



Observer efficiency in free-localization tasks with correlated noise

Craig K. Abbey* and Miguel P. Eckstein

Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA, USA

Edited by:

Rémy Allard, Université Pierre et Marie Curie, France

Reviewed by:

Markus Lappe, Universität Münster, Germany

Peter Neri, University of Aberdeen, UK

***Correspondence:**

Craig K. Abbey, Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA 93106, USA
e-mail: abbey@psych.ucsb.edu

The efficiency of visual tasks involving localization has traditionally been evaluated using forced choice experiments that capitalize on independence across locations to simplify the performance of the ideal observer. However, developments in ideal observer analysis have shown how an ideal observer can be defined for free-localization tasks, where a target can appear anywhere in a defined search region and subjects respond by localizing the target. Since these tasks are representative of many real-world search tasks, it is of interest to evaluate the efficiency of observer performance in them. The central question of this work is whether humans are able to effectively use the information in a free-localization task relative to a similar task where target location is fixed. We use a yes-no detection task at a cued location as the reference for this comparison. Each of the tasks is evaluated using a Gaussian target profile embedded in four different Gaussian noise backgrounds having power-law noise power spectra with exponents ranging from 0 to 3. The free localization task had a square 6.7° search region. We report on two follow-up studies investigating efficiency in a detect-and-localize task, and the effect of processing the white-noise backgrounds. In the fixed-location detection task, we find average observer efficiency ranges from 35 to 59% for the different noise backgrounds. Observer efficiency improves dramatically in the tasks involving localization, ranging from 63 to 82% in the forced localization tasks and from 78 to 92% in the detect-and-localize tasks. Performance in white noise, the lowest efficiency condition, was improved by filtering to give them a power-law exponent of 2. Classification images, used to examine spatial frequency weights for the tasks, show better tuning to ideal weights in the free-localization tasks. The high absolute levels of efficiency suggest that observers are well-adapted to free-localization tasks.

Keywords: free-localization tasks, ideal observer theory, power-law noise, observer efficiency, image statistics

INTRODUCTION

The concept of calculation efficiency, which we refer to simply as efficiency, in the presence of image noise has been used extensively as a method for understanding visual processing since its seminal introduction by Barlow (Barlow, 1977, 1978; Barlow and Reeves, 1979). At the core of this measure is comparison with an optimal decision maker, the ideal observer, for a given task. The use of the ideal observer as yardstick for human performance implicitly controls for the relevant information present in stimuli used to perform a task. This topic has a long history in vision science, as well as areas of applied vision such as medical imaging. In the realm of vision science, there are many examples where efficiency is used to reveal the presence (or absence) of limitations and constraints in visual processing (Barlow, 1978; Barlow and Reeves, 1979; Burgess et al., 1981; Pelli, 1985; Legge et al., 1987; Geisler, 1989; Tjan et al., 1995). In imaging applications, efficiency is used to identify opportunities for image processing or other methodological changes that lead to improved performance in visual tasks (Myers et al., 1985; Wagner and Brown, 1985; Insana and Hall, 1994; Siewerdsen and Jaffray, 2000; Abbey et al., 2006).

Studies evaluating efficiency have often relied on experimental paradigms where the location of a target, if it is present, is explicitly defined through the use of location cues. Forced-choice paradigms, with two or more specified locations that serve as possible target locations, are a common choice (Burgess and Ghandeharian, 1984). These studies do involve spatial (or temporal) search, but it is a limited search that is confined to choosing between distinct, cued locations. The use of independent noise masking the target at each location makes the computation of the ideal observer considerably easier. Studies that have analyzed the ideal observer in tasks with location uncertainty on a quasi-continuous scale (i.e., limited to the pixelation of the stimulus) have generally utilized a detection or discrimination response that did not involve localizing the target (Park et al., 2005; Tjan and Nandy, 2006; Neri, 2010).

However, recent analysis by Khurd and Gindi (2005) have demonstrated how an ideal observer may be evaluated when targets can be located anywhere within a search region, and the task requires localizing targets to within a fixed distance, or more general acceptance region. This general paradigm has been used previously in medical imaging studies (Burgess et al., 2001;

Bochud et al., 2004) due to the similarity with many clinical tasks that require identifying a location in the body for further assessment. However, these studies did not have the benefit of an ideal observer. The Khurd and Gindi approach leads to the definition of an optimal decision function, from which ideal observer performance can be extracted via simulation studies, as we do below. Extensions to the theory (Khurd et al., 2010) include methods for evaluating the presence of multiple targets that are beyond the scope of this work. There has been some use of this analysis to evaluate the role of regularization in emission computed tomography (Liu et al., 2009). However, we are not aware of any use of the ideal observer for examining efficiency in more general free-localization tasks.

The main focus of this work is a comparison of fixed-location detection tasks—where a single target location is well cued—to free-localization tasks, where the subject must indicate the location of a target that can be anywhere in a defined search region. **Figure 1** gives examples of the stimulus displays for each task. For the detection tasks, subjects render a decision on whether a Gaussian “bump” target profile is present or not at the cued location. In the localization tasks, the subject is required to indicate the location of the target profile, which will always be present somewhere in the search region. We are interested in comparing the efficiency of human observers in these two tasks, and understanding the mechanisms that can explain differences.

We are also interested in the role of the background image statistics on this process. For this reason, we evaluate four different Gaussian image textures. These are defined by their power spectra, which are constrained to be a power-law parameterized by the power-law exponent β . We evaluated four different background textures, based on β -values ranging from 0 (white noise) to 3 (See **Figure 2** below). Natural scenes are often modeled as power-law processes with exponents that vary around $\beta = 2$ (Burton and Moorhead, 1987; Field, 1987). Various forms of x-ray images, for breast imaging in particular, have also been

modeled as power-law processes with exponents ranging from less than 2 for computed tomography reconstructions (Metheany et al., 2008; Chen et al., 2012, 2013) to 3 or more for tomosynthesis (Engstrom et al., 2009) or projection images (Bochud et al., 1999; Burgess et al., 2001).

In addition to the main comparison of detection and localization tasks, two smaller follow-up studies were conducted to give additional insight on issues that arose from the primary study. One issue was the different nature of the two tasks, given that a yes-no type detection task requires maintaining some sort of detection criterion from trial to trial for choosing the response. The free-localization tasks do not require this. To investigate the effect of a detection criterion on free-localization tasks, we evaluated a detect-and localize (D&L) task, in which the target profile appeared at a random location in the search region in 50% of the trials, and was not present in the other 50% if the trials. As an alternative to indicating the target location, the subjects could also respond “not present” in these experiments. In this way, the requirement of maintaining a task criterion in the detection task was matched in a task with substantial spatial uncertainty.

The second follow-up study concerned the white-noise ($\beta = 0$) background condition, where lower efficiency than other power-law backgrounds was observed in both detection and localization tasks. In this case, we were interested in whether processing the images to have more favorable background statistics could improve performance. We evaluated task performance after filtering these images to have a power-law power spectrum with $\beta = 2$, which also modified the profile of the target.

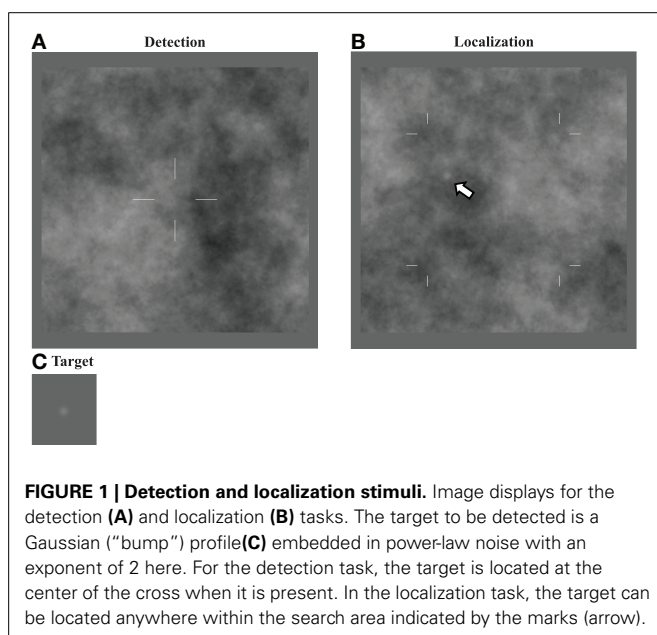
METHODS

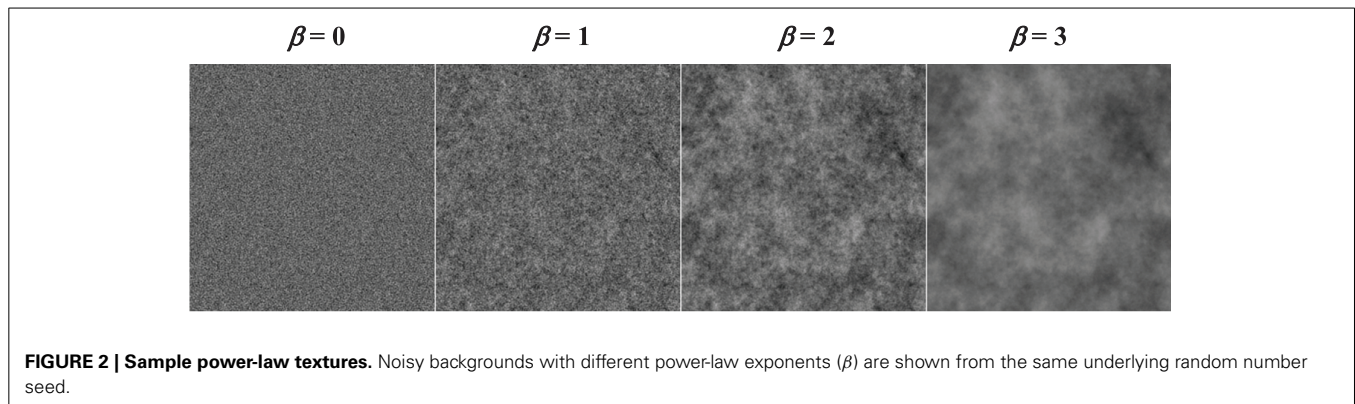
A total of 5 subjects participated in the primary comparison of efficiency between detection and forced-localization tasks. On subject (S1) was a coauthor of this work, and the other 4 were naïve to the purposes of the research and compensated for their participation. Of these, 3 subjects (S2, S4, and S5) participated in the secondary detect-and-localize experiments, and 3 subjects (S3, S4, and S5) participated in the secondary image-processing experiments.

STIMULUS AND DISPLAY PROPERTIES

A monochrome CRT display (Imaging Systems, Minnetonka, Minnesota) with a dedicated controller (DOME, NDS Inc., San Jose, CA), was used for all experiments, which were conducted in a darkened room. The monitor was photometer-calibrated to an 8-bit linear lookup table (LUT) that ranged from 0.02–40 Cd/m². Viewing distance of the subjects was not constrained. Subjects had normal vision or wore corrective lenses. After becoming familiar with the display procedure, and completing several sessions of experiments, measurements of each subject’s comfortable viewing distance were made. The average viewing distance used was 64 cm, with a range of 51–70 cm. This average distance was used for all subsequent calculations of visual angle. The stimuli were generated as 256 by 256 pixel images, and these were magnified by a factor of two for display, making the effective pixel size 0.052° (0.583 mm).

The experiments used a Gaussian “bump” as a target added to stationary noise with a power-law power spectrum. The spatial





standard deviation of the target was 3 pixels giving the displayed target a FWHM of 0.37° . A mean background level of 100 gray levels (gl) was added as well and the noise was scaled to have a pixel variance of 400 gl^2 , which is equivalent to a 20% RMS contrast on the linearized display that was used for the psychophysical studies. Target contrast varied over the different experiments, as described below.

Noise backgrounds were generated by filtering white noise to achieve a power-law power spectrum in the spatial-frequency domain, $S_\beta(f) = C_\beta/f^\beta$, which we will identify by the power-law exponent, β . To avoid the singularity at $f = 0$, the DC component is set to the value of the first harmonic. The normalization constant, C_β , is set so that the RMS contrast of the background is fixed at 20%. The noise generation filter for each background condition was set to be the square-root of $S_\beta(f)$. Examples of the different noise textures for the four values of β that we used are seen in **Figure 2**.

For detection tasks using the different backgrounds, targets had a 50% probability of being present in any given trial. Subjects were informed that this was the target probability. When present, the target was always located in the center of the image with the location indicated by cross-hairs, as shown in **Figure 1**. The observer response was obtained by capturing a mouse click outside the image area. Feedback (correct/incorrect) was given after each trial. While separate performance measures were determined for target present and target absent images (hit rate and false alarm rate), for purposes of fitting psychometric functions the proportion of correct responses was also used.

For localization tasks, the target was randomly located in the central region of the image, with borders delineated by hash marks. Subjects were informed that this was the search region. The central region consisted of 128 by 128 pixels (6.7 by 6.7°), and thus constituted one quarter of the total image area. The large border region was chosen to minimize any edge effects as well as effects from “wrap-around” from the filtering operation. Observers responded by clicking with a mouse on their selected location. Mouse-clicks that were 5 pixels (0.26°) or less from the center of the target were considered “correct,” and subject performance was measured as the proportion of correct responses. Subjects received feedback (correct/incorrect and true target location) after each trial.

Slight modifications to the experimental protocol above were used in the two follow-up studies. For the D&L tasks, there was a 50% probability that the target was present somewhere in the search region. Subjects responded with a mouse-click on the target location to indicate target presence at that location, or by a mouse-click outside the image to indicate target not present. Subjects received feedback (correct/incorrect and target location if applicable) after each trial.

In the image processing study, white noise images (i.e., $\beta = 0$) were filtered after the target profile was added to have a $\beta = 2$ power-law spectrum. The frequency profile of the filter was $\sqrt{S_2(f)}$. As a result of this filtering operation, the target was no longer a Gaussian profile, which reflects the practical reality that image processing alters the properties and appearance of both target and background. Examples of the an image before and after filtering, as well as a plot showing the effect on the target profile, are seen in **Figure 3**. Detection and localization tasks on the processed images were run as described above.

THE IDEAL OBSERVER

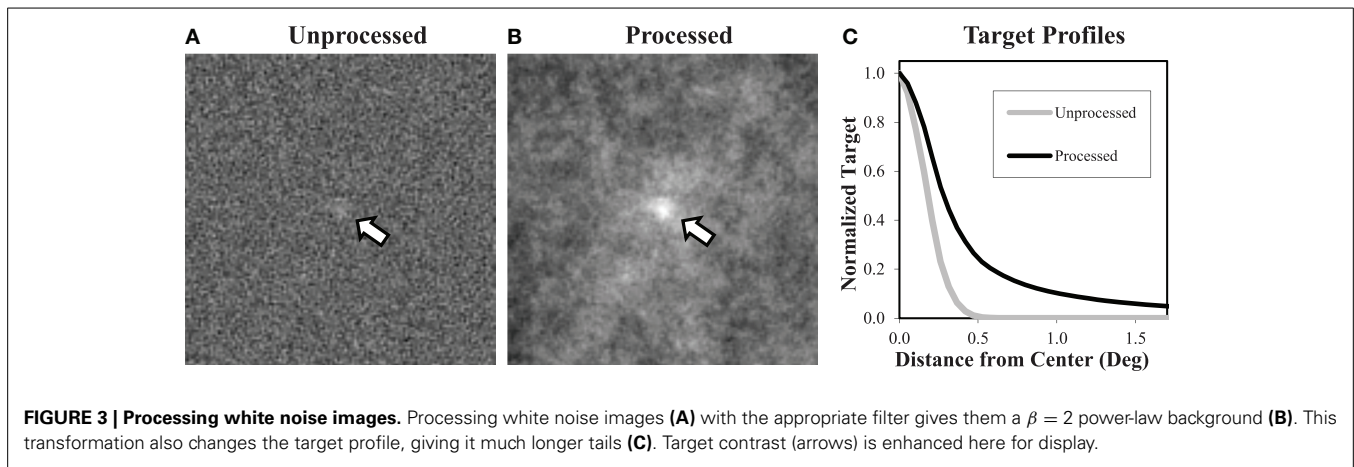
Task efficiency with respect to the ideal observer is the fundamental calculation used in this work. In this section we describe how the ideal observer analysis is implemented, leading to an efficiency estimate.

Detection tasks

For the yes-no detection task we identify target-present images as one hypothesis (or class), H_1 , and the target-absent images as the other possible hypothesis, H_0 . We will refer to the images generically as \mathbf{g} , a column vector of pixel values, with the assumption that the mean background intensity of the stimuli (100 gl in our case) has been subtracted off of the pixel intensities. The Gaussian noise in the images is specified by a multivariate normal distribution (MVN) with a covariance matrix that depends on the power-law exponent of the noise texture, Σ_β . The conditional distributions of the resulting images are given by

$$\begin{aligned} p(\mathbf{g}|H_0) &= \text{MVN}(\mathbf{0}, \Sigma_\beta) \\ p(\mathbf{g}|H_1) &= \text{MVN}(\mathbf{s}, \Sigma_\beta). \end{aligned} \quad (1)$$

Under these conditions, it is well known that the ideal observer can be implemented as a weighted sum of the pixel intensities



(Green and Swets, 1966). Let the vector $\mathbf{w}_{\text{IO},\beta}$ represent these weights, which are defined in terms of the statistical properties of the images as

$$\mathbf{w}_{\text{IO},\beta} = \Sigma_{\beta}^{-1} \mathbf{s}. \quad (2)$$

The resulting Ideal observer strategy is implemented by comparing the weighted sum of the image, with mean background subtracted, to a detection threshold,

$$\begin{aligned} H_0 &: \text{if } \mathbf{w}_{\text{IO},\beta}^T \mathbf{g} < t_{\text{crit}} \\ H_1 &: \text{if } \mathbf{w}_{\text{IO},\beta}^T \mathbf{g} > t_{\text{crit}}. \end{aligned} \quad (3)$$

The value of the threshold, t_{crit} , determines the tradeoff between hits and false alarms. In principle this term should be set on the basis of outcome utilities. However, we leave it as a free parameter to be fit to the human observer data.

With human observer data, we obtain the equivalent contrast for the ideal observer by adjusting contrast and t_{crit} until the hit rate and false-alarm rate equal the human observer's. Let $C_{\text{Obs},\beta}^{\text{D}}$ be the target contrast used for the human observer study, and let $C_{\text{IO},\beta}^{\text{D}}$ be the equivalent ideal observer contrast. The efficiency of the observer is defined in terms of a squared ratio of contrast thresholds following Kersten (1987) as

$$\eta_{\text{obs},\beta}^{\text{D}} = \left(\frac{C_{\text{IO},\beta}^{\text{D}}}{C_{\text{Obs},\beta}^{\text{D}}} \right)^2. \quad (4)$$

Standard errors are determined by calculating efficiency on a session-by-session basis, and then computing the standard error across sessions.

Localization tasks

As mentioned in the Introduction, the theory we use for ideal observers in a free-localization task comes from the work of Khurd and Gindi (2005). Here we present a somewhat simplified derivation that is adequate for our purposes. In this case, we have a conditional probability of the data for every possible location of the target. Let s_l represent the profile of the target, when it is centered on the pixel with index l , which can be anywhere in

the search region (i.e., 128^2 possible locations). The conditional likelihood of the data given a particular target location is

$$p(\mathbf{g}|l) = \text{MVN}(\mathbf{s}_l, \Sigma_{\beta}). \quad (5)$$

The basis for localization by the ideal observer is the posterior distribution on possible locations, $p(l|\mathbf{g})$. For a uniform prior distribution on target locations, the posterior distribution is proportional to the likelihood. Under the Gaussian assumptions of our images, we have

$$p(l|\mathbf{g}) = N_{\mathbf{g}} e^{\mathbf{s}_l^T \Sigma_{\beta}^{-1} \mathbf{g}}, \quad (6)$$

where $N_{\mathbf{g}}$ is a normalization constant that ensures that $p(l|\mathbf{g})$ sums to 1 over all possible locations.

The task specifies that any response within 5 pixels of the target center is considered a correct response. The ideal observer will therefore choose the location that maximizes the probability of a correct answer. For each point under consideration, the ideal observer adds up the probabilities of all points within a 5-pixel radius, to get a final score for the location. The point with the largest score is then chosen as the ideal observer's response. It is worth noting that the ideal observer response at a given location is very similar to an ideal detector with spatial uncertainty (Pelli, 1985), where uncertainty is confined to the acceptance region around a given location.

The ideal observer decision function can be implemented using convolutions to speed up the computationally intensive steps. For example, the stationary nature of the noise covariance matrix allows the computations of $\mathbf{s}_l^T \Sigma_{\beta}^{-1} \mathbf{g}$ to be implemented by convolving the ideal observer template, defined in Equation 2, with the mean-subtracted background. Similarly the computation of the final score at each location in the image can be computed by convolving a disk of radius 5 pixels with the normalized posterior distribution in Equation 6. The recipe for computing the ideal observer begins with pre-computing the ideal observer filter by dividing the Fourier transform of the target by the power-spectrum of the noise. Then for each image, (1) this filter is used in a convolution after the mean background has been subtracted; (2) the result is exponentiated; (3) pixels outside the

search region are set to zero; (4) pixels in the search region are scaled so that they sum to 1; (5) the posterior is convolved with a disk of radius 5 pixels; and (6) the maximum point is chosen.

With a case-by-case ideal observer algorithm, the performance of the ideal observer is estimated to arbitrary accuracy using large sets of sample images. We use this approach to build LUTs of ideal observer performance as a function of target contrast. The LUTs for each β are determined in contrast increments of 0.01 from 0 until PC rises above 94%. The functions are plotted for each β in **Figure 4**. Each point is based on 2000 sample images, which results in standard errors that are less than 1% near the 80% correct level that is used in the experiments. Inverting these functions allows for us to determine the contrast threshold required by the ideal observer to achieve the specified level of PC. For an observer that achieves a proportion correct of PC_{Obs} in a localization task with a target contrast of $C_{Obs,\beta}^L$, efficiency is again defined (Kersten, 1987) as the squared contrast ratio

$$\eta_{Obs,\beta}^L = \left(\frac{C_{IO,\beta}^L(PC_{Obs})}{C_{Obs,\beta}^L} \right)^2. \quad (7)$$

Standard errors are determined by calculating efficiency on a session-by-session basis, and then computing the standard error across sessions.

The D&L task uses a similar process as the localization task, except that in the last step a threshold is applied. If the maximum score is above the detection threshold, the location of the score is

selected for detecting and localizing the target. If the maximum score is below the detection threshold, the ideal observer selects the “target-absent” response. For matching human observer data, the threshold contrast and detection criterion are adjusted to match the rate of correct detect-and-localize responses and the false positive (FP) rate. Efficiency is then calculated as the squared ratio of this contrast to the contrast used in the experiment, as in Equations 4 and 7.

CLASSIFICATION IMAGES

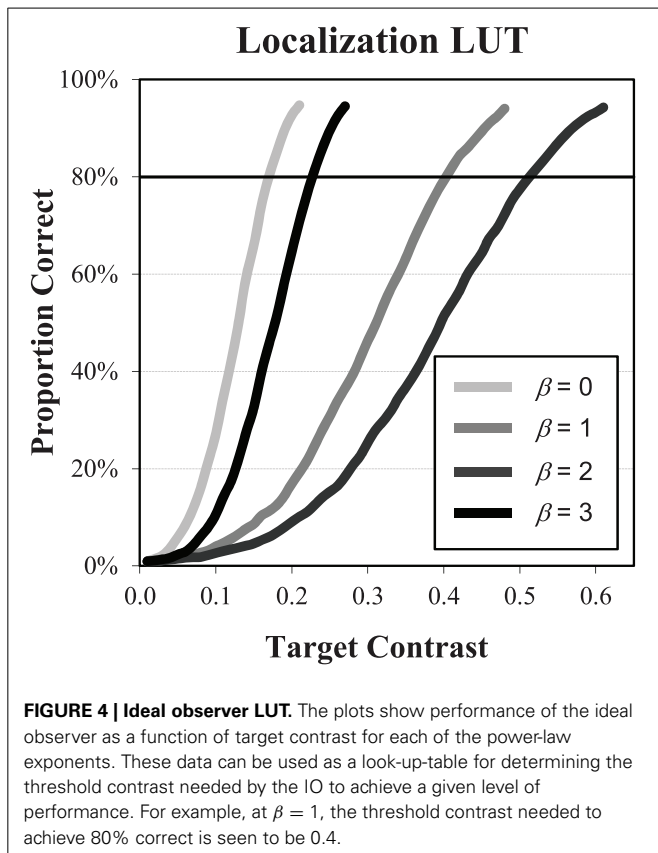
In addition to efficiency, we will use classification images as a way to investigate how visual processing affects task efficiency. This approach is straightforward for the detection tasks, where the classification image analysis has been well developed by Ahumada (2002) and others (Gold et al., 2000; Chauvin et al., 2005; Victor, 2005; Tjan and Nandy, 2006; Murray, 2011). Let \mathbf{n} represent the noise field for a given trial, with no target profile or mean background. Let us define the quantity \mathbf{q} as the product of the inverse covariance and the noise field, $\mathbf{q} = \Sigma_{\beta}^{-1}\mathbf{n}$. The classification image is given by

$$\mathbf{w}_{CI}^D = \bar{\mathbf{q}}_{FP} - \bar{\mathbf{q}}_{TN} + \bar{\mathbf{q}}_{TP} - \bar{\mathbf{q}}_{FN}, \quad (8)$$

where the $\bar{\mathbf{q}}$ are the average \mathbf{q} over the FP, true-negative (TN), true-positive (TP), and false-negative (FN) noise fields. Under the (strong) assumption of a linear template as the mechanism for detecting the target, the classification image will provide an unbiased estimate of the template. If the observer does not follow the linear assumption, the resulting classification image may be distorted, depending on the degree of violation (Ahumada, 2002).

Tjan and Nandy (2006) have analyzed discrimination tasks in the presence of target location uncertainty using classification images. Their approach utilizes the concept of a “clamped signal,” in which the noise field masking the target profile in an incorrect response is analyzed. This approach was found to work well in various two-class detection and discrimination tasks with targets that could be subject to spatial uncertainty. Additionally, Neri (2010) has used early static nonlinearities as a way to model performance in such tasks. In principle, our free-localization task can be considered a classification task with 128^2 possible response categories (and a somewhat ambiguous definition of a correct response that includes neighboring locations). However, in this work we have pursued a different approach for classification images in which the noise at the location of an incorrect response is used rather than the noise that masked the unchosen target. In this regard, our approach is similar to a previous study by Rajashekar et al. (2006) that used eye-tracking to estimate gaze-contingent classification images, as well as studies that have used the classification-image approach in multiple-alternative forced choice studies (Caspi et al., 2004; Eckstein et al., 2007; Dai and Micheyl, 2010).

Let \mathbf{n}^A represent a “response-aligned” noise field, in which the image noise field is shifted so that the location selected by the observer is translated to the center of the image. Let $\mathbf{q}^A = \Sigma_{\beta}^{-1}\mathbf{n}^A$, which is analogous to a response-aligned version of \mathbf{q} defined above. For classification images in the localization tasks, we use



the average of the response-aligned \mathbf{q} vectors when the subject incorrectly localizes (IL) the target

$$\mathbf{w}_{CI}^I = \bar{\mathbf{q}}_{FL}^A. \quad (9)$$

In these cases, the response is entirely driven by the form of the noise at the response location. We will see below that this leads to a strong classification image relative to detection, even though the detection task uses all noise fields in the image and this approach for the free-localization uses approximately 20% of the trials in which a false-localization response is given.

As a simple test of the classification image approach for localization tasks, we have used it to evaluate the ideal observer. **Figure 5A** shows the frequency weights of the ideal observer, derived analytically from $\Sigma_{\beta}^{-1}\mathbf{s}$. In **Figure 5B**, we see the estimated frequency weights for 2000 trials of the ideal observer using Equation 9, when the target contrast is set so that PC = 80%. While there are some areas of apparent bias, particularly at the lowest spatial frequencies for $\beta = 0$, there is generally good agreement between the actual frequency weights used to perform the task and the estimated weights.

RESULTS AND DISCUSSION

PSYCHOMETRIC FUNCTIONS

Contrast thresholds were determined for each subject in each condition from fitted psychometric functions. After an initial training of 5 runs of increasing difficulty totaling 210 trials, psychometric data was acquired in 20 runs of 50 trials at five different contrast levels for a total of 200 trials at each contrast level. The contrast levels used were determined from pilot data. Cumulative Gaussian distribution functions were fit to the proportion of correct responses over the range of contrasts, and contrast thresholds were determined from the contrast that produced 80% correct. An example of the psychometric functions (Subject 4, $\beta = 1$) is

shown in **Figure 6**. There was generally good agreement between the subject data and the cumulative Gaussian fitting function.

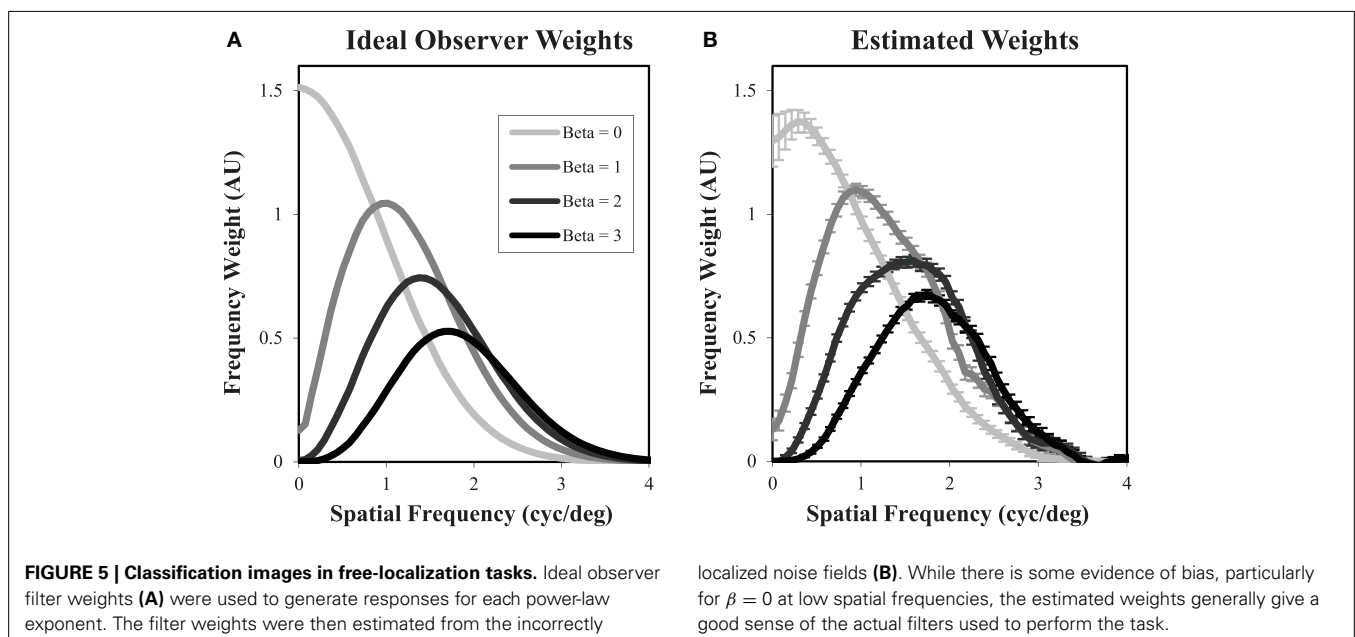
The average threshold contrast for each task and background type is plotted in **Figure 6B**. Thresholds within each task peak for $b = 2$. The thresholds are substantially higher for the localization task, with roughly a factor of two increase for each background.

CHARACTERIZING TASK PERFORMANCE

After each contrast threshold was determined from the psychometric data, subjects performed a total of 40 runs of 50 trials, for a total of 2000 trials at the subject's threshold contrast. Efficiency with respect to the ideal observer was estimated from this data. The efficiency results are described below in Detect-And-Localize Efficiency. Here, we will describe other measurements that provide additional information to characterize task performance.

Performance in the efficiency data is reasonably close to the nominal 80% correct levels derived from the psychometric functions. **Figure 7A** plots average PC across subjects from the efficiency data as a function of the power-law exponent of the background. Overall, PC values averaged 81.9% in the detection experiments and 80.3% in the localization experiments. The slight increases across subjects may be due to learning effects that occurred over the 2000 trials. The largest observed deviation from 80% correct for a single subject in a single condition was 7.3%. These results give us some confidence that efficiency was measured at contrasts near the actual 80% correct threshold.

While reaction time is not an endpoint of our study, this data is recorded as part of the experimental procedure. Reaction time is defined as the time from stimulus onset to the acquisition of a subject response. Median reaction times, given in **Table 1**, are mostly larger for the free localization task. This is not surprising since the subject need to search an area $6.7 \times 6.7^\circ$ in the localization task. Given the size of this area, the 48% average increase in reaction times seems rather modest. It is worth noting that the



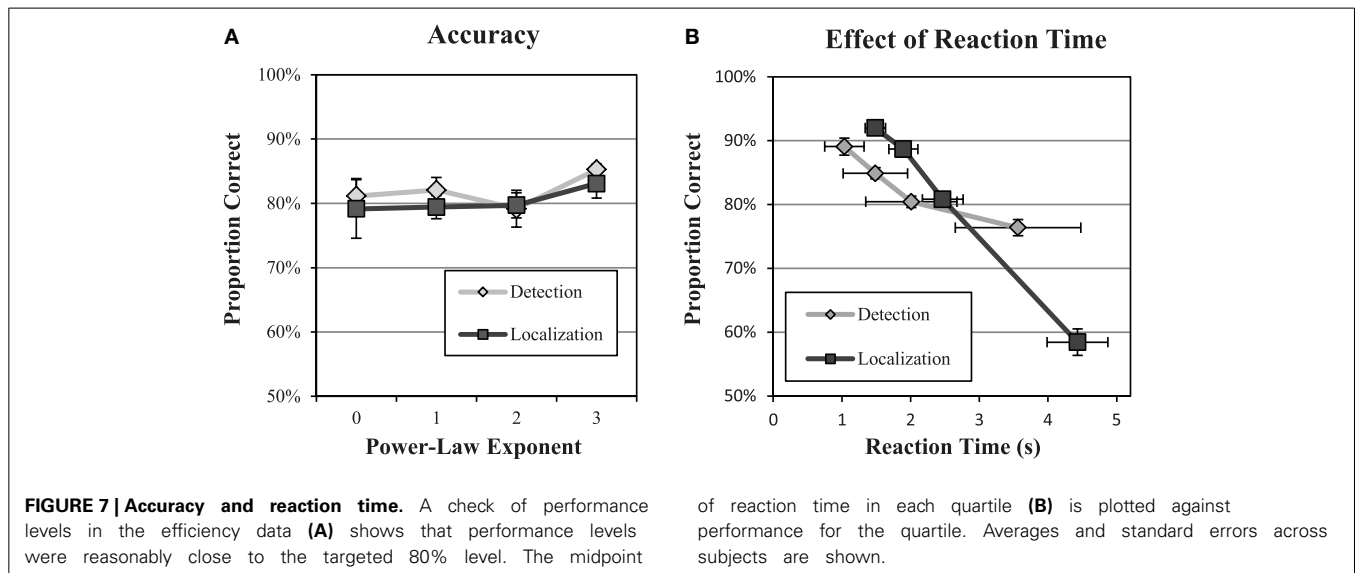
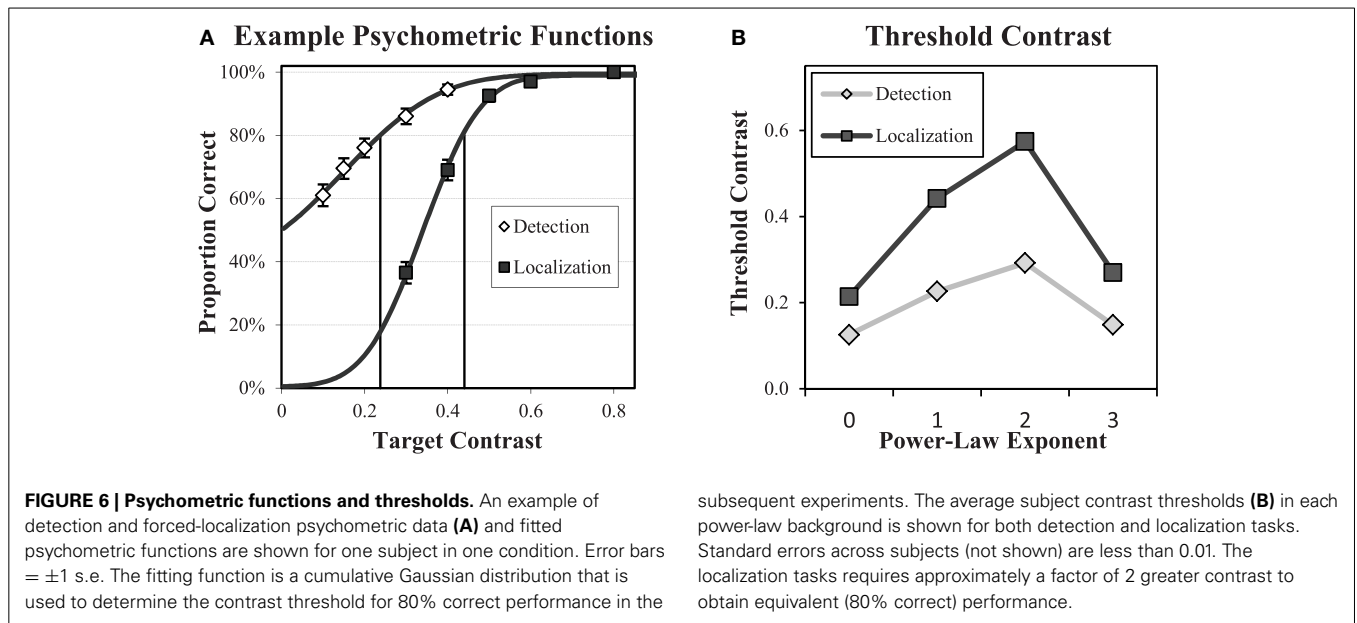


Table 1 | Reaction times.

Subjects	Detection RT	Localization RT	Rel. dif. (%)
S1	1.35	2.61	94
S2	1.11	1.80	62
S3	1.17	1.77	51
S4	3.97	2.85	-28
S5	1.01	1.65	64
Ave	1.72	2.13	48

Median reaction times (RTs) are given for each subject as well as the relative difference between the detection and localization tasks.

increase in median response times is not uniform over the subjects. One subject (S4) is markedly slower in the detection task.

It is also of interest to compare the effect of reaction time and performance as shown in a representative example in **Figure 7B**.

We divided the data into quartiles of 500 trials according to reaction time, and then computed proportion correct in each quartile. The figure plots proportion correct as a function of the median reaction time for the quartile. All subjects exhibited a similar trend of decreased performance with greater reaction times in both tasks. This finding is the opposite of what might be expected from a speed-accuracy tradeoff, where slower speeds allow for more effective task performance. However, decreased performance for longer reaction times has been found previously (Eckstein et al., 2001), and is thought to reflect the effects of a noise limited task where longer reaction times are associated with noise masks that make the task more difficult.

Unlike the detection task, the localization response requires careful positioning of the cursor using the mouse. The accuracy of this process has consequences both for overall accuracy in the task, if mis-positioning the cursor causes the localization response

Table 2 | Localization accuracy.

Abs. dev.	$\beta = 0$	$\beta = 1$	$\beta = 2$	$\beta = 3$	Uniform
Pixels	1.96 ± 0.08	1.69 ± 0.11	1.52 ± 0.14	1.41 ± 0.15	3.40
Degrees	0.102 ± 0.004	0.088 ± 0.006	0.079 ± 0.007	0.074 ± 0.008	0.180

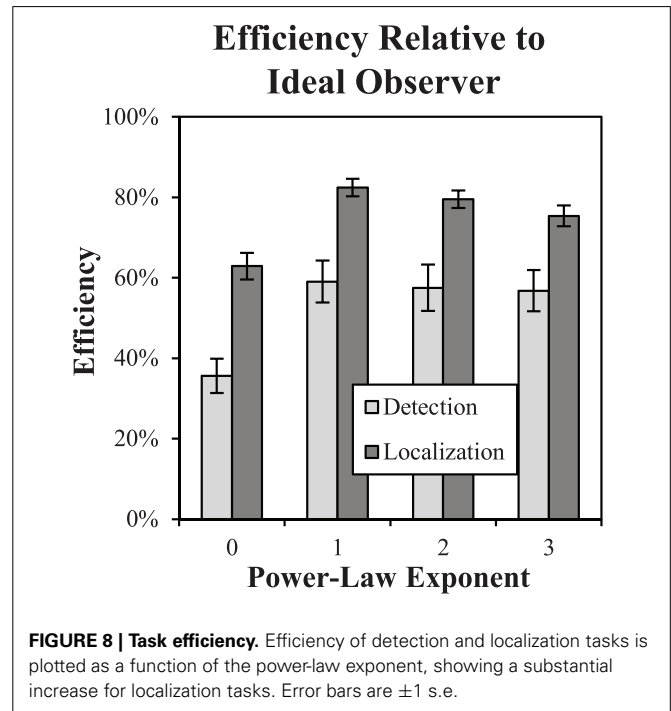
Average absolute deviation (\pm s.e. across subjects) of correct localization responses relative to the target center. Data is given both in pixel units as well as degrees of visual angle. For reference, the absolute deviation assuming a uniform distribution within the acceptance region is also given.

to fall outside the acceptance region, and for aligning the noise fields for the classification image analysis. To get some sense of the accuracy of the localization responses, we have evaluated the deviation of the responses, which is defined as the distance of the subject mouse clicks from the target location for responses that fall within the acceptance region of 5 pixels from the target center. **Table 2** gives the average deviation across subjects, in both pixels and degrees of visual angle, and well as the deviation assuming a uniform distribution of responses over the acceptance region. The deviations are all substantially smaller than the uniform distribution would predict, suggesting that there is considerable additional accuracy in the localization response. In addition, there is a consistent decrease in the deviation as β increases. The error represented by the absolute deviation contains both the effects of subject's misperception of the target center, as well as motor noise in the subject's response. Of these two, motor noise will be detrimental to the classification image methodology, since it will lead to misalignment of the selected noise fields. The observed deviations in **Table 2** act as an upper bound on motor noise in the subject responses, and suggest that these effects may be modest.

TASK EFFICIENCY

The primary performance result we are interested in for these studies is observer efficiency, as plotted in **Figure 8**. Efficiency with respect to the ideal observer appears to be substantially higher for localization tasks than detection tasks. A Two-Way ANOVA with the five subjects considered as replications finds significant effects for both the task [$F_{(1,32)} = 63.4, p < 0.0001$] as well as the background exponent [$F_{(3, 32)} = 11.7, p < 0.0001$]. The interaction between task and exponent was not found to be significant [$F_{(3, 32)} = 0.39, p > 0.76$]. It should be noted that average efficiency near 80% for β -values of 1–3, is considered quite high. In the classic experiments by Burgess et al. (1981), efficiency as high as 70% was observed with averages across observers closer to 50%. These experiments used a spatial forced choice methodology and white noise ($\beta = 0$). Experiments in low-pass noise similar to the $\beta = 3$ condition used here (Abbey and Eckstein, 2007), found efficiency in the 40 to 60% range. These are consistent with our findings in the detection task, all of which utilize aperiodic “bump” targets. Efficiency of oscillatory targets are typically lower (Legge et al., 1987). The increased efficiency we find in the localization tasks represents a substantial gain from these fixed-location tasks, and suggests that subjects have little room for sub-optimal computations in performing these tasks.

Efficiency is somewhat lower for $\beta = 0$ in both the detection and localization tasks. We consider this case further in Efficiency of Image Processing for $\beta = 0$. below. We also note that these



efficiency values appear to be relatively stable with the acceptance radius. We observed less than a 1% difference in observed efficiency varying the acceptance region from 4 pixels to 7.

These efficiency results show that in spite of larger thresholds for the free-localization relative to detection, as shown in **Figure 6B**, overall efficiency is substantially higher. This means that thresholds for the ideal observer increase proportionally even more than the human subjects' did. Our findings are consistent with the uncertainty hypothesis (Tanner, 1961; Pelli, 1985), which posits imperfect use of the location cues in detection tasks, and leave the observer with some residual uncertainty regarding the location of the target that can reduce performance. The ideal observer is not subject to this phenomenon, which results in a somewhat lower contrast threshold. In the free-localization task, where uncertainty is intrinsic to the task, the ideal observer does not have the advantage of precise knowledge of location, and contrast thresholds rise relative to the human observers as a result. However, other explanations for the large difference in efficiency are possible. For example, detection tasks require that the subject use some sort of criterion that dichotomizes responses. If this criterion drifts or is prone to jitter, performance will be reduced. This possibility motivated the detect-and-localize study.

DETECT-AND-LOCALIZE EFFICIENCY

A subset of three subjects performed the detect-and-localize experiments, which were all run after the detection and localization data were acquired. Threshold target contrast from the localization tasks were used as target contrasts for these experiments. The proportion of correct responses dropped modestly from an average of 80.3% in the localization tasks to 76% in the D&L tasks. **Figure 9** plots shows the efficiency data for the detection task, localization task, and D&L tasks as a function of the power-law exponent for the subset of subjects that participated in all three studies. The average efficiency values for the D&L are all well above both the detection and localization tasks. In fact several observed values are near 90% efficiency, which is again quite high for tasks masked by luminance noise. These findings are close to the highest reported efficiency we are aware of for visual tasks limited by noise (Manjeshwar and Wilson, 2001).

EFFICIENCY OF IMAGE PROCESSING FOR $\beta = 0$

Figure 8 shows reduced efficiency in the $\beta = 0$ condition of both the detection and localization tasks. After finding this effect, we were interested in whether it might be mitigated by processing the images to have a background power-law of $\beta = 2$, where efficiency was generally better. As described above in Stimulus and Display Properties, this is accomplished by filtering the images with a kernel that has a $1/f$ spectrum, which will modify both the background statistics and the target profile, as shown in **Figure 3**.

Figure 10 shows that the effect of processing is to bring efficiency in the $\beta = 0$ condition up to 66% in the detection task

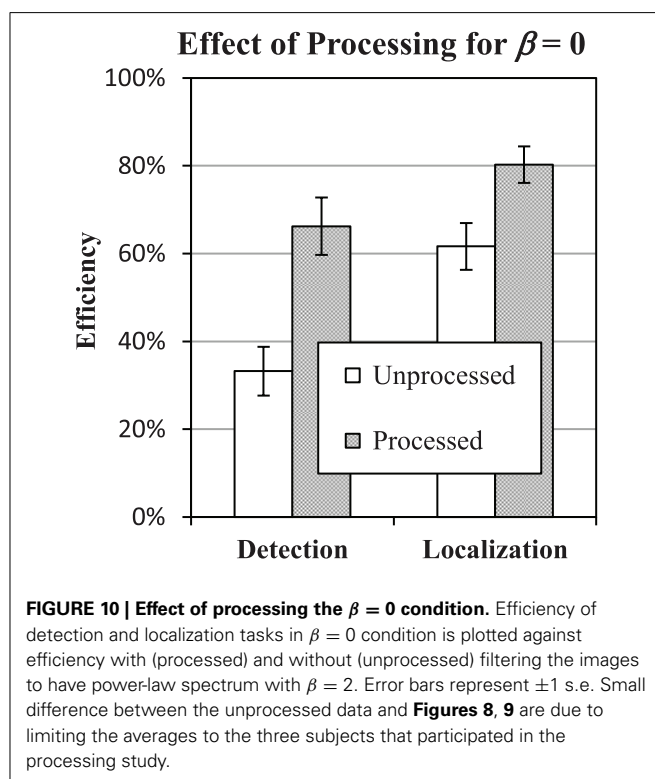
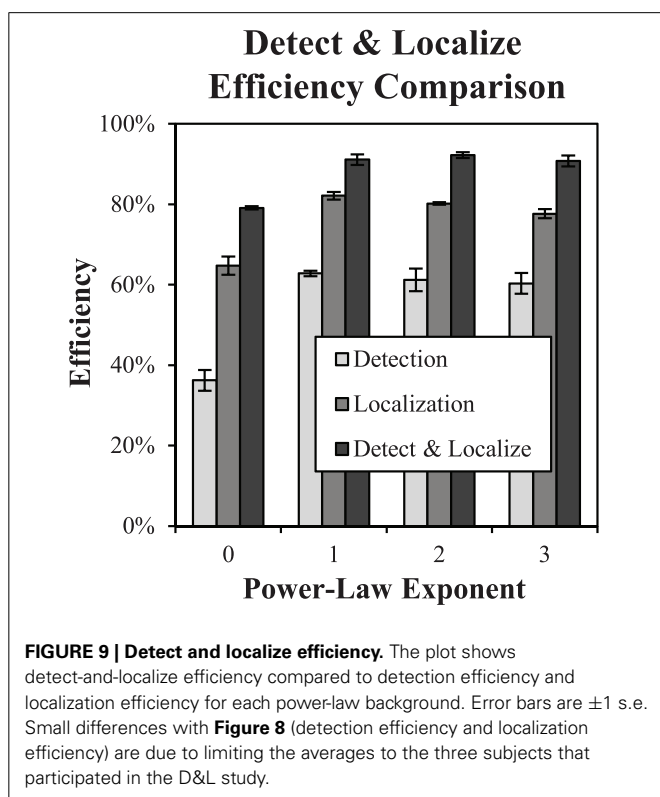
and 80% in the localization task. These levels are consistent with efficiency levels found for β in the range of 1–3.

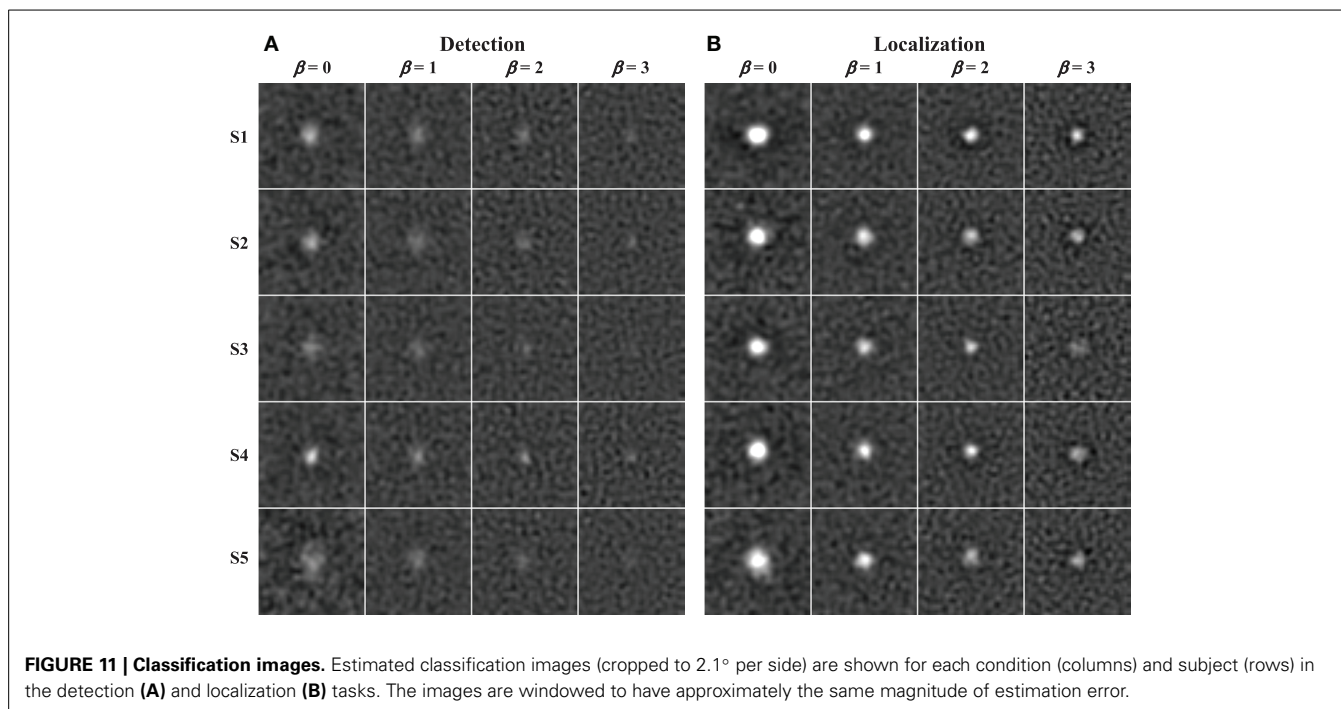
CLASSIFICATION IMAGES

Figure 11 shows the classification images for each subject in each background condition for both the detection and localization tasks. The images are cropped to the central 2.1° of visual angle (40 pixels). Outside of this area, there are no discernable features beyond what appears to be estimation error in the classification images. To mitigate the effects of noise, the classification images have been low-pass filtered with a 4th-order Butterworth filter, with the roll-off parameter set to 5.6 cyc/deg (0.29 cyc/pixel). This was well beyond the point at which the spatial frequency plots below appear to decay to zero.

The images in **Figure 11** were windowed to have approximately the same mean background and error magnitude. Thus, the intensity of the features in the observed classification images gives some sense of their signal-to-noise ratio (SNR). The generally brighter appearance of classification images in the localization tasks relative to the corresponding detection tasks suggests that search process may lead to methodological advantages for estimating classification images, even though the localization classification images are estimated from approximately 20% of the subjects responses in which an incorrect localization response is given. There also appears to be some differences in the intensity of the classification images going from $\beta = 0$ to $\beta = 3$, and there are clearly individual differences between subjects.

In addition to the overall intensity of the classification images, we are also interested in the profile of these decision weights.





Based on previous experience, we find that differences between classification images in different conditions are most clearly depicted for radial averages in the spatial-frequency domain. **Figure 12** plots the classification frequency weights averaged over subjects and normalized so that the weight at the peak frequency is 1. To reduce the effects of noise in the classification images, a Butterworth spatial window with a cutoff of 1.05° (20 pixels) was applied before the Fourier transform and radial averaging. For reference, we have plotted the classification weights of the ideal observer as well. In all conditions, the average frequency weights assume a bandpass form, peaking at frequencies between 0.7 cyc/deg and 1.6 cyc/deg as β goes from 0 to 3. As has been found previously (Abbey and Eckstein, 2007; Conrey and Gold, 2009), the classification weights here give evidence of visual processing that is changing with the different power-law textures in the background. But this process is not as extreme as the adaptation that occurs in the ideal observer, where peak frequencies move from 0 to 1.7 cyc/deg.

In the $\beta = 0$ condition, we observe substantial underweighting of low spatial frequencies relative to the ideal observer. Of interest for the comparison of detection and localization tasks, there is less low-frequency suppression in the localization task compared to the detection task. As β increases, we see that the low-frequency profiles come together, but now they do not suppress low frequencies as much as the ideal observer. Also, as β increases, the classification weight frequency profiles begin to diverge at higher spatial frequencies above the peak values. Here the profiles from the localization tasks have higher weights that are closer to the ideal observer.

Figure 13 shows the frequency plots in the $\beta = 0$ condition using responses from the processed and unprocessed data averaged over the three subjects that participated in these studies.

The plots show processing effectively modifies the weighting profile that subjects use. In both tasks, the effect of processing is to increase the low-frequency weighting so that the average subject classification weights more closely match the ideal observer. Thus, the classification image profiles give a visual mechanism for the improved efficiency found in **Figure 10**.

SUMMARY AND CONCLUSIONS

We find human observers substantially improve in performance relative to the ideal observer in free-localization tasks compared to fixed-location detection tasks, in spite of increased contrast thresholds. This occurs in all four power-law textures that were investigated. In a follow-up study investigating a detect-and-localize task, we find the highest measured efficiency in our experiments, suggesting that our efficiency results are not simply a consequence of a general inability to maintain detection criteria. Our findings are consistent with spatial uncertainty as a limiting effect in the presence of location cues.

While it is clear from the classification images that observers are able to tune their visual templates to the statistics of the noise in the images, there is also evidence that this process is limited in both fixed and free-localization tasks. Despite a common target profile, the different power-law textures require different frequency tuning to achieve optimal performance. We do find some evidence of such tuning in the classification images estimated from the subject responses. Peak spatial frequency weights change by roughly a factor of two going from a power-law exponent of $\beta = 0$ to $\beta = 3$ (0.72–1.59 cyc/deg). However, on average the subject frequency weights exhibited some clear departures from optimal tuning as defined by the ideal observer. At $\beta = 0$, we find human observer frequency weights shifted to higher spatial frequencies relative to the ideal observer. For $\beta > 0$,

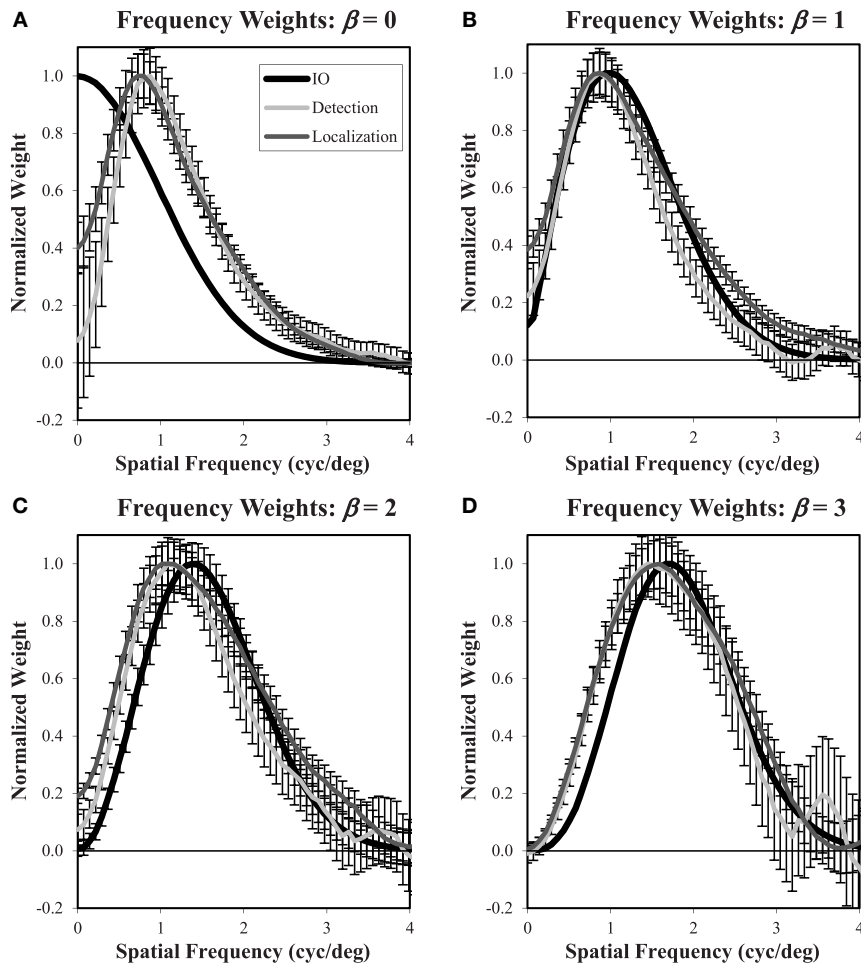


FIGURE 12 | Frequency weights derived from Classification images. Radial frequency profiles are shown for each of the four power-law textures (A–D) with normalization so that the maximum weight is one.

The ideal observer profile is derived from theory. The detection and localization plots are averaged across the five subjects. Error bars are ± 1 s.e. averaged across subjects. The legend (A) applies to all plots.

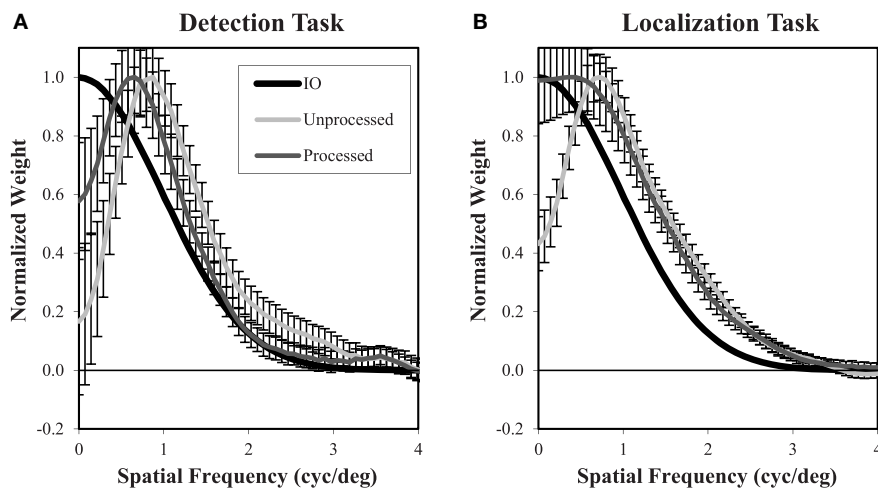


FIGURE 13 | Frequency weights for processed and unprocessed images. These plots are similar to Figure 12 and show estimated weights from the $\beta = 0$ images using the responses to processed and unprocessed images. In

both the detection (A) and localization (B) tasks, the effect of image processing is to increase the estimated weights at low spatial frequency, bringing them closer to the ideal observer weights. The legend (A) applies to both plots.

human-observer classification weights peak at lower spatial frequencies than the ideal observer.

Frequency tuning of subjects in the white-noise condition was most different from the ideal observer. This condition also led to the lowest efficiency in performance. Since $\beta = 0$ was the power-law exponent furthest from that found in natural scenes ($\beta = 2$), this finding is consistent with the idea that the human visual system is somewhat adapted to the statistics of natural images. The follow-up study investigating processed images supports this connection by finding uniformly improved performance when the white-noise images were filtered to have $\beta = 2$. Filtering the images was also seen to effectively improve frequency tuning of the subjects in the white-noise condition.

While we do not attempt to explicitly model the visual system to explain our findings, we do believe that our findings may be relevant in such attempts, for the same reasons given originally by Burgess (Burgess et al., 1981). The finding of high efficiency in free-localization and detect-and-localize tasks suggest that models of vision in these tasks cannot be very different, at a computational level, from the ideal observer, and thus may provide a valuable constraint to such efforts in future studies.

ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the U.S. National Institutes of Health (NEI-EY015925).

REFERENCES

- Abbey, C. K., and Eckstein, M. P. (2007). Classification images for simple detection and discrimination tasks in correlated noise. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 24, B110–B124. doi: 10.1364/JOSAA.24.00B110
- Abbey, C. K., Zemp, R. J., Liu, J., Lindfors, K. K., and Insana, M. F. (2006). Observer efficiency in discrimination tasks simulating malignant and benign breast lesions imaged with ultrasound. *IEEE Trans. Med. Imaging* 25, 198–209. doi: 10.1109/TMI.2005.862205
- Ahumada, A. J. Jr. (2002). Classification image weights and internal noise level estimation. *J. Vis.* 2, 121–131. doi: 10.1167/2.1.8
- Barlow, H. (1977). Retinal and central factors in human vision limited by noise. *Vertebrate Photoreception* 337, C358.
- Barlow, H. B. (1978). The efficiency of detecting changes of density in random dot patterns. *Vision Res.* 18, 637–650. doi: 10.1016/0042-6989(78)90143-8
- Barlow, H. B., and Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Res.* 19, 783–793. doi: 10.1016/0042-6989(79)90154-8
- Bochud, F. O., Abbey, C. K., and Eckstein, M. P. (2004). Search for lesions in mammograms: statistical characterization of observer responses. *Med. Phys.* 31, 24–36. doi: 10.1118/1.1630493
- Bochud, F. O., Valley, J. F., Verdun, F. R., Hessler, C., and Schnyder, P. (1999). Estimation of the noisy component of anatomical backgrounds. *Med. Phys.* 26, 1365–1370. doi: 10.1118/1.598632
- Burgess, A. E., and Ghandeharian, H. (1984). Visual signal detection. II. Signal-location identification. *J. Opt. Soc. Am. A* 1, 906–910. doi: 10.1364/JOSAA.1.000906
- Burgess, A. E., Jacobson, F. L., and Judy, P. F. (2001). Human observer detection experiments with mammograms and power-law noise. *Med. Phys.* 28, 419–437. doi: 10.1118/1.1355308
- Burgess, A. E., Wagner, R. F., Jennings, R. J., and Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science* 214, 93–94. doi: 10.1126/science.7280685
- Burton, G. J., and Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Appl. Opt.* 26, 157–170. doi: 10.1364/AO.26.000157
- Caspi, A., Beutter, B. R., Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Proc. Natl. Acad. Sci. U.S.A.* 101, 13086–13090. doi: 10.1073/pnas.0305329101
- Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., and Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *J. Vis.* 5, 659–667. doi: 10.1167/5.9.1
- Chen, L., Abbey, C. K., and Boone, J. M. (2013). Association between power law coefficients of the anatomical noise power spectrum and lesion detectability in breast imaging modalities. *Phys. Med. Biol.* 58, 1663–1681. doi: 10.1088/0031-9155/58/6/1663
- Chen, L., Abbey, C. K., Nosrati, A., Lindfors, K. K., and Boone, J. M. (2012). Anatomical complexity in breast parenchyma and its implications for optimal breast imaging strategies. *Med. Phys.* 39, 1435–1441. doi: 10.1118/1.3685462
- Conrey, B., and Gold, J. M. (2009). Pattern recognition in correlated and uncorrelated noise. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 26, B94–B109. doi: 10.1364/JOSAA.26.000B94
- Dai, H., and Michéyl, C. (2010). Psychophysical reverse correlation with multiple response alternatives. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 976–993. doi: 10.1037/a0017171
- Eckstein, M. P., Beutter, B. R., Pham, B. T., Shimozaki, S. S., and Stone, L. S. (2007). Similar neural representations of the target for saccades and perception during search. *J. Neurosci.* 27, 1266–1270. doi: 10.1523/JNEUROSCI.3975-06.2007
- Eckstein, M. P., Beutter, B. R., and Stone, L. S. (2001). Quantifying the performance limits of human saccadic targeting during visual search. *Perception* 30, 1389–1401. doi: 10.1068/p3128
- Engstrom, E., Reiser, I., and Nishikawa, R. (2009). Comparison of power spectra for tomosynthesis projections and reconstructed images. *Med. Phys.* 36, 1753–1758. doi: 10.1118/1.3116774
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4, 2379–2394. doi: 10.1364/JOSAA.4.002379
- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychol. Rev.* 96, 267–314. doi: 10.1037/0033-295X.96.2.267
- Gold, J. M., Murray, R. F., Bennett, P. J., and Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Curr. Biol.* 10, 663–666. doi: 10.1016/S0960-9822(00)00523-6
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York, NY: Wiley.
- Insana, M. F., and Hall, T. J. (1994). Visual detection efficiency in ultrasonic imaging: a framework for objective assessment of image quality. *J. Acoust. Soc. Am.* 95, 2081. doi: 10.1121/1.408669
- Kersten, D. (1987). Statistical efficiency for the detection of visual noise. *Vision Res.* 27, 1029–1040. doi: 10.1016/0042-6989(87)90016-2
- Khurd, P., and Gindi, G. (2005). Decision strategies that maximize the area under the LROC curve. *IEEE Trans. Med. Imaging* 24, 1626–1636. doi: 10.1109/TMI.2005.859210
- Khurd, P., Liu, B., and Gindi, G. (2010). Ideal AFROC and FROC observers. *IEEE Trans. Med. Imaging* 29, 375–386. doi: 10.1109/TMI.2009.2031840
- Legge, G. E., Kersten, D., and Burgess, A. E. (1987). Contrast discrimination in noise. *J. Opt. Soc. Am. A* 4, 391–404. doi: 10.1364/JOSAA.4.000391
- Liu, B., Zhou, L., Kulkarni, S., and Gindi, G. (2009). The efficiency of the human observer for lesion detection and localization in emission tomography. *Phys. Med. Biol.* 54, 2651–2666. doi: 10.1088/0031-9155/54/9/004
- Manjeshwar, R. M., and Wilson, D. L. (2001). Hyperefficient detection of targets in noisy images. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 18, 507–513. doi: 10.1364/JOSAA.18.000507
- Metheany, K. G., Abbey, C. K., Packard, N., and Boone, J. M. (2008). Characterizing anatomical variability in breast CT images. *Med. Phys.* 35, 4685–4694. doi: 10.1118/1.2977772
- Murray, R. F. (2011). Classification images: a review. *J. Vis.* 11, 1–25. doi: 10.1167/11.5.2
- Myers, K. J., Barrett, H. H., Borgstrom, M. C., Patton, D. D., and Seeley, G. W. (1985). Effect of noise correlation on detectability of disk signals in medical imaging. *J. Opt. Soc. Am. A* 2, 1752–1759. doi: 10.1364/JOSAA.2.001752
- Neri, P. (2010). Visual detection under uncertainty operates via an early static, not late dynamic, non-linearity. *Front. Comput. Neurosci.* 4:151. doi: 10.3389/fncom.2010.00151
- Park, S., Clarkson, E., Kupinski, M. A., and Barrett, H. H. (2005). Efficiency of the human observer detecting random signals in random backgrounds. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 22, 3–16. doi: 10.1364/JOSAA.22.000003

- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *J. Opt. Soc. Am. A* 2, 1508–1532. doi: 10.1364/JOSAA.2.001508
- Rajashekar, U., Bovik, A. C., and Cormack, L. K. (2006). Visual search in noise: revealing the influence of structural cues by gaze-contingent classification image analysis. *J. Vis.* 6, 379–386. doi: 10.1167/6.4.7
- Siewerdsen, J. H., and Jaffray, D. A. (2000). Optimization of x-ray imaging geometry (with specific application to flat-panel cone-beam computed tomography). *Med. Phys.* 27, 1903–1914. doi: 10.1118/1.1286590
- Tanner, W. P. Jr. (1961). Physiological implications of psychophysical data. *Ann. N.Y. Acad. Sci.* 89, 752–765. doi: 10.1111/j.1749-6632.1961.tb20176.x
- Tjan, B. S., Braje, W. L., Legge, G. E., and Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Res.* 35, 3053–3069. doi: 10.1016/0042-6989(95)00070-G
- Tjan, B. S., and Nandy, A. S. (2006). Classification images with uncertainty. *J. Vis.* 6, 387–413. doi: 10.1167/6.4.8
- Victor, J. D. (2005). Analyzing receptive fields, classification images and functional images: challenges with opportunities for synergy. *Nat. Neurosci.* 8, 1651–1656. doi: 10.1038/nn1607
- Wagner, R. F., and Brown, G. G. (1985). Unified SNR analysis of medical imaging systems. *Phys. Med. Biol.* 30, 489–518. doi: 10.1088/0031-9155/30/6/001
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 27 January 2014; paper pending published: 06 March 2014; accepted: 02 April 2014; published online: 01 May 2014.
- Citation: Abbey CK and Eckstein MP (2014) Observer efficiency in free-localization tasks with correlated noise. *Front. Psychol.* 5:345. doi: 10.3389/fpsyg.2014.00345
- This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.
- Copyright © 2014 Abbey and Eckstein. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.