Perception of dim targets on dark backgrounds in MRI

M. Dylan Tisdall and M. Stella Atkins

Computing Science, Simon Fraser University, Burnaby, BC, Canada

ABSTRACT

Some diagnostic tasks in MRI involve determining the presence of a faint feature (target) relative to a dark background. In MR images produced by taking pixel magnitudes it is well known that the contrast between faint features and dark backgrounds is reduced due to the Rician noise distribution. In an attempt to enhance detection we implemented three different MRI reconstruction algorithms: the normal magnitude, phase-corrected real, and a wavelet thresholding algorithm designed particularly for MRI noise suppression and contrast enhancement.

To compare these reconstructions, we had volunteers perform a two-alternative forced choice (2AFC) signal detection task. The stimuli were produced from high-field head MRI images with synthetic thermal noise added to ensure realistic backgrounds. Circular targets were located in regions of the image that were dark, but next to bright anatomy. Images were processed using one of the three reconstruction techniques. In addition we compared a channelized Hotelling observer (CHO) to the human observers in this task. We measured the percentage correct in both the human and model observer experiments.

Our results showed better performance with the use of magnitude or phase-corrected real images compared to the use of the wavelet algorithm. In particular, artifacts induced by the wavelet algorithm seem to distract some users and produce significant inter-subject variability. This contradicts predictions based only on SNR. The CHO matched the mean human results quite closely, demonstrating that this model observer may be used to simulate human response in MRI target detection tasks.

Keywords: Magnetic resonance imaging, Fat separation, Phase error estimation, Denoising, Dixon imaging

1. INTRODUCTION

MR images are the result of processing the complex-valued raw data with a reconstruction operation to produce a real-valued image that can be displayed in greyscale. The most common reconstruction takes the magnitude of the complex-valued pixel as the real-valued estimate. The raw complex data is corrupted with complex additive white Gaussian noise (AWGN),[1] so taking the magnitude produces an image with Rician noise.[2] Images with this noise distribution suffer from reduced contrast in dark regions. To alleviate this problem phase-corrected real reconstruction can be employed to produce an image with Gaussian noise.[3–6] Due to the nature of the Rician noise distribution, these two approaches should produce approximately the same estimate in situations where the signal-to-noise ratio (SNR) is greater than 3. Noting this similarity, in clinical imaging the magnitude estimate is preferred because there is no possibility of error in its computation, unlike phase-correction.

Since the noise in MR images is often considered detrimental to viewers, many reconstruction techniques have been proposed that include filters for noise reduction.[7–9] However, in the literature presenting and comparing these algorithms the quality of the output is usually represented in terms of summary statistics such as SNR, contrast-to-noise ratio (CNR), or mean-to-standard-deviation ratio (MSR) calculated over a set of example images.[6–9]

There is a substantial body of literature using human observers and mathematical models of human task performance on simple detection tasks using x-ray,[10,11] nuclear-medicine,[12,13] and MR images.[14] Noting this, we used both human and model observers to compare three MRI reconstruction algorithms. As in our previous work on MRI reconstruction quality, we used high-SNR MR images as backgrounds combined with complex AWGN to produce our simulated low-SNR raw MRI data and then reconstructed it using three techniques.[14]

Further author information: (Send correspondence to M.D.T.)
M.D.T.: E-mail: mtisdall@cs.sfu.ca, Telephone: 1 604 291 5509
M.S.A.: E-mail: stella@cs.sfu.ca, Telephone: 1 604 291 4288

Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment,
edited by Yulei Jiang, Berkman Sahiner, Proc. of SPIE Vol. 6515
651513, (2007) · 1605-7422/07/$18 · doi: 10.1117/12.709738

In our previous work, we compared human and model observers in MR images where a faint target feature was overlaid on bright regions of anatomy.[14] The present experiment considers images where the target feature was faint and overlaid in a dark region of the image next to anatomy. With this new data set, our detection task was more highly sensitive to the image contrast and noise distribution. As before, our human observers were volunteers without radiological experience.

We compared the humans' performance against a channelized Hotelling observer (CHO)[15,16] in a two-alternative forced-choice (2AFC) signal-known-exactly (SKE) detection task. We 'tuned' the model observers to the human observers' performance on magnitude MR images and then compared the ability of the model to extrapolate the human performance in the other two reconstructions. Experiments structured this way would allow researchers developing new reconstruction algorithms to quantify the results of their work without needing new human studies for each new reconstruction.

In Section 2 we present a simple model of the complex MRI signal and describe the three reconstruction techniques used. In Section 3 we present the process used to generate the synthetic data. Section 4 describes the experiment performed by the human observers and describes the CHO used for comparison. The results of our experiments are presented and discussed in Section 5. Section 6 contains our conclusions.

2. MRI RECONSTRUCTION

The image acquired from an MRI scanner is initially complex-valued and can be modeled by the equation

$$\mathbf{Y}[\boldsymbol{x}] = \mathbf{S}[\boldsymbol{x}] \exp(\mathrm{i}\boldsymbol{\Theta}[\boldsymbol{x}]) + \mathbf{N_r}[\boldsymbol{x}] + \mathrm{i}\mathbf{N_i}[\boldsymbol{x}]\,, \tag{1}$$

where $\boldsymbol{x}$ is the 2D index of a pixel, $\mathbf{Y}$ is the matrix of complex-valued image pixels, $\mathbf{S}$ is the matrix of real-valued signals, $\boldsymbol{\Theta}$ is the matrix of signal phases, and $\mathbf{R}$ and $\mathbf{I}$ are matrices of samples from a normal distribution $\mathcal{N}(0,\sigma)$ representing thermal noise in the real and imaginary components respectively.[17] It is important to note that $\boldsymbol{\Theta}$ is not the phase of the recorded pixel, but the phase of the signal component in the recorded pixel. The complex AWGN represented with $\mathbf{N_r}$ and $\mathbf{N_i}$ will alter the phase of the recorded pixel. For a more detailed model of MRI signal and noise, see Macovski.[18]

In order to display the image in greyscale, a real-valued matrix of pixels, $\mathbf{Y}'$, must be produced. Ideally we would display the real-valued $\mathbf{S}$, but since we only know $\mathbf{Y}$, we attempt to calculate and display a real-valued $\mathbf{Y}'$ that is as close as possible to the unknown $\mathbf{S}$.

2.1. Magnitude reconstruction

The most common reconstruction algorithm in clinical practice is the magnitude transform.[19] In this algorithm, each pixel's real value, $\mathbf{Y}'_m[\boldsymbol{x}]$, is set to the magnitude of the recorded complex pixel. This approach benefits from discarding the effects of the signal phase, $\boldsymbol{\Theta}$. However, the principle difficulty with the magnitude transform is that the complex AWGN in $\mathbf{Y}$ is transformed to a Rician noise in $\mathbf{Y}'_m$ with the attendant problem of bias in low-signal areas reducing image contrast.[2]

2.2. Wavelet reconstruction

A wavelet transform can be applied and particular wavelet coefficients thresholded to enhance contrast and edges in an image. MRI reconstruction algorithms involving wavelets can be applied to the real and imaginary components of $\mathbf{Y}$ before applying a magnitude transform[9] or applied to the real-valued image after the magnitude transform.[7] Algorithms using the wavelet-then-magnitude format will smooth Gaussian noise in each component, but may be affected by the phase of the signal. The alternative, magnitude-then-wavelet approach has the advantage of working on data from which the phase has been discarded, but must be tailored to Rician noise.

We used a wavelet filter developed specifically for Rician noise by Nowak.[7] Since this filter is designed for the magnitude-then-wavelet approach, our wavelet reconstruction process begins with the magnitude reconstruction described in the previous section. We then applied the wavelet filter to the real-valued magnitude data to produce our final real-valued image, $\mathbf{Y}'_w$. In particular, we implemented Nowak's algorithm using the full-scale Harr wavelet transform without any shift-invariant approximations. We chose this filter because it is often cited in the MRI noise reduction literature as a point of comparison for new techniques.

2.3. Phase-corrected real reconstruction

The phase-corrected real reconstruction is an alternative to using the magnitude transform either alone or with wavelets.[3–6] This reconstruction first estimates $\boldsymbol{\Theta}$ with $\tilde{\boldsymbol{\Theta}}$. Assuming that $\tilde{\boldsymbol{\Theta}} \simeq \boldsymbol{\Theta}$ and noting that the thermal noise distribution is rotationally invariant, we can write

$$\mathbf{Y}[\boldsymbol{x}] \exp(-\mathrm{i}\,\tilde{\boldsymbol{\Theta}}[\boldsymbol{x}]) \simeq \mathbf{S}[\boldsymbol{x}] + \mathbf{N_r}[\boldsymbol{x}] + \mathrm{i}\,\mathbf{N_i}[\boldsymbol{x}]\,. \tag{2}$$

The values of the noise samples will have changed, but their distribution will be the same.

Taking just the real component of $\mathbf{Y}[\boldsymbol{x}] \exp(-\mathrm{i}\,\tilde{\boldsymbol{\Theta}}[\boldsymbol{x}])$ gives our final definition for $\mathbf{Y}'_p[\boldsymbol{x}]$, the phase-corrected real reconstructed image:

$$\mathbf{Y}'_p[\boldsymbol{x}] = \mathrm{Re}\left(\mathbf{Y}[\boldsymbol{x}] \exp(-\mathrm{i}\,\tilde{\boldsymbol{\Theta}}[\boldsymbol{x}])\right) \simeq \mathbf{S}[\boldsymbol{x}] + \mathbf{N_r}[\boldsymbol{x}]\,. \tag{3}$$

This shows that, assuming we have estimated $\boldsymbol{\Theta}$ closely, the result of a phase-corrected real reconstruction is an image containing the signal and a real AWGN.

It is important to note that while $\mathbf{Y}'_p$ is an unbiased estimator of $\mathbf{S}$, when $\mathbf{Y}'_p$ is displayed on a monitor the result is that dark regions will be biased positively simply due to the fact that a monitor cannot display negative intensities. However, the displayed bias in $\mathbf{Y}'_p$ is less than that in $\mathbf{Y}'_m$ and so $\mathbf{Y}'_p$ will be closer to $\mathbf{S}$ than $\mathbf{Y}'_m$.

3. SYNTHETIC IMAGES

When attempting to locate a target feature in an MR image there are two major sources of distraction: non-target patient anatomy and thermal noise. In terms of Eq. (1) the anatomy and target feature combine to form the image signal, $\mathbf{S}$, while the thermal noise is the complex AWGN process, $\mathbf{N_r} + \mathrm{i}\,\mathbf{N_i}$. By producing synthetic images with both of these components we propose that target feature detection in our synthetic images will have approximately the same results as similar tasks in clinical MR images when using the reconstructions described in section 2.

One important factor in comparing these reconstruction is the multiplicative feature of Rician noise. As previously noted, when $\frac{\mathbf{S}[\boldsymbol{x}]}{\sigma_{\mathbf{N}}} \geq 3$ we find that $\mathbf{Y}'_m[\boldsymbol{x}] \simeq \mathbf{Y}'_p[\boldsymbol{x}]$.[2] Thus, if we locate the target feature in such a way that it sums with bright anatomy underneath, we will expect very little difference in detection outcomes between the magnitude and phase-corrected images. This was confirmed in our previous work.[14] In order to ensure that our experiment considered the case where there is an expected difference, our synthetic images were designed so that the target feature would be overlaid on a dim region of the image but immediately next to bright anatomy. We simulated the task of detecting a faint anatomical feature next to bright anatomy in order to consider the distracting effect of the background anatomy while still testing our particular combination of faint feature and dark background.

We create a complex-valued image without a target feature via

$$\mathbf{Y} = \mathbf{B} + \mathbf{N_r} + \mathrm{i}\,\mathbf{N_i}\,, \tag{4}$$

where $\mathbf{B}$ is a real-valued background anatomy image, and $\mathbf{N_r}$ and $\mathbf{N_i}$ are images containing real-valued AWGN. Similarly, we produce a complex-valued image with a target feature using

$$\mathbf{Y} = \mathbf{T} + \mathbf{B} + \mathbf{N_r} + \mathrm{i}\,\mathbf{N_i}\,, \tag{5}$$

where $\mathbf{T}$ is the real-valued image containing only the target.

Equations (4) and (5) do not include signal phases. This is a substantial deviation from the model in Eq. (1), but can be be justified by considering the processing that will be applied to these images. The result of the magnitude transform on either feature-present or feature-absent images will be invariant to signal phase since the reconstruction explicitly discards phase information. Similarly, because the wavelet transform being considered operates on images after the magnitude transform, its result is also correct without needing a simulated signal phase.

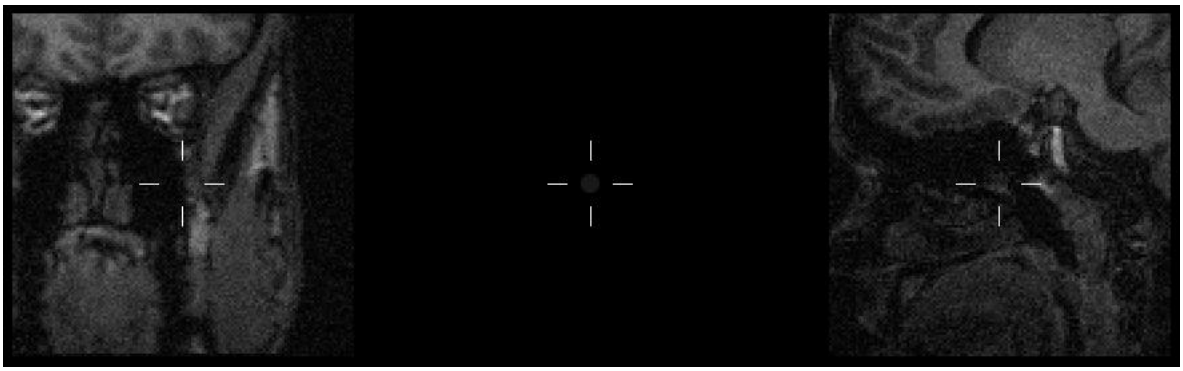

Figure 1. Example of the user interface used in human observer study. In this case the target feature is in the right image.

The phase-corrected real reconstruction does rely on an estimator of the signal phase and thus potentially on a simulated signal phase. However, we note that in the best case, $\tilde{\mathbf{\Theta}}$ is exactly the same as $\mathbf{\Theta}$ and so we can simulate the best case result of phase correction dropping the $\mathrm{i}\,\mathbf{N_i}$ term from equations (4) and (5).

We also note that in practice, if a phase estimation scheme over-fits the data (i.e., if the phase due to noise is fit, instead of just the signal phase) the result will be an approximation to the magnitude transform. Alternatively, if the phase estimation under-fits the data there will be spatially varying signal intensity but the noise power will remain the same. This will have the same effect as lowering the target feature intensity relative to the thermal noise. Thus, while we do not simulate the failure conditions directly in our experiment, the effects of both these types of failures will be discernible from the experiment results because we cover both the magnitude and phase-corrected reconstructions at a range of target feature intensities and thermal noise powers.

As in our previous work, our target feature, $\mathbf{T}$, was an anti-aliased circular object located in the center of the feature-present image (see Fig. 1).[14] For this experiment we set the radius of the circle to three pixels and selected target brightnesses of $b \in \{\frac{3}{100}, \frac{1}{20}, \frac{1}{12}\} \simeq \{0.03, 0.05, 0.083\}$ for each image. The complex-valued thermal noise was simulated by first selecting $\sigma_{\mathbf{N}} \in \{\frac{9}{200}, \frac{3}{40}\} \simeq \{0.045, 0.075\}$. Based on this choice, we randomly generated two 128×128 pixel images for each synthetic MRI. Each pixel in these noise images was sampled from $\mathcal{N}(0, \sigma_{\mathbf{N}})$. One of these images was taken as $\mathbf{N_r}$ and the other as $\mathbf{N_i}$.

To simulate distracting anatomy, $\mathbf{B}$, we used regions of slices from high-SNR MR head images of healthy volunteer. These volunteers were scanned using a 3D inversion recovery pulse sequence on a Philips Gyroscan Intera 3.0 Tesla MRI scanner. Each volume was reconstructed using the magnitude transform to give real-valued images. These real-valued volumes were then sliced along the axial, coronal, and sagittal directions to produce a library of 2D images from which appropriate 128×128 pixel sub-images were selected. 560 backgrounds were produced in each of the 36 possible conditions (three reconstructions, three signal powers, two noise powers, target feature present or absent).

As noted in section 2, the background images, $\mathbf{B}$, produced by cropping and normalization had Rician noise because the magnitude reconstruction was used to produce the real values. Since we used these images as our real-valued data, this Rician noise was added to the Gaussian $\mathbf{N_r}$ and included in all of the synthetic image reconstructions. However, because our anatomical images were scanned at 3 Teslas, the magnitude images, after being normalized to the range $(0, 1)$, had a standard deviation of approximately 0.002 measured in regions of air. Since the lowest standard deviation, $\sigma_{\mathbf{N}}$ used for our simulated thermal noise was more than 20 times greater than the inherent noise of our anatomical backgrounds, it is unlikely that the spurious noise included in the anatomical images had any effect on our results.

4. OBSERVERS

4.1. Human observers

Eight volunteer observers without any previous radiological training were recruited to participate in our study. The experimental software presented participants with three images aligned horizontally (see Fig. 1). The center image showed the target feature and the two exterior images represented choices in the 2AFC test. Since this was an SKE task, crosshairs were superimposed over the images in order to reduce the possibility of confusion about the target feature location. The crosshairs could be toggled on and off by the user to reduce visual distraction. Participants were instructed that in every display, one of the exterior images would contain the target feature and that they should use the mouse to click on whichever exterior image they felt most probably contained the target. They were allowed to take as long as they wanted to reach a decision on each image pair. Once a participant clicked on an exterior image, the screen was made completely black for 0.5 seconds and then the next set of images was shown and the process repeated.

Each participant was given two training sets, composed equally of all 18 possible combinations of noise power, target feature power, and reconstruction algorithm. If the training took less than ten minutes, they were then instructed to wait until ten minutes had elapsed in order to ensure a constant dark adaptation time across all participants. After the training and delay, the participants then proceeded through 16 experiment sets composed equally of the 18 possible combinations for a total of 288 image pairs. In order to minimize order effects, the ordering of the image pairs was selected randomly for each participant from a constant distribution of all possible orderings. As the experiment lasted up to an hour, fatigue was reduced by displaying a black screen after every 18 image pairs and instructing users to take as long a break as they desired while the display was dark.

The experiments were conducted in a completely darkened room using a CRT monitor (SGI CMNB024B) as a display. All participants viewed the monitor from approximately 50 cm away while wearing any corrective lenses they would normally use for computer viewing. When displayed on the screen, the images had a diameter of approximately 8.5 cm and so occupied an angle of approximately 10 degrees from the viewer's eye position. All negative image pixel intensities were truncated at 0 in the reconstructed images since negative intensities are not displayable. The maximum pixel intensity in the set was found (~ 1.2925) and all pixels' intensities were then scaled by the same value (~ 197.29) to ensure the entire set fell on the range (0,255) for greyscale display. By imposing a consistent scaling on all images, some (e.g., those with a low-intensity target feature and low noise power) did not use the entire range of display intensities. Using the above scaling we see that target feature intensity for $b = \frac{3}{100}$ was mapped to 5.92 on the greyscale range, $b = \frac{1}{20}$ was mapped to 9.86, and $b = \frac{1}{12}$ was mapped to 16.44.

4.2. Channelized Hotelling observer

We applied the CHO[15,16] with Gabor channels to our synthetic images to compare the results with human performance. Gabor channels in particular were used because it has been suggested they are a useful approximation for the grating response of the human visual system.[10,11,20] We have used a setup with forty channels, based on the example of Eckstein,[10,11] with a filter width of 0.8825 octaves, angles of $\{0, \frac{2\pi}{5}, \frac{4\pi}{5}, \frac{6\pi}{5}, \frac{8\pi}{5}\}$, and central frequencies of 2, 4, 8, or 16 cycles per degree. The $16,384 \times 40$ channel matrix, $\mathbf{C}$, is produced by rearranging each channel, $\mathbf{G}$, as a $16,384 \times 1$ vector and making them each a column of $\mathbf{C}$.

In order to derive the CHO for each of the 18 experimental conditions, we must also compute the specific covariance matrix, $\mathbf{K}_{\boldsymbol{u}}^{c}$, for each condition, c. We first note that

$$\mathbf{K}_{\boldsymbol{u}}^{c} = \frac{1}{2}(\mathbf{K}_{\boldsymbol{u},0}^{c} + \mathbf{K}_{\boldsymbol{u},1}^{c}) + \mathbf{K}_{\boldsymbol{\epsilon}}^{c} \tag{6}$$

where $\mathbf{K}_{\boldsymbol{u},0}^{c}$ and $\mathbf{K}_{\boldsymbol{u},1}^{c}$ are the covariance matrices of the channel responses in condition c's target feature-absent and -present cases respectively, and $\mathbf{K}_{\boldsymbol{\epsilon}}^{c}$ is the covariance matrix of the observer's internal noise process in condition c. This process is assumed to add noise independently to each response channel by sampling from a Gaussian with zero mean and variance depending on the channel. Following the example of Eckstein, we define $\mathbf{K}_{\boldsymbol{\epsilon}}^{c} = \alpha_{\mathrm{CHO}} \operatorname{Diag}\left(\frac{1}{2}(\mathbf{K}_{\boldsymbol{u},0}^{c} + \mathbf{K}_{\boldsymbol{u},1}^{c})\right)$ where Diag() zeroes all the off-diagonal elements of its argument and α_{CHO} is a proportionality constant that can be varied to reduce the absolute performance of the model observer.[11]

This leaves the problem of determining $\mathbf{K}^c_{\boldsymbol{u},0}$ and $\mathbf{K}^c_{\boldsymbol{u},1}$. Given that we have closed forms for neither the pixel covariance of the anatomical backgrounds, nor the effects of the wavelet filter, we opted to estimate $\mathbf{K}^c_{\boldsymbol{u},0}$ and $\mathbf{K}^c_{\boldsymbol{u},1}$ from the synthetic data. Noting that $\mathbf{K}^c_{\boldsymbol{u},0}$ and $\mathbf{K}^c_{\boldsymbol{u},1}$ are each 40×40 matrices, we used the channel responses from 400 signal-present and signal-absent images respectively to estimate each for each condition. Since we had 560 target feature-present and 560 target feature-absent images in each of the 18 experimental conditions, this left 160 image pairs for testing each condition.

Having derived all the necessary components, we can write the template, $\boldsymbol{w}^c$, applied by the observer in condition c as

$$\boldsymbol{w}^c = \mathbf{C}\left(\mathbf{K}^c_{\boldsymbol{u}}\right)^{-1}\left(\langle \boldsymbol{u}^c_1\rangle - \langle \boldsymbol{u}^c_0\rangle\right), \tag{7}$$

where $\langle \boldsymbol{u}^c_1\rangle$ and $\langle \boldsymbol{u}^c_0\rangle$ are the sample mean target feature-present and target feature-absent channel response vectors for condition c. This template can be used to calculate the response, λ, to a reconstructed image $\mathbf{Y}'$. Reordering the 128×128 matrix $\mathbf{Y}'$ to the $16,384 \times 1$ vector $\boldsymbol{y}'$ we write

$$\lambda = (\boldsymbol{w}^c)^t \boldsymbol{y}', \tag{8}$$

where $\boldsymbol{y}'$ was produced with conditions c.

The final step is to add the internal noise of the observer. Rather than compute a noise for each channel, we can modify the computed response by adding a single sample, ϵ, from a zero-mean Gaussian distribution with variance

$$\sigma^2_\epsilon = \left[\left(\mathbf{K}^c_{\boldsymbol{u}}\right)^{-1}\left(\langle \boldsymbol{u}^c_1\rangle - \langle \boldsymbol{u}^c_0\rangle\right)\right]^t \mathbf{K}_{\boldsymbol{\epsilon}}\left(\mathbf{K}^c_{\boldsymbol{u}}\right)^{-1}\left(\langle \boldsymbol{u}^c_1\rangle - \langle \boldsymbol{u}^c_0\rangle\right). \tag{9}$$

Adding this noise sample gives our final estimate of the score a human observer would assign to the image:

$$\lambda' = \lambda + \epsilon\,. \tag{10}$$

The image with the greater λ' in each pair is considered the CHO selection in the 2AFC trial.

5. RESULTS AND DISCUSSION

5.1. Human observer study

For each participant, we computed the percentage correct, Pc, in each of the 18 experimental conditions. We computed the median Pc over all the participants as well as the first and third quartiles. It can be shown that in a 2AFC task, the Pc is also an estimator for the area under the curve (AUC) of the experiment's ROC.[21] Noting this, we have plotted the median, first quartile, and third quartile of the AUC in Fig. 2. Although the width of the quartile measurements imply a large amount of inter-subject variability this is also an effect of our experiment design. The resolution of our AUC quartile measurements is only $\frac{1}{16}$ because we opted to cover a variety of experimental cases and thus show each participant each condition only 16 times.

In each condition, we summed the number of correct and incorrect trials across all 8 participants to produce the success and failure rate of the average participant. To compare equivalent conditions between two reconstructions, A and B, we use the one-sided Fisher's Exact test to evaluate the null hypothesis that viewers do not perform better on average with A than B. We found that there was a significant performance gain to a particular reconstruction when compared at almost all pairs of equivalent conditions (see Table 1). In particular, we found that between magnitude and phase-corrected images in all cases there was either no significant difference, or a significant difference in favour of phase correction at $p \leq 0.05$. Similarly, we found that there was a statistically significant difference at $p \leq 0.05$ in favour of magnitude or phase correction over the wavelet processing in almost every case.

The human results confirm the theoretical assertion that phase-corrected images should improve signal detection in dark regions of the image. However, these results also conflict with the assertion that the wavelet algorithm should enhance signal detection. Thus, we see that these human studies can be a valuable tool for evaluating MRI reconstructions since they provide information that is not necessarily apparent from mean square error (MSE) or other image distance metrics.

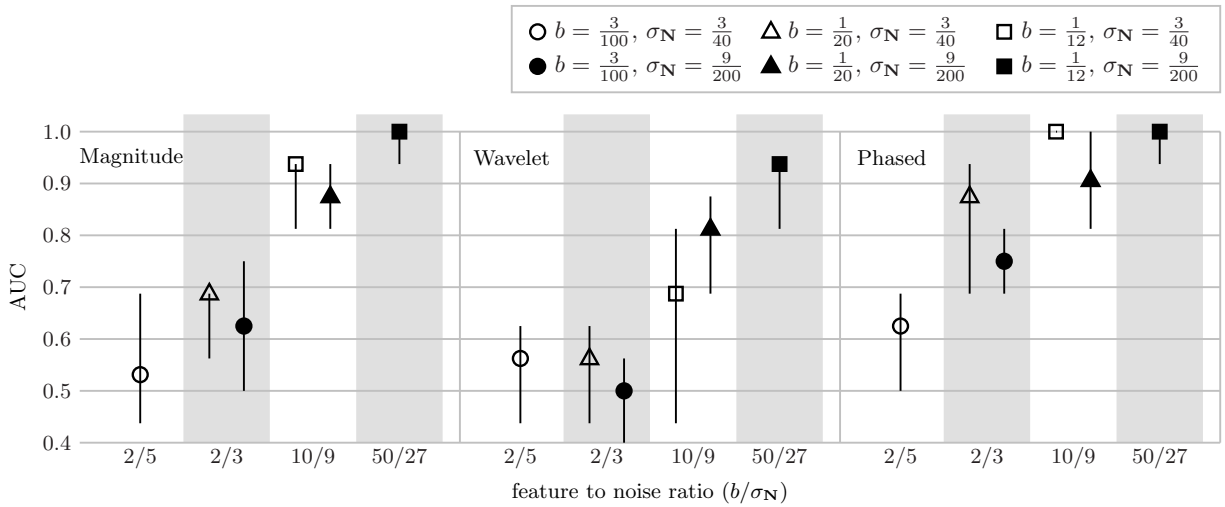

Figure 2. Plot of the median, first quartile, and third quartile of AUC for the human observers in each of the 18 experimental conditions. The y-axis is the AUC score. Each experimental condition is represented by a symbol (one of three shapes, either filled or unfilled) that locates the median AUC and two vertical lines that represent the first and third quartiles of the AUC. The symbol's shape represents a different configuration of target feature and noise power, as explained in the legend at the top-right of the chart. The x-axis is divided into thirds, with each third containing results from one of the reconstruction technique as labeled in the top-left corner of each third. Inside of each third, the x-axis is further divided by the ratio of target feature to thermal noise standard deviation (complex feature-SNR). In some cases, two experimental conditions have the same feature to noise ratio (e.g,. $b = \frac{1}{12}, \sigma_{\mathbf{N}} = \frac{3}{40}$ and $b = \frac{1}{20}, \sigma_{\mathbf{N}} = \frac{9}{200}$) and so appear in the same band on the diagram.

p-values for the one-sided Fisher's Exact test

Condition			Phased/Wavelet	Phased/Magnitude	Magnitude/Wavelet
○	$b = \frac{3}{100}$	$\sigma_{\mathbf{N}} = \frac{3}{40}$	$\mathbf{6.46 \times 10^{-2}}$	**0.262**	**0.226**
●	$b = \frac{3}{100}$	$\sigma_{\mathbf{N}} = \frac{9}{200}$	1.40×10^{-5}	4.82×10^{-2}	7.87×10^{-3}
△	$b = \frac{1}{20}$	$\sigma_{\mathbf{N}} = \frac{3}{40}$	6.08×10^{-6}	6.88×10^{-3}	3.55×10^{-2}
▲	$b = \frac{1}{20}$	$\sigma_{\mathbf{N}} = \frac{9}{200}$	7.06×10^{-3}	**0.215**	6.77×10^{-3}
□	$b = \frac{1}{12}$	$\sigma_{\mathbf{N}} = \frac{3}{40}$	1.26×10^{-11}	8.75×10^{-3}	7.19×10^{-6}
■	$b = \frac{1}{12}$	$\sigma_{\mathbf{N}} = \frac{9}{200}$	5.35×10^{-3}	**0.5**	1.50×10^{-2}

Table 1. p-values for the one-sided Fisher's Exact test in each in the 18 pairs of equivalent experimental conditions. The columns are the pairs reconstructions that were compared, testing the null hypothesis that the first reconstruction does not produce superior mean detection rate compared to the second. The rows indicate the level of signal and noise at which the null hypothesis was tested. We rejected the null hypothesis at $p \leq 0.05$ except in cases that are highlighted in bold. Note that the comparison between phased and magnitude reconstructions fails at $b = \frac{1}{12}$ $\sigma_{\mathbf{N}} = \frac{9}{200}$ (black square) because of the design of the trial: Figure 2 shows that the majority of users were perfectly accurate in both trials. This choice of noise and signal power does, however, allow us to distinguish the wavelet reconstruction from the other two.

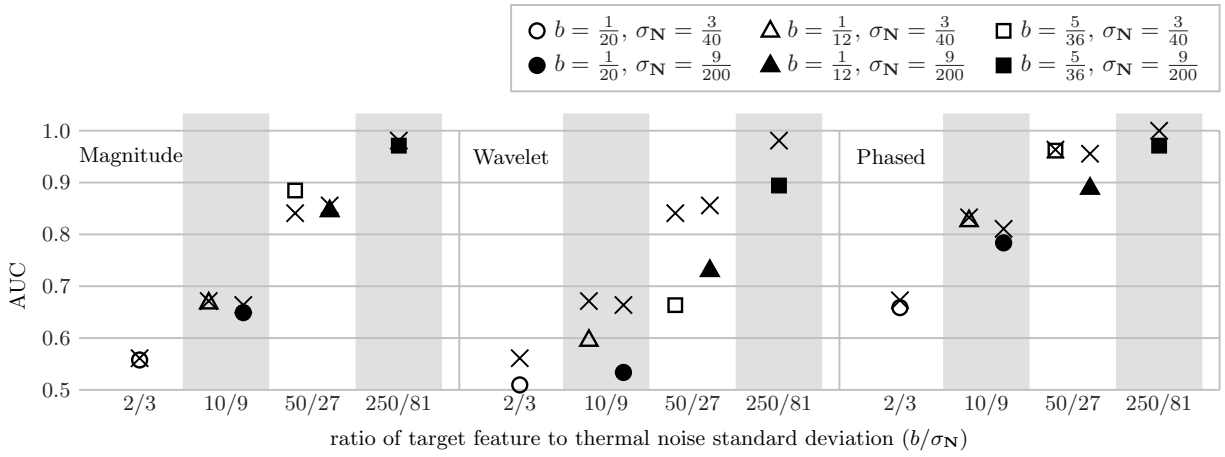

Figure 3. Plot of the mean AUC of the the model and mean AUC of the 8 human observers in each of the 18 experimental conditions. The y-axis is the AUC score. Each experimental condition is represented by a symbol (one of three shapes, either filled or unfilled) that locates the mean human observer AUC. The symbol's shape represents a different configuration of target feature and noise power, as explained in the legend at the top-right of the chart. Either directly above or below the symbol will be one x-mark representing the CHO result for the experimental condition. The x-axis is divided into thirds, with each third containing results from one of the reconstruction technique as labeled in the top-left corner of each third. Inside of each third, the x-axis is then further divided by the ratio of target feature to thermal noise standard deviation (complex feature-SNR). In some cases, two experimental conditions have the same PSNR (e.g,. $b = \frac{1}{12}, \sigma_{\mathbf{N}} = \frac{3}{40}$ and $b = \frac{1}{20}, \sigma_{\mathbf{N}} = \frac{9}{200}$) and so appear in the same band on the diagram.

5.2. Model observer study

Ideally, we would compute the AUC or some similar metric directly from the description of our model observer and the image statistics. However, due to the fact that we have neither stationary image backgrounds, nor Gaussian-distributed model observer responses, λ, we calculated the AUC for the CHO by applying the template to each synthetic image, computing λ' using equation (10), and then to calculating the Pc. We performed this operation for 50 separate instances of each model observer, and then computed the result of the mean observer in each of the experimental conditions. As in the human observer case, the Pc values were taken as estimates of the AUC.

The mean AUC of the model observers was fit to the mean human observer data by searching for the α_{CHO} that minimized the MSE between the human and model results in the magnitude reconstructed experimental conditions. This was chosen to test the ability for the observers to extrapolate to other reconstructions with different statistical properties based on human results with only the magnitude reconstruction. Using humans to calibrate the models based on only magnitude images was a feasible model for how an experimenter might proceed to test a new MRI reconstruction. With this approach the experimenter must only record human results once for the magnitude reconstructions of a given data set and then could proceed to explore as many other reconstructions as they desired using only the models to evaluate as long as the statistical properties of the data set were unchanged.

For the CHO, we determined the best fit to the mean human results on magnitude images to be $\alpha_{\text{CHO}} = 0.9$ by searching on an initially coarse range of α_{CHO} values and then gradually refining the range. The mean AUC of the CHO is shown in comparison to the mean human results in Fig. 3 (note that Fig. 2 displays the median human results while Fig. 3 displays the mean). Overall, the CHO shows a good match with the human study. The orderings both between reconstructions and between conditions are approximately maintained with the exception of the wavelet reconstruction. We see that the CHO has substantially over-estimated human performance for the wavelet reconstruction. Noting that many of our human volunteers complained of 'blurriness' in the wavelet reconstructed images, we suspect that the CHO may be more successful than humans when compensating for the smoothing due to its use of the covariance matrix.

6. CONCLUSION

Among our human observers, we found a statistically significant difference between the three reconstructions at almost all of the target feature amplitude and noise variance levels tested. Interestingly, the preferences displayed in our human studies supported the theoretical suggestion that phase-corrected reconstructions are an improvement over the magnitude while disagreeing with the assertion that the wavelet filter would improve detection. It is important to note that this difference was only found for a very specific task in MRI: the detection of faint objects next to bright anatomy. While this was clearly the 'best-case' scenario for applying phase-correction, it is unclear if a scenario exists where the wavelet filtering would be superior to the magnitude reconstruction.

The model observers matched the human results well in all conditions when fit to only the magnitude human results. In general, the models indicated the same preference for phase-correction and decrease in detection with the wavelet filter. The over-estimate of human performance by the CHO in the wavelet reconstructions likely has to do with the correlations introduced by the smoothing of the wavelet filter.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Burkhard Maedler at Philips Medical Systems for providing the high-field MRI images. The authors would like to acknowledge the National Science and Engineering Research Council for their support of this research.

REFERENCES

1. W. A. Edelstein, G. H. Glover, C. J. Hardy, and R. W. Redington, "The intrinsic signal-to-noise ratio in NMR imaging," *Magnetic Resonance in Medicine* **3**, pp. 604–618, 1986.
2. H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magn. Reson. Med.* **34**, pp. 910–914, 1995.
3. M. A. Bernstein, D. M. Thomasson, and W. Perman, "Improved detectability in low signal-to-noise ratio magnetic resonance images by means of a phase-corrected real reconstruction," *Med. Phys.* **16**, pp. 813–817, Sept. 1989.
4. D. Noll, D. Nishimura, and A. Macovski, "Homodyne detection in magnetic resonance imaging," *IEEE Trans. Med. Imag.* **10**, pp. 154–163, 1991.
5. M. D. Tisdall and M. S. Atkins, "MRI denoising via phase error estimation," in *Proc. SPIE*, **5747**, pp. 646–654, 2005.
6. Z. Chang and Q.-S. Xiang, "Nonlinear phase correction with an extended statistical algorithm," *IEEE Trans. Med. Imag.* **24**, pp. 791–798, June 2005.
7. R. D. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," *IEEE Trans. Image Processing* **8**, pp. 1408–1418, Oct. 1999.
8. M. E. Alexander, R. Baumgartner, A. R. Summers, C. Windischberger, M. Klarhoefer, E. Moser, and R. L. Somorjai, "A wavelet-based method for improving signal-to-noise ratio and contrast in MR images," *Magn. Reson. Imaging* **18**, pp. 169–180, 2000.
9. P. Bao and L. Zhang, "Noise reduction for magnetic resonance images via adaptive multiscale products thresholding," *IEEE Trans. Med. Imag.* **22**, pp. 1089–1099, Sept. 2003.
10. M. P. Eckstein, C. K. Abbey, and J. S. Whiting, "Human vs model observers in anatomic backgrounds," in *Proc. SPIE*, **3340**, pp. 16–26, 1998.
11. M. P. Eckstein, C. K. Abbey, F. O. Bochud, J. L. Bartoff, and J. S. Whiting, "The effect of image compression in model and human performance," in *Proc. SPIE*, **3663**, pp. 243–252, 1999.
12. J. P. Rolland and H. H. Barrett, "Effect of random background inhomogeneity on observer detection performance," *J. Opt. Soc. Am. A* **9**, pp. 649–658, May 1992.
13. A. E. Burgess, "Statistically defined backgrounds: performance of a modified nonprewhitening observer model," *J. Opt. Soc. Am. A* **11**, pp. 1237–1242, Apr. 1994.
14. M. D. Tisdall and M. S. Atkins, "Using human and model performance to compare MRI reconstructions," *IEEE Trans. Med. Imag.* **25**, pp. 1510–1517, Nov. 2006.

15. K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model," *J. Opt. Soc. Am. A* **4**, pp. 2447–2457, Dec. 1987.
16. C. K. Abbey and F. O. Bouchud, *Handbook of Medical Imaging*, vol. 1, ch. 11 - Modelling Visual Detection Tasks in Correlated Image Noise with Linear Model Observers, pp. 629–654. SPIE - The International Society for Optical Engineering, 2000.
17. R. M. Henkelman, "Measurement of signal intensities in the presence of noise in MR images," *Med. Phys.* **12**, pp. 232–233, Mar. 1985.
18. A. Macovski, "Noise in MRI," *Magn. Reson. Med.* **36**, pp. 494–497, 1996.
19. V. Kuperman, *Magnetic Resonance Imaging - Physical Principles and Applications*, Academic Press, 2000.
20. S. Park, "Efficiency of the human observer detecting random signals in random backgrounds," *J. Opt. Soc. Am. A* **22**, pp. 3–16, Jan. 2005.
21. H. H. Barrett, C. K. Abbey, and E. Clarkson, "Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating function," *J. Opt. Soc. Am. A* **15**, pp. 1520–1535, June 1998.