Bayesian Model for Multisensory Integration and Segregation

Xiangyu Ma[1], **He Wang**[1], **Min Yan**[1], **Wen-Hao Zhang**[2], **and K. Y. Michael Wong**[1].
[1]Hong Kong University of Science and Technology, [2]University of Pittsburgh.

Abstract

Multisensory integration and segregation are important for the survival of higher animals. Experimental data indicate that the brain processes information in a Bayesian way. We consider a recently proposed model that is able to perform both multisensory integration and segregation concurrently using congruent and opposite groups of neurons in each sensory module. We show that the model is able to yield estimates with excellent agreement with Bayesian inference in the weak stimulus limit, and fairly good agreement in stronger fields. When the prior consists of a correlated component and an independent component, we show that Bayesian inference can be achieved by incorporating an additional layer of neuron groups.

Keywords: Bayesian inference; multisensory integration and segregation; continuous attractor neural network

Introduction

Our brains process information from different sensory modalities. If two cues are received from the same source, the nervous system will integrate those sensory signals. Otherwise, they should be segregated. For instance, when you are walking along a street, the signals originating in the visual and auditory stimuli will integrate together or segregate apart to improve the accuracy of perceptions of the moving objects. Multisensory integration and segregation are important in neuroscience since the give the ability for mammals to build coherent perceptions of the external world. Previously, the experimental data suggested that in the dorsal medial superior temporal (MSTd) area and the ventral intraparietal (VIP) area, there exist two types of neurons, congruent and opposite cells. Based on previous studies, the congruent cells are responsible for multisensory integration. In our recent studies, (Zhang et al., 2019) proposed that the opposite cells are relevant to the multisensory segregation (Zhang et al., 2019). In the proposed model, the neural circuit consists of two modules, each contains two groups of excitatory neurons, congruent and opposite neurons. It was shown that the proposed network yields the Bayesian posterior estimate in a broad range of parameters, but there are also parameter ranges that the inference can only be approximately Bayesian. Hence, in this paper we will approach the dynamics analytically and analyze the conditions for Bayesian inference. Furthermore, the Bayes-optimality in (Zhang et al., 2019) is based on a prior distribution of stimuli that is fully correlated. In practice, there are many other scenarios described by priors with more than one component. For example, studies in causal inference consider prior distributions with a correlated and an independent component (Körding et al., 2007; Sato, Toyoizumi, & Aihara, 2007; Shams & Beierholm, 2010). In the second half of the paper, we propose a neural circuit with additional modules to cover these cases.

A Model of Multisensory Integration and Segregation

We consider a neural network model (Zhang et al., 2019) receiving external inputs of modality 1 and 2 with $I_m^{ext}(y,t)$, $m=1,2$, where y is an angular variable in the range $(\pi,\pi]$ and t is time. The two inputs are fed into two separate modules. Each module has two groups of excitatory and opposite neurons. The recurrent connections within each group are excitatory and dependent on the positions of the neurons through a bump-shaped function of the position separation, and there are global inhibitory connections among both groups, thus forming a continuous attractor neural network. The congruent groups of each module are connected in a congruent manner, that is, neurons receiving inputs at position x of each module are reciprocally connected to each other. Likewise, the opposite groups of each module are connected in an opposite manner, that is, neurons receiving inputs at position x of one module are reciprocally connected to those at position $x+\pi$. Let $\psi_m(x,t)$ and $\bar{\psi}_m(x,t)$ be the synaptic input at position x and time t for the congruent and opposite groups respectively in module m, and denote as $\bar{m}$ the other module of module m. Then the neuronal dynamics of the congruent group is given by

$$\tau\frac{\partial\psi_m(y,t)}{\partial t}=-\psi_m(y,t)+\frac{J_{rc}}{D_m}\sum_{y'=-\pi}^{\pi}V(y-y',a_0)\psi_m^2(y',t)$$
$$+\frac{J_{rp}}{D_{\bar{m}}}\sum_{y'=-\pi}^{\pi}V(y-y',a_0)\psi_{\bar{m}}^2(y',t)+I_m^{ext}(y,t) \quad (1)$$

where J_{rc} and J_{rp} represent the strengths of the recurrent and reciprocal couplings respectively, $D_m\equiv 1+\omega\left[\sum_y\psi_m^2(y,t)+J_{int}\sum_y\bar{\psi}_m^2(y,t)\right]$ is the global inhibition acting on the congruent group in module m. $V(y-y',a_0)$ is the von Mises function given by

$$V(y-y',a_0)\equiv\frac{\exp[a_0\cos(y-y')]}{2\pi I_0(a_0)} \quad (2)$$

where $I_0(a_0)$ is introduced to normalize the integration of the von Mises function; $I_0(a_0)$ is known as the modified Bessel function of order 0.

In Eq. (1), we assign the external inputs to be the sum of a bump-shaped input of amplitude I_m multiplied by the von Mises function centered at position x_m and a constant input

of amplitude I_b, plus their noisy components characterized by the Fano factor F_0,

$$I_m^{ext}(y,t) = I_m V(y-x_m, \frac{a_0}{2}) + I_b + \sqrt{F_0 I_1 V(y-x_m, a_0/2)}\xi_m(y,t) + \sqrt{F_0 I_b}\varepsilon_m(y,t) \quad (3)$$

where $\xi_m(y,t)$ and $\varepsilon_m(y,t)$ are Gaussian white noise of zero mean and variance satisfying $\langle \xi_m(y,t), \xi_{m'}(y',t')\rangle = \delta_{mm'}\delta(y-y')\delta(t-t')$ and $\langle \varepsilon_m(y,t), \varepsilon_{m'}(y',t')\rangle = \delta_{mm'}\delta(y-y')\delta(t-t')$.

On the other hand, the neuronal dynamics of the opposite group is given by

$$\tau\frac{\partial \bar{\psi}_m(y,t)}{\partial t} = -\bar{\psi}_m(y,t) + \frac{J_{rc}}{D_m}\sum_{y'=-\pi}^{\pi} V(y-y',a_0)\bar{\psi}_m^2(y',t) + \frac{J_{rp}}{D_{\bar{m}}}\sum_{y'=-\pi}^{\pi} V(y-y'+\pi,a_0)\bar{\psi}_{\bar{m}}^2(y',t) + I_m^{ext}(y,t). \quad (4)$$

These equations can be solved by numerical simulations to obtain the means and variances of the firing rates. In the framework of probabilistic population coding ("Bayesian inference with probabilistic population codes", 2006), the posterior estimates of the external inputs can be derived from these quantities. The analytical technique that we adopt is the projection method (Fung, Wong, & Wu, 2010). We first attempted first order perturbation using the von Mises function and its derivative as the basis. They represent distortions of the height and position of the bump-shaped firing rate distributions respectively, but the calculated means and variances of the distributions deviated from the numerical results. A careful inspection of the firing rate distributions showed that their profiles were not calculated accurately. Hence, higher order perturbations describing the distortions in the height, position, width and skewness have to be introduced. We approximate the solution to the dynamical equations to be

$$\psi_m = u_{m0} + u_{m1}\cos(y_1 - s_1) + u_{m2}\cos 2(y_1 - s_1) + u_{m3}\sin 2(y_1 - s_1)\ m = 1,2. \quad (5)$$

The background, height, position, width and skewness are largely determined by the coefficients u_{m0}, u_{m1}, s_m, u_{m2} and u_{m3} respectively. Multiplying both sides of Eqs. (1) and (4) by 1, $\cos(y-s_m)$, $\cos 2(y-s_m)$ and $\sin 2(y-s_m)$ in turn and integrating over y, we obtain the steady state equations for this set of coefficients after averaging over noise. By considering the linear perturbation around the steady state, the variance $\hat{\sigma}_m^2$ of the peak positions $\hat{s}_m$ can be found.

Bayesian Inference

Consider the task of inferring the stimuli s_m $(m = 1,\ 2)$ from the received cues x_m $(m = 1,\ 2)$. It has been shown that for uniform distributions of s_m and cues x_m, the condition for the network inference to produce Bayesian inference is that

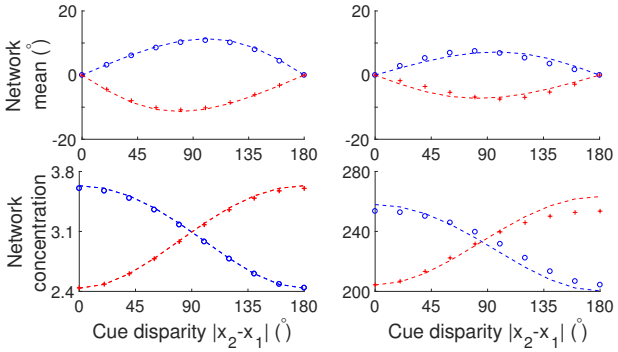

Figure 1: Weak input (left column) and strong input (right column). Symbols: network results; dash lines: Bayesian prediction. The blue and red colors represent congruent and opposite groups in module 1 respectively.

the marginal posterior distribution of s_1 given the direct cue x_1 and indirect cue x_2 by (Zhang et al., 2019)

$$p(s_1|x_1,x_2) \propto p(s_1|x_1)p(s_1|x_2) \quad (6)$$

where the marginal posterior distribution $p(s_1|x_2)$ of s_1 given the indirect cue x_2 depends on the prior distribution $p(s_1,s_2)$ via

$$p(s_1|x_2) \propto \int ds_2 p(x_2|s_2)p(s_1,s_2). \quad (7)$$

The expressions of the marginal posterior distribution of s_2 are similar; hereafter we will focus on the result for s_1. To explain the role of congruent and opposite neurons, it is convenient to consider likelihood distributions in the form of the von Mises distribution with concentration κ_m,

$$p(x_m|s_m) = V(x_m - s_m, \kappa_m). \quad (8)$$

Hence, to verify whether the congruent groups of the proposed network is able to make Bayesian predictions, one may use it to estimate the posterior distributions of s_1 when it receives cue 1 only, cue 2 only, and cues 1 and 2 combined, and test whether the result for the combined cues agrees with those predicted from the single cues according to Eq. (6).

We first consider a prior that the stimuli s_1 and s_2 are correlated. In particular, we consider the prior

$$p(s_1,s_2) = V(s_1 - s_2, \kappa_s). \quad (9)$$

When this von Mises prior is substituted into Eq. (7), the integration yields a new von Mises function $V(s_1 - x_2, \kappa_{2s})$, where ,$\kappa_{2s} \equiv A^{-1}[A(\kappa_2)A(\kappa_s)]$ with the function A related to the modified Bessel functions via $A(\kappa) \equiv I_1(\kappa)/I_0(\kappa)$ (V. Mardia & E Jupp, 2000). In turn, when Eq. (7) is substituted into Eq. (6), the product of two von Mises distributions yields another von Mises distribution whose mean and concentration are given by a vector sum rule. Thus, the condition (6) for Bayesian inference reduces to

$$\hat{\kappa}_1 e^{j\hat{s}_1}|_{I_1,I_2} = \hat{\kappa}_1 e^{j\hat{s}_1}|_{I_1} + \hat{\kappa}_{2s} e^{j\hat{s}_2}|_{I_2}, \quad (10)$$

where j is the imaginary number $\sqrt{-1}$. The subscripts represent the non-vanishing stimuli applied to the network, and the hat accents represent the network estimates.

To segregate the information from the two cues, we consider the disparity information of stimulus 1 defined to be

$$p_d(s_1|x_1,x_2) \propto \frac{p(s_1|x_1)}{p(s_1|x_2)}. \tag{11}$$

Noting that the cosine function satisfies $\cos(y-y'-\pi) = \cos(y-y)$, we obtain

$$p_d(s_1|x_1,x_2) \propto V(s_1-x_1,\kappa_1)V(s_1-x_2-\pi). \tag{12}$$

Hence, the mean $\Delta\hat{s}_1$ and concentration $\Delta\hat{\kappa}_1$ of the disparity information, to be estimated by the p

$$\hat{\kappa}_1 e^{j\hat{s}_1}|_{I_1,I_2} = \hat{\kappa}_1 e^{j\hat{s}_1}|_{I_1} - \hat{\kappa}_{2s} e^{j\hat{s}_2}|_{I_2}. \tag{13}$$

Equations (10) and (13) are used to generate Bayesian predictions based on the single-input estimates and compare with the combined estimates generated from network simulations (the estimated concentration will be compared with the inverse of the variance). In our simulations, each network consist of $N = 180$ congruent and opposite neurons each. The synaptic range $a_0 = 3$ and the time step is 0.01τ using Eulers method with τ rescaled to 1. The strength of the background input is $I_b = 1$. The strength of the divisive normalization is $\omega = 3\times10^4$ and $J_{int} = 1$. The Fano factor is set to 0.5. We fix $x_1 = 0$, $J_{rc} = 0.3J_c$ and $J_{rp} = 0.5J_{rc}$, where J_c is the minimal recurrent strength for the formation of a persistent bump, given by $J_c = \sqrt{\frac{8\pi I_0(a_0/2)^2\omega(1+J_{int})}{I_0(a_0)\rho}}$ (Fung et al., 2010). U_0 is the rescaled input strength $U_0 = \frac{J_c e^{a_0/2}}{2\pi\omega(1+J_{int})I_0(a_0/2)}$. As shown in Fig. 1, we find that the network can implement Bayesian inference in the weak field limit. When the external inputs are strong, the network prediction starts to deviate from the Bayesian inference, but the estimates remain reasonably close.

Priors with an Independent Component

So far we have considered the prior in Eq. (8) in which the two stimuli are correlated. However, there are many other scenarios described by priors with an additional independent component. Those priors are often used in causal inference tasks, in which the subject is required to determine whether the two stimuli originate from the same cause or they are independent (Körding et al., 2007; Sato et al., 2007; Shams & Beierholm, 2010). Hence, we consider the following prior,

$$p(s_1,s_2) = \frac{p_0}{2\pi}V(s_1-s_2,\kappa_s) + \frac{1-p_0}{(2\pi)^2}. \tag{14}$$

Using Eq. (7), the marginal posterior distribution $p(s_1|x_2$ becomes

$$\begin{aligned} p(s_1|x_1,x_2) =& p_0 C V(s_1-x_1,\kappa_1)V(s_1-x_2,\kappa_{2s}) \\ &+ (1-p_0)V(s_1-x_1,\kappa_1), \end{aligned} \tag{15}$$

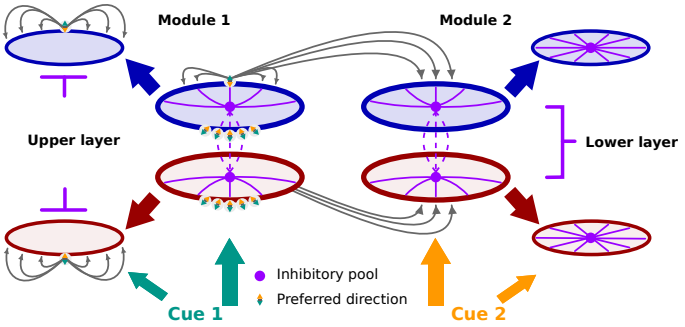

Figure 2: Illustration of decentralized model after modification. New modules in the second layer. No connection between two groups of neurons in each module. The inputs and feedforward inputs from the first layer are rescaled by $1-p_0$ and p_0 respectively. Each group of neurons has their own inhibition pool and there's no reciprocal coupling between two modules.

where C is the normalization constant. The first term corresponds to the integrated information output from the congruent group of neurons described in the previous section, whereas the second term corresponds to the direct input from the stimulus. Hence, we propose the network architecture in Fig. 2, in which the congruent and opposite groups of the modules form the first layer, whose outputs are fed into the congruent groups of the second layer, which also receive the direct stimulus input. The dynamics of this group of neurons is given by

$$\begin{aligned} \tau\frac{\partial\psi^u_m(y,t)}{\partial t} =& -\psi^f_m(y,t) + \frac{J_{rc}}{D_m}\sum_{y'=-\pi}^{\pi} V(y-y',a_0)[\psi^u_m(y',t)]^2 \\ &+ p_0\frac{c_k}{D_m}\sum_{y'=-\pi}^{\pi}\cos(y-y')[\psi_m(y',t)]^2 + I^{ext}_m(y,t). \end{aligned} \tag{16}$$

Note that the neuron groups in the second layer do not have reciprocal connections from the other module. Hence, their output will be the weighted sum of the two types of input. Thus, for the case of combined cues, the output of the congruent group in module 1 will become $p_0\kappa e^{js_1} + (1-p_0)\kappa'_1 e^{jx_1}$. The first component is a complex number since it contains position information. Meanwhile, κ'_1 does not change when this network only receives direct stimulus 1. The output will then be $p_0\kappa_1 e^{jx_1} + (1-p_0)\kappa'_1 e^{jx_1}$. When the network only receives stimulus 2, the final output of the congruent group will be $p_0\kappa_{12}e^{jx_2}$. So in summary, the network has a Bayesian behaviour in all cases. Figure 3 shows the vector diagram showing that information integration can be achieved.

In Fig. 4 we compare the network behaviour in weak and strong inputs, corresponding to $I = 0.01U_0$ and $I = 0.07U_0$ respectively. The outputs from the second layer behave in a Bayesian way in the weak stimulus limit. Although the prior is the sum of two von Mises functions, the output is not double-peaked since the position disparity between inputs and feedforward inputs is small. Next, we consider information segre-

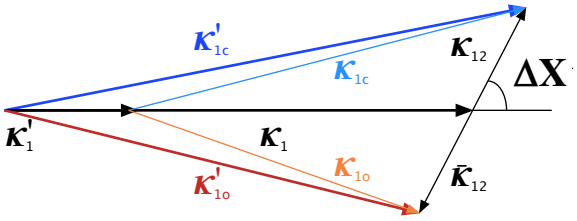

Figure 3: Geometric interpretation of vector space. Colored vectors represent the sum of other vectors (black). Outputs from first layer: κ_{1c} (congruent) and κ_{1o} (opposite), second layer: κ'_{1c} (congruent) and κ'_{1o} (opposite). Note that the vectors (black) have been rescaled by p_0 and $1-p_0$.

gation. Using Eq. (11), the disparity information is given by

$$p_d(s_1|x_1,x_2) = \frac{V(s_1-x_1,\kappa_1}{p_0 CV(s_1-x_2,\kappa_{2s})+(1-p_0)}. \quad (17)$$

Although p_d is hard to obtain from this neuronal network, however, this expression is equivalent to that of p_d^{-1} below, written as

$$\begin{aligned} p_d(s_1|x_1,x_2)^{-1} =& p_0 CV(s_1-x_2,\kappa_{2s})V(s_1-x_1+\pi,\kappa_1) \\ &+(1-p_0)V(s_1-x_1+\pi,\kappa_1), \end{aligned} \quad (18)$$

where C is the normalization constant. Note that the stimulus position is shifted by π for the opposite group. Hence, we see that the opposite group in the second layer has the same structure as that of the congruent group, except that the positions of the outputs from the second layer is shifted by π.

Conclusion and Discussion

In this work, we have developed a novel method for revealing the dynamics inside the neural circuit. We discuss the conditions for decentralized model to achieve Bayesian inference, which is helpful in biological implementation. However, the functions of opposite groups are not fully understood, and the details of the structure on causal inference are missing in this model. That could be our future work.

References

Bayesian inference with probabilistic population codes. (2006). *Nature Neuroscience*, *9*(11), 1432–1438. doi: 10.1038/nn1790

Fung, C. C. A., Wong, K. Y. M., & Wu, S. (2010). A moving bump in a continuous manifold: A comprehensive study of the tracking dynamics of continuous attractor neural networks. *Neural Computation*, *22*(3), 752-792. (PMID: 19922292) doi: 10.1162/neco.2009.07-08-824

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLOS ONE*, *2*(9), e943.

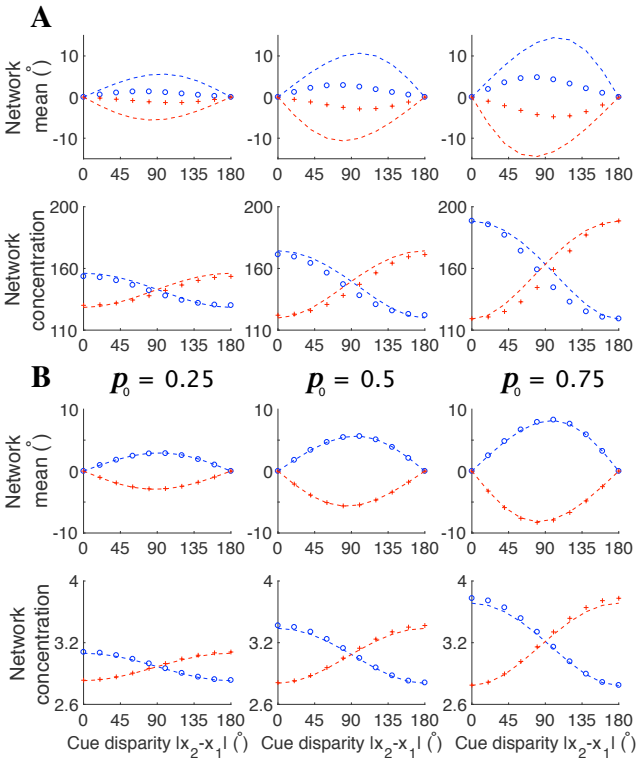

Figure 4: Outputs from second layer with varying probability p_0. Illustration of the population response of congruent and opposite groups in module 1 applying weak inputs (A) and strong inputs (B). Symbols: network results; dash lines: Bayesian prediction. The blue and red colors represent congruent and opposite groups in module 1 respectively.

Sato, Y., Toyoizumi, T., & Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: Identification of common sources of audiovisual stimuli. *Neural Comput*, *19*(12), 3335–3355.

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends Cogn Sci*, *14*(9), 425–432.

V. Mardia, K., & E Jupp, P. (2000). *Directional statistics*. doi: 10.1002/9780470316979.ch11

Zhang, W.-H., Wang, H., Chen, A., Gu, Y., Lee, T. S., Wong, K. Y. M., & Wu, S. (2019, may). Complementary congruent and opposite neurons achieve concurrent multisensory integration and segregation. , *8*, e43753.