# Optimal Eye Movement Strategies in Visual Search (Supplement)

Jiri Najemnik and Wilson S. Geisler
*Center for Perceptual Systems and Department of Psychology, University of Texas at Austin, Austin TX 78712*

Here we derive the ideal searcher for the case of dynamic (temporally uncorrelated) external and internal noise, and we describe (without proof) the ideal searcher for the case of static (temporally correlated) external noise and dynamic (temporally uncorrelated) internal noise. As a check on the accuracy of the formulas for these ideal searchers, they were derived via two separate methods: a direct derivation from the joint probability distributions (WG) and a derivation based upon the theory of Kalman filtering (JN).

**The visibility map**

To derive the ideal searcher it is necessary to consider the meaning of the visibility map in more detail. To do this we describe the ideal detector for a known sine wave target presented at a known location in a single interval forced choice task.

First, consider an ideal detector with direct access to the retinal image; that is, an ideal detector for a non-foveated visual system. Each trial of a single interval forced choice task consists either of background noise alone or background noise plus the sine wave target. The ideal detector multiplies the retinal image with a template of the sine wave target and then integrates the product to obtain a template response $W$ (i.e., the template response is the cross correlation of the target with the retinal stimulus). The magnitude of this template response is then compared to a criterion; the optimal behavior is to respond "yes" if the template response exceeds the criterion and "no" otherwise. The accuracy of this ideal detector is determined by the signal-to-noise ratio, $d'$, which is the average difference in the template response to the background plus target and background, divided by the standard deviation of the template response[1,2]. The expected value of the template response to background plus target is proportional to the target contrast and the variance of the template response is proportional to the noise contrast power of the background, and hence

$$d'(c,e_n)^2 = \frac{Ac^2}{Be_n} \tag{S1}$$

where, $c$ is the RMS contrast of the target, $e_n$ is the contrast power of the noise background, and $A$ and $B$ are proportionality constants.

Now consider an ideal detector in a foveated visual system. In this case, the ideal detector does not have direct access to the retinal image, but instead to a representation that is degraded by variable spatial resolution and neural noise. Reduced spatial resolution due to spatial filtering will effectively reduce the contrast of a sine wave target, but will have little effect on the shape of the target, thus the same template can be used in regions of reduced resolution, although the responses to the target and background will be smaller. (We have verified that for our target and for eccentricities as large as the radius of our display, the appropriate template shape changes negligibly for transfer functions that match human contrast sensitivity functions.) Furthermore, the neural noise will add a term $C(c, e_n, \varepsilon)$ to the variance of the template responses. In general, this neural noise term may depend on the target contrast, the background noise power, and the eccentricity. Therefore, the visibility map of an ideal detector with a foveated visual system is given by

$$d'(c, e_n, \varepsilon)^2 = \frac{A(\varepsilon)c^2}{B(\varepsilon)e_n + C(c, e_n, \varepsilon)} \tag{S2}$$

where the proportionality factors on the template responses to the target and background now vary with eccentricity. Without loss of generality this formula can be simplified by dividing numerator and denominator by $A(\varepsilon)$:

$$d'(c, e_n, \varepsilon)^2 = \frac{c^2}{\alpha e_n + \beta(c, e_n, \varepsilon)} \tag{S3}$$

where $\alpha = B(\varepsilon)/A(\varepsilon)$ and $\beta(c, e_n, \varepsilon) = C(c, e_n, \varepsilon)/A(\varepsilon)$. Because the same shaped template is used at different eccentricities, the value of $\alpha$ is a constant that does not change with eccentricity. Equation (S3) gives the visibility map of the ideal detector, once the values of $\alpha$ and $\beta$ are specified for all eccentricities. We note, however, that in the case of dynamic (temporally uncorrelated) external and internal noise, the two noise sources are of the same type and so their combined effect is given directly by the psychophysically measured visibility map $d'(c, e_n, \varepsilon)$. In the case of static external noise and dynamic internal noise, the two noises have different effects and so it is necessary to specify $\alpha$ and $\beta(c, e_n, \varepsilon)$ separately (see later).

**Ideal searcher for dynamic external and internal noise**

Let $n$ be the number of potential target locations, and let $W_{ik(t)}$ be the Gaussian-distributed template response from the $i^{th}$ potential target location, at fixation $t$, with fixation location $k(t)$. Without loss of generality, we can set the expected value of $W_{ik(t)}$ to 0.5, when the target is present at that location, and to -0.5, otherwise. This has no affect on the predictions as long as we set the standard deviation of the template responses so as to preserve the value of $d'$ specified by the visibility map. Thus, the mean value of the template response is

$$u_{ik(t)} = \begin{cases} 0.5 & i = \text{ target location} \\ -0.5 & i \neq \text{ target location} \end{cases} \tag{S4}$$

and the variance of the template response is

$$\sigma^2_{ik(t)} = \frac{1}{d'^{\,2}_{ik(t)}} \tag{S5}$$

***Updating posterior probabilities***. We now derive the following formula, which gives the posterior probability that the target is at location $i$ after $T$ fixations:

$$p_i(T) = \frac{prior(i)\exp\left[\sum_{t=1}^{T} d'^{\,2}_{ik(t)} W_{ik(t)}\right]}{\sum_{j=1}^{n} prior(j)\exp\left[\sum_{t=1}^{T} d'^{\,2}_{jk(t)} W_{jk(t)}\right]} \tag{S6}$$

This formula (which is equation (1) in the main body of the report) shows that to perfectly integrate information across fixations it is sufficient to keep a running weighted sum of the template responses at each potential target location, where the weights are the squares of the visibility map values.

Note first that all the template responses collected at fixation $t$ can be represented by a vector, $\mathbf{W}(t) = \left(W_{1k(t)}, \cdots, W_{nk(t)}\right)$. Using Bayes' formula, it follows immediately that the posterior probability that the target is at location $i$ after $T$ fixations is given by

$$p_i(T) = \frac{prior(i)\, p\left(\mathbf{W}(1), \cdots, \mathbf{W}(T) \middle| i\right)}{\sum_{j=1}^{n} prior(j)\, p\left(\mathbf{W}(1), \cdots, \mathbf{W}(T) \middle| j\right)} \tag{S7}$$

Both the external noise and internal noise are statistically independent over time, and for potential target locations separated sufficiently in space (as in the present search experiments) the external noise and internal noise are also statistically independent over space. Thus,

$$p_i(T) = \frac{prior(i)\prod_{t=1}^{T}\prod_{q=1}^{n} p\left(W_{qk(t)} \middle| i\right)}{\sum_{j=1}^{n} prior(j)\prod_{t=1}^{T}\prod_{q=1}^{n} p\left(W_{qk(t)} \middle| j\right)}$$

Given equations (S4) and (S5), and the fact that the template responses are Gaussian distributed, we have

$$p_i(T) = \frac{prior(i)\frac{(2\pi)^{-T/2}}{\prod_{t=1}^{T}\sigma_{ik(t)}}\exp\left(-\sum_{t=1}^{T}\frac{1}{2}\frac{\left(W_{ik(t)}-0.5\right)^2}{\sigma_{ik(t)}^{2}}\right)\prod_{q\neq i}\left(\frac{(2\pi)^{-T/2}}{\prod_{t=1}^{T}\sigma_{qk(t)}}\exp\left(-\sum_{t=1}^{T}\frac{1}{2}\frac{\left(W_{qk(t)}+0.5\right)^2}{\sigma_{qk(t)}^{2}}\right)\right)}{\sum_{j=1}^{n}prior(j)\frac{(2\pi)^{-T/2}}{\prod_{t=1}^{T}\sigma_{jk(t)}}\exp\left(-\sum_{t=1}^{T}\frac{1}{2}\frac{\left(W_{jk(t)}-0.5\right)^2}{\sigma_{jk(t)}^{2}}\right)\prod_{q\neq j}\left(\frac{(2\pi)^{-T/2}}{\prod_{t=1}^{T}\sigma_{qk(t)}}\exp\left(-\sum_{t=1}^{T}\frac{1}{2}\frac{\left(W_{qk(t)}+0.5\right)^2}{\sigma_{qk(t)}^{2}}\right)\right)}$$

Dividing the numerator and denominator by the numerator gives,

$$p_i(T) = \frac{1}{1+\sum_{j\neq i}\frac{prior(j)}{prior(i)}\frac{\exp\left(-\sum_{t=1}^{T}\frac{1}{2}\frac{\left(W_{jk(t)}-0.5\right)^2}{\sigma_{jk(t)}^{2}}\right)}{\exp\left(-\sum_{t=1}^{T}\frac{1}{2}\frac{\left(W_{ik(t)}-0.5\right)^2}{\sigma_{ik(t)}^{2}}\right)}\frac{\exp\left(-\sum_{t=1}^{T}\frac{1}{2}\frac{\left(W_{ik(t)}+0.5\right)^2}{\sigma_{ik(t)}^{2}}\right)}{\exp\left(-\sum_{t=1}^{T}\frac{1}{2}\frac{\left(W_{jk(t)}+0.5\right)^2}{\sigma_{jk(t)}^{2}}\right)}}$$

After combining terms we have,

$$p_i(T) = \frac{1}{1+\sum_{j\neq i}\frac{prior(j)}{prior(i)}\exp\left(\sum_{t=1}^{T}\frac{W_{jk(t)}}{\sigma_{jk(t)}^{2}}-\sum_{t=1}^{T}\frac{W_{ik(t)}}{\sigma_{ik(t)}^{2}}\right)}$$

Finally, equation (S6) is obtained by using equation (S5) to substitute for the variances, and then by multiplying the numerator and denominator by $prior(i)\exp\left[\sum_{t=1}^{T}d'^{\,2}_{ik(t)}W_{ik(t)}\right]$.

***Selecting next fixation location***. To compute the optimal next fixation point, $k_{opt}(T+1)$, the ideal searcher considers each possible next fixation and picks the location that, given its knowledge of the current posterior probabilities and the visibility map, will maximize the probability of correctly identifying the location of the target after the next fixation is made:

$$k_{opt}(T+1) = \underset{k(T+1)}{\arg\max}\left(p\left(C\middle|k(T+1)\right)\right)$$

Conditioning on the target location gives the following equation, which is equation (2) in the main body of the report:

$$k_{opt}(T+1) = \underset{k(T+1)}{\arg\max}\left(\sum_{i=1}^{n} p_i(T)\, p\left(C \middle| i, k(T+1)\right)\right) \tag{S8}$$

Here we derive a version of this equation that is practical to evaluate in computer simulations.

The probability of each possible target location is given by equation (S6), thus our job is to derive an expression for $p\left(C \middle| i, k(T+1)\right)$, the probability of being correct given that the true target location is $i$, and the location of the next fixation is $k(T+1)$. Let $Z_j$ be the hypothetical template response from the $j^{th}$ location after the next fixation is made; i.e., $Z_j = W_{jk(T+1)}$. After making the next fixation, the decision rule that would maximize accuracy would be to pick the location with the maximum posterior probability. If one uses that decision rule, then the percent correct is equal to the probability that the posterior probability at location $i$ will be greater than that at all other locations:

$$p\left(C \middle| i, k(T+1)\right) = p\left(p_i(T+1) \geq p_1(T+1), \cdots, p_i(T+1) \geq p_n(T+1) \middle| i, k(T+1)\right)$$

or equivalently,

$$p\left(C \middle| i, k(T+1)\right) = p\left(L_{i1} \geq 1, \cdots, L_{in} \geq 1 \middle| i, k(T+1)\right) \tag{S9}$$

where $L_{ij}$ is the ratio of the posterior probabilities for location $i$ and location $j$. From equation (S6) we have

$$L_{ij} = \frac{prior(i)\exp\left[\sum_{t=1}^{T+1} {d'_{ik(t)}}^2 W_{ik(t)}\right]}{prior(j)\exp\left[\sum_{t=1}^{T+1} {d'_{jk(t)}}^2 W_{jk(t)}\right]}$$

which can be partitioned into the currently known posterior probabilities $p_i(T)$ and $p_j(T)$, and the currently unknown template responses that will occur after the next fixation. Thus,

$$L_{ij} = \frac{\exp\left[d'^{\,2}_{ik(T+1)} Z_i\right] p_i(T)}{\exp\left[d'^{\,2}_{jk(T+1)} Z_j\right] p_j(T)}$$

Note that the $L_{ij}$ are statistical dependent, but become statistically independent if we condition on the value of $Z_i$. Thus,

$$p\left(C \middle| i, k(T+1)\right) = \int_{-\infty}^{\infty} p(z_i) \prod_{i \neq j} p\left(\frac{\exp\left[d'^{\,2}_{ik(T+1)} z_i\right] p_i(T)}{\exp\left[d'^{\,2}_{jk(T+1)} Z_j\right] p_j(T)} \geq 1\right) dz_i$$

or,

$$p\left(C \middle| i, k(T+1)\right) = \int_{-\infty}^{\infty} p(z_i) \prod_{i \neq j} p\left(Z_j < \frac{\ln\left(\frac{p_j(T)}{p_i(T)}\right) + d'^{\,2}_{ik(T+1)} z_i}{d'^{\,2}_{jk(T+1)}}\right) dz_i$$

Expressing this equation in terms of the standard normal distribution we have

$$p\left(C \middle| i, k(T+1)\right) = \int_{-\infty}^{\infty} d'_{ik(T+1)} \phi\left(d'_{ik(T+1)}(z_i - 0.5)\right) \prod_{i \neq j} \Phi\left(\frac{\ln\left(\frac{p_j(T)}{p_i(T)}\right) \frac{1}{d'^{\,2}_{jk(T+1)}} + \frac{d'^{\,2}_{ik(T+1)}}{d'^{\,2}_{jk(T+1)}} z_i + 0.5}{\frac{1}{d'_{jk(T+1)}}}\right) dz_i$$

where $\phi(x)$ is the standard normal density function and $\Phi(x)$ is the standard normal integral function: $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$, $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) dy$. Finally, making the change of variable $w = d'_{ik(T+1)}(z_i - 0.5)$, we have

$$p\left(C\middle|i,k\left(T+1\right)\right)=\int_{-\infty}^{\infty}\phi\left(w\right)\prod_{j\neq i}\Phi\left(\frac{-2\ln\frac{p_j\left(T\right)}{p_i\left(T\right)}+{d'_{jk(T+1)}}^2+2d'_{ik(T+1)}w+{d'_{ik(T+1)}}^2}{2d'_{jk(T+1)}}\right)dw \tag{S10}$$

Although this formula contains an integral it can be evaluated rapidly using numerical integration, because the standard normal density function approaches zero rapidly away from the origin. This formula reduces to the well-known formula for accuracy in the n-alternative forced choice task[2], in the special case that all the priors are equal and all the values of $d'$ are equal. In sum, equations (S6), (S8) and (S10), and the measured visibility map (equation S3) can be used to simulate optimal visual search in the dynamic noise case.

**Ideal searcher for static external noise and dynamic internal noise**

The ideal searcher for static external noise is basically the same as for dynamic external noise, except that we must consider the external and internal noise separately. Note first that equation (S3) partitions the noise that limits detection performance into two parts, the static part due to the external noise ($\alpha e_n$) and the dynamic part due to the internal noise ($\beta\left(c,e_n,\varepsilon\right)$). We directly determined the value of $\alpha$ (0.0218) by measuring the template responses to the target and by measuring the variance of the template responses to very large number of samples of the actual background noise used in the experiments. Given the value of $\alpha$, the variance of the dynamic internal noise can be determined from the measured visibility maps (the values of $d'\left(c,e_n,\varepsilon\right)$), using equation (S3).

To simulate the ideal searcher's performance, we generated a static noise sample of standard deviation $\sqrt{\alpha e_n}/c$ for each potential target location, before the first fixation. Then, on each fixation, for each potential target location, we generated a dynamic noise sample of standard deviation $\sqrt{\beta\left(c,e_n,\varepsilon\right)}/c$, which we added to the static noise sample at that location. Otherwise, the simulation was run in the same fashion as for dynamic case (see the methods section in the report). We now state (without proof) the formulas for the ideal searcher when the external noise is static.

***Updating posterior probabilities***. The optimal integration of the template responses across fixations is given by the following three equations:

$$p_i\left(T\right)=\frac{prior\left(i\right)\exp\left[\sum_{t=1}^{T}g_{ik(t)}\left(T\right)W_{ik(t)}\right]}{\sum_{j=1}^{n}prior\left(j\right)\exp\left[\sum_{t=1}^{T}g_{jk(t)}\left(T\right)W_{jk(t)}\right]} \tag{S11}$$

$$g_{ik(t)}(T) = \frac{c^2}{e_n \alpha \beta_{ik(t)} v_i(T) + \beta_{ik(t)}} \tag{S12}$$

$$v_i(T) = \sum_{r=1}^{T} \frac{1}{\beta_{ik(r)}} \tag{S13}$$

These equations reflect the fact that the ideal searcher averages out the internal noise over time (because it is dynamic), and hence obtains an ever-improving estimate of the static noise and static noise plus target.

***Selecting next fixation location***. Equation (S8) still applies in the static noise case. However, the formulas for $p\left(C \middle| i, k(T+1)\right)$ are more complicated.

$$p\left(C \middle| i, k(T+1)\right) = \int_{-\infty}^{\infty} \phi(w) \prod_{j \neq i} \Phi\left( \frac{g_{ik(T+1)}(T+1)\left(\sigma_{ik(T+1)} w + \mu_{ik(T+1)}\right) + h_{ij}(T+1)}{g_{jk(T+1)}(T+1)\sigma_{jk(T+1)}} - \frac{\mu_{jk(T+1)}}{\sigma_{jk(T+1)}} \right) dw \tag{S14}$$

$$\mu_{ik(T+1)} = 1 + \frac{c e_n \alpha \sum_{t=1}^{T} \frac{W_{ik(t)} - 0.5}{\beta_{ik(t)}}}{1 + e_n \alpha v_i(T+1)} \tag{S15}$$

$$\mu_{jk(T+1)} = \frac{c e_n \alpha \sum_{t=1}^{T} \frac{W_{jk(t)} + 0.5}{\beta_{jk(t)}}}{1 + e_n \alpha v_j(T+1)}, \; j \neq i \tag{S16}$$

$$\sigma_{jk(T+1)} = \sqrt{\beta_{jk(T+1)} \frac{1 + e_n \alpha v_j(T+1)}{1 + e_n \alpha v_j(T)}}, \quad \forall j \tag{S17}$$

$$h_{ij}(T+1) = \ln\left(\frac{prior(i)}{prior(j)}\right) + \sum_{t=1}^{T}\left(g_{ik(t)}(T+1) W_{ik(t)} - g_{jk(t)}(T+1) W_{jk(t)}\right) \tag{S18}$$

In sum, equations (S11) - (S18), and the measured visibility map (equation S3) can be used to simulate optimal visual search in the static noise case. Although these formulas are more complicated than for the dynamic noise case, they compute at almost the same speed.

***Looking into the future.*** The sub-optimal searcher that always fixates the location with maximum posterior probability (the MAP searcher) does not consider potential information that will be gained after making eye movements, and hence does not look into the future. The ideal searcher described here considers the probability of localizing the target after the next eye movement, and hence it looks one fixation into the future. Because the ideal searcher looks only one fixation into the future it is a so-called 'greedy' algorithm. We have derived formulas for ideal searchers that look more fixations into the future. These algorithms are not practical because of the combinatorial explosion of fixation locations to consider; however, we have compared (for a single search condition) the present ideal searcher with one that is able to look two fixations into the future. This searcher performs slightly better (approximately a quarter of a fixation faster) than the searcher that looks ahead one fixation, which in turn performs better (approximately one fixation faster) than the MAP searcher that looks ahead no fixations. These results demonstrate rapidly diminishing returns for looking further into the future, and thus the present ideal searcher appears to be very close to the true global optimum.


1. Peterson, W. W., Birdsall, T. G. & Fox, W. C. The theory of signal detectability. *Transactions of the Institute of Radio Engineers, Professional Group on Information Theory* **4**, 171-212 (1954).
2. Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).