We will suppose that evidence is in the form of a Gaussian likelihood function for two cues x and y which may have come from a common source. If x and y do not come from a common source then they are uncorrelated and p(x,y) = p(x)\*p(y). If they do come from a common source then p(x,y) is highly correlated. In the simplest case p(x,y) = delta(x-y)\*p(x), but here we will assume that p(x,y) is just a highly correlated Gaussian with correlation coeff 1-epsilon where epsilon is small. Thus the generative model is given by



where



So that



A useful quantity is the difference between these two precision matrices, which is approximately given by,



Now, having observed rx and ry we wish to infer p(x,y) having marginalized out c. This is complicated by the fact that the posterior p(x,y,c|rx,ry) is a mixture distribution. Specifically



Where **x**=[x y]^T and



The VBEM approximation is q(x,y|mu,Sigma)\*q(c|pi) = Normal(mu,Sigma)\*Binomal(pi) where



Note that so that if the predicted means are far away then the log likelihood that they came from the same source is very negative.

Inference is performed by iteratively implementing the equations above (1.6) until a fixed point is reached. Alternatively we could build a continuous time dynamical system with the same fixed point and iterate that. For a linear PPC, the chief difficulty in this case is that we have to compute the actual covariance matrix and mean while a linear PPC usually represents only the inverse of the covariance matrix and the inverse covariance matrix times the mean.

One way around this is to only track the Diagonal of the covarnaice matrix Sigma. This is equlvalent to assuming that the variational posterior fully factorized, i.e. q(x,y,c)=q(x)\*q(y)\*q(c). In this case Sigma and Sigma^{-1} in 1.6 are constrained to be diagonal. This makes Sigma easy to compute (and mu as well) and demonstrates that the non-linearity needed to estimate log(pi/(1-pi)) is a rational polynomial of rank 2 over 2 (because of the mu\*mu^T terms)

Regardless, I like to break down this set of 3 equations (1.6) into 5 coupled equations in which each of the terms on the right hand side of the first line is computed independently. Specifically we define matrices A and B so that



These definition allow us to construct a dynamical system which has the same fixed points as the algorithm or variational inference and also has the property that it is a it is at most quadratic in the natural parameters of the posterior distribution (log(pi/(1-pi)), Sigma^{-1}, and Sigma^{-1}mu) with the exception of log(pi/(1-pi)) which must pass through a logistic nonlinearity in the equation for Sigma^{-1}



Anyway, thats the kind of approach i have been taking recently.

What can we say about g1’,g2’=f(g1,g2). Where g1’ and g2’ are the amplitudes of the spatial localization variables????? Also for Pi.

Fix gain of one and change the other.

For both variational approx and exact. Assume Unimodal Tuning. Plot the plethoral of ‘single cell responces’

Spatial Principle: the less effective stimulus from one modality suppresses a highly effective stimulus from another modality. Figures 2,3,4 in Greg’s 2011 paper. Nature Neuro.

Inverse effectiveness.

Multisensory neurons.

Show effects of gain for both C and S.

For variational approximation, how bad it the approximation in terms of behaviour (percent correct) when you use a fixed (best) decision criterion across multiple gains.

CALCULATING PTRUE



Where C=1 or C=0. We compute p(c|r\_x,r\_y) in the usually way. The full joint distribution on x and y with c marginalized out can be obtained by simply summing,



This results in mixture of Gaussians posterior distribution for x and y. While I have my preferred way to represent a mixture of Gaussians, it seems more appropriate to just use the ‘magic property’ that the PPC encodes the inverse variance of the posterior in its amplitude. We therefore compute the mean and variance of the posterior of x and y. This is actually pretty easy.



The same holds for the raw second moment (but not the central). This implies that,



The OR Neuron

A population of OR neurons encodes the presence of an object in the receptive fields of either modality. There are two cases to consider, c=0 and c=1. The c=1 case is easy. There is only one object. In the c=0 case, its somewhat more complicated. The idea is that in the c=0 case, we consider p(x,y|rx,ry,c=0) where x is the position of the object in modality 1 and y is the position of an object in modality 2. If we define z as the position of a particle in either modality, then the probability that there is an object in either modality is given by



This means that



The variance can then be computed in a straight forward manner.