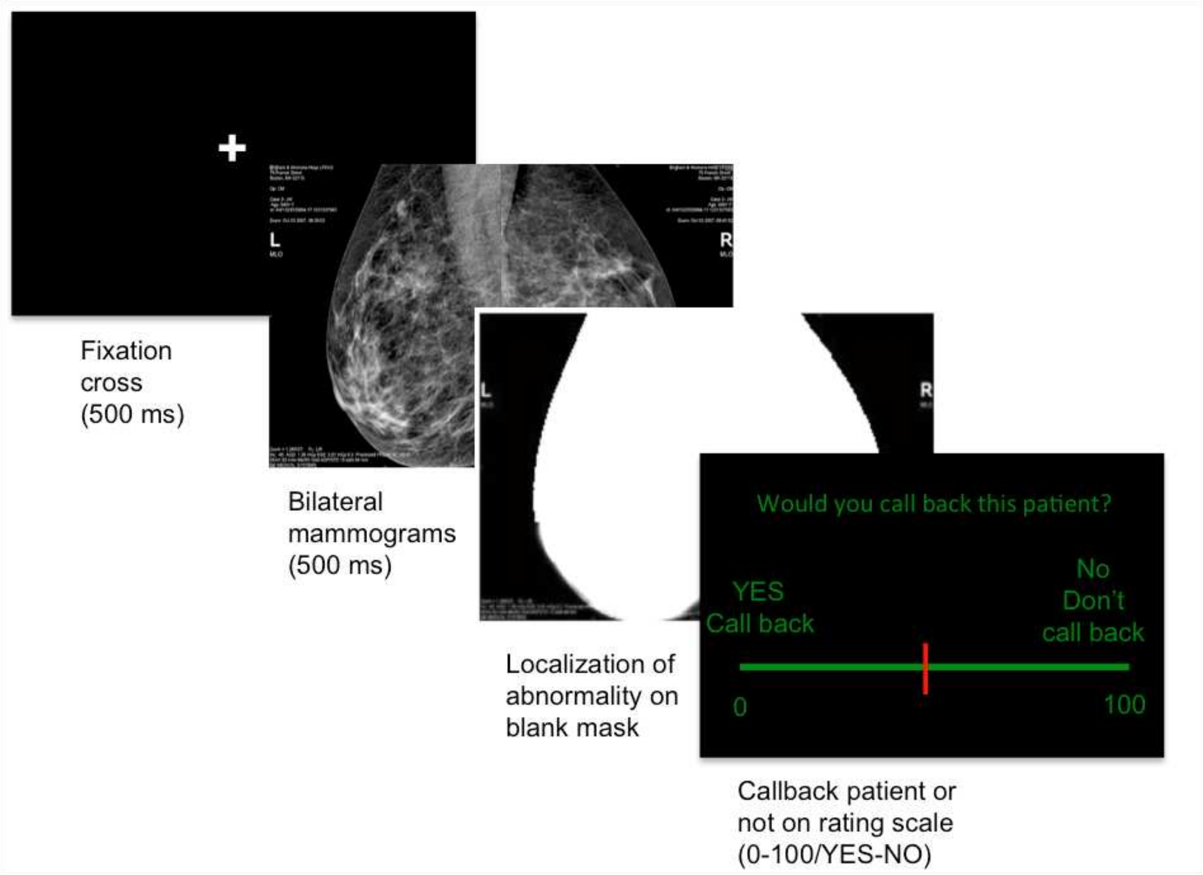

Fixation cross (500 ms)
L
R
Bilateral mammograms (500 ms)
L
R
Localization of abnormality on blank mask
Would you call back this patient?
YES Call back
No Don't call back
0
100
Callback patient or not on rating scale (0-100/YES-NO)

Figure 2

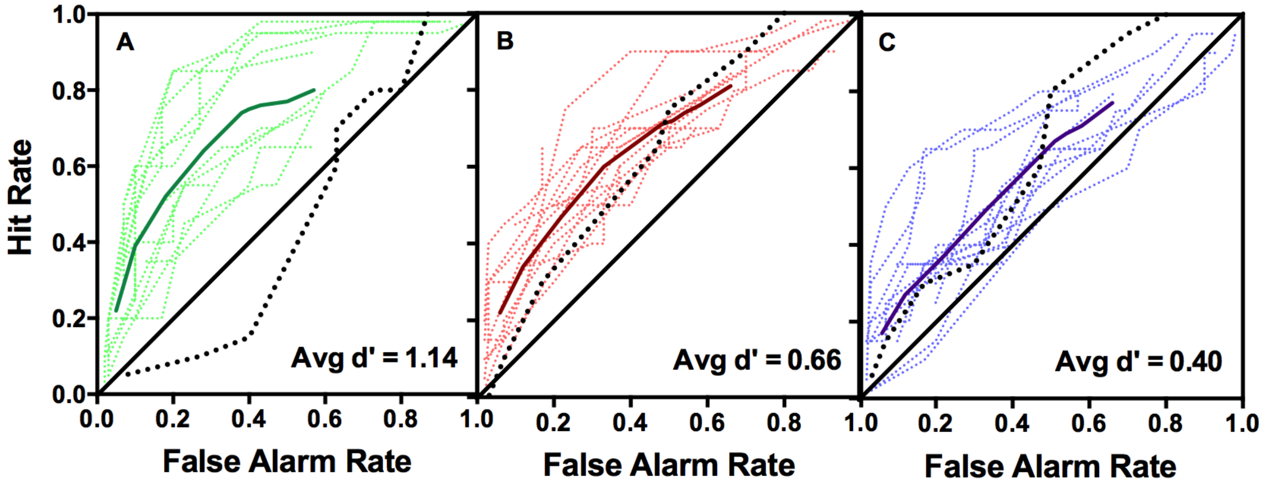

A
B
C
Hit Rate
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
False Alarm Rate
Avg d' = 1.14
Avg d' = 0.66
Avg d' = 0.40

Figure 3

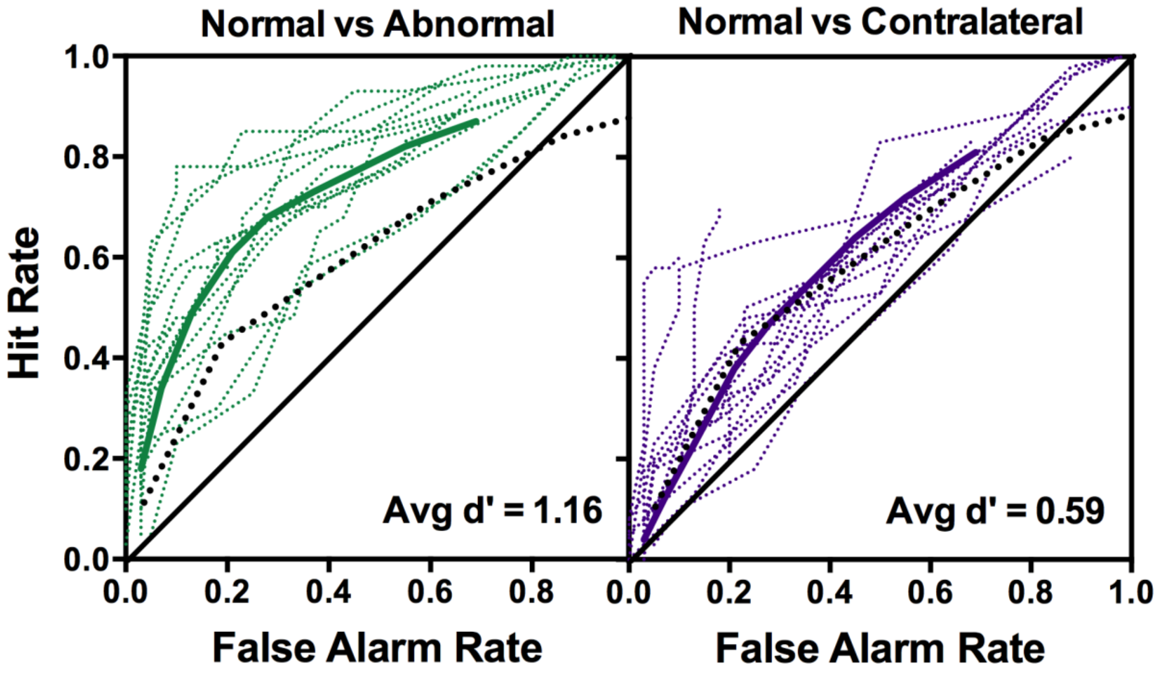
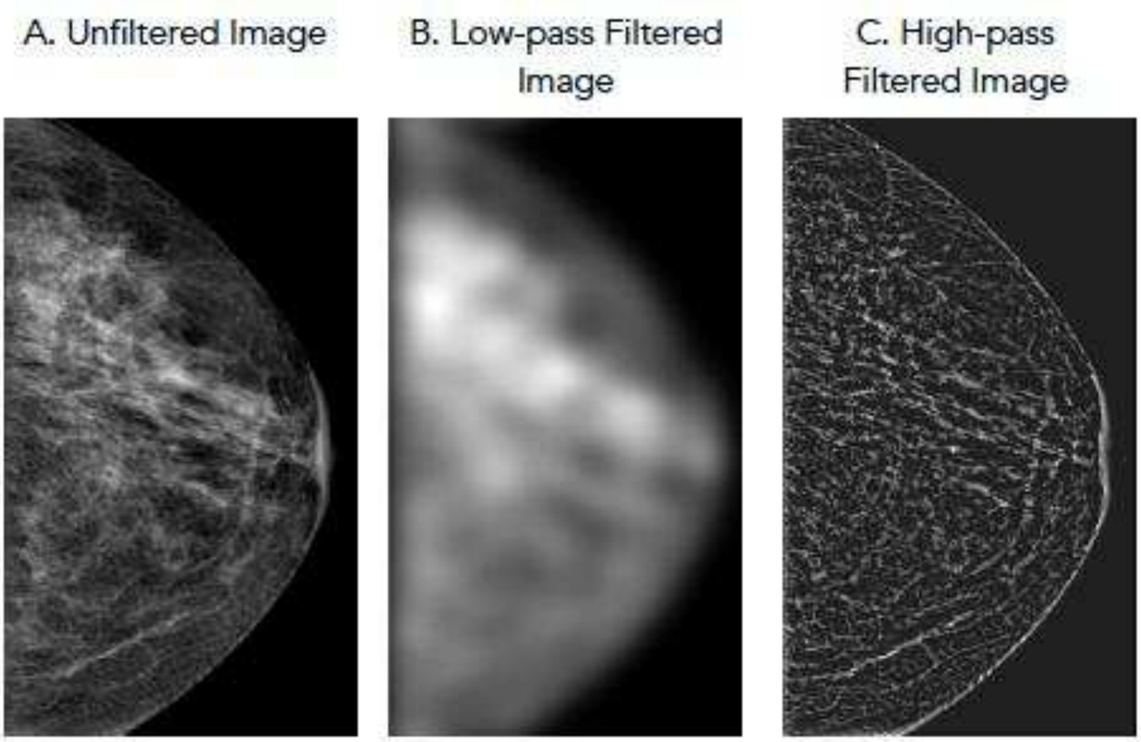

Normal vs Abnormal
Normal vs Contralateral
Hit Rate
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
1.0
False Alarm Rate
False Alarm Rate
Avg d' = 1.16
Avg d' = 0.59

Figure 4 A


A. Unfiltered Image
B. Low-pass Filtered Image
C. High-pass Filtered Image

4 B

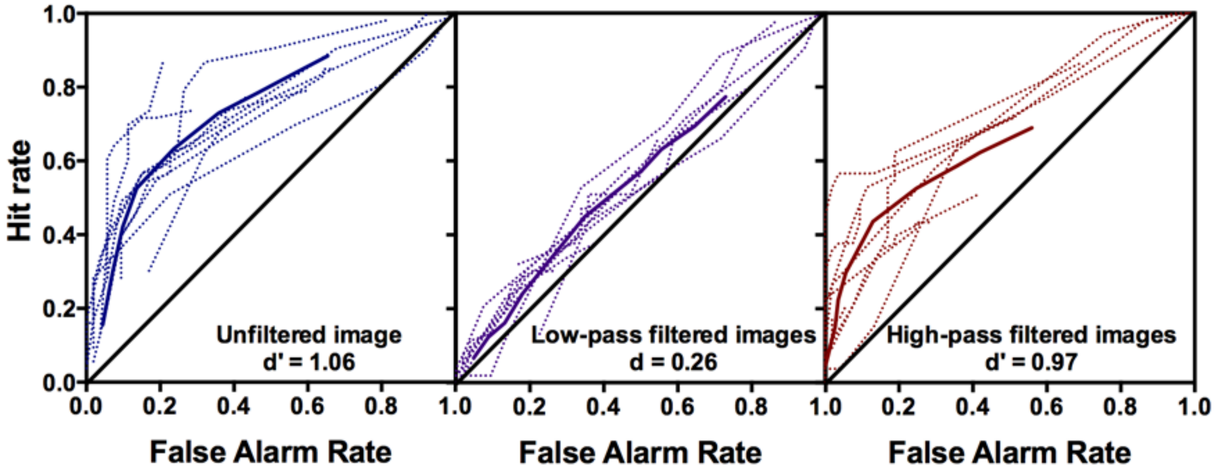
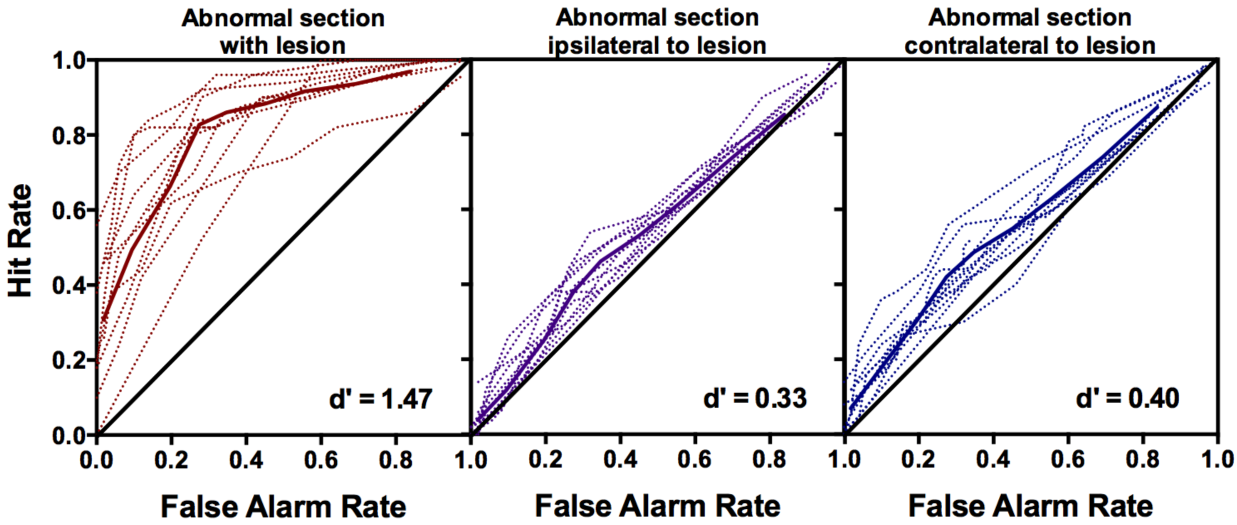

Hit rate
1.0
0.8
0.6
0.4
0.2
0.0
Unfiltered image
d' = 1.06
Low-pass filtered images
d = 0.26
High-pass filtered images
d' = 0.97
0.0
0.2
0.4
0.6
0.8
1.0
False Alarm Rate
False Alarm Rate
False Alarm Rate

Figure 5


Abnormal section
with lesion
Abnormal section
ipsilateral to lesion
Abnormal section
contralateral to lesion
Hit Rate
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
d' = 1.47
d' = 0.33
d' = 0.40
False Alarm Rate
False Alarm Rate
False Alarm Rate