

A Decentralized Architecture for Multisensory Neural Information Integration

Wen-hao Zhang^{†‡}, K.Y. Michael Wong[‡], Malte J. Rasch[†] and Si Wu[†]

[†]State Key Laboratory of Cognitive Neuroscience and Learning,
 IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China

[‡]Department of Physics, Hong Kong University of Science & Technology,
 Clear Water Bay, Kowloon, Hong Kong, China

Email: {wenhaoz, phkywong}@ust.hk, {malte.rasch, wusi}@bnu.edu.cn

Abstract—Multisensory integration is important in our brain. However, understanding how the brain integrates multiple sensory cues in its neural circuitry remains a challenge. In this study, using biologically realistic neural network models, we propose a novel mechanism of how multisensory information might be integrated in a distributed fashion across interconnected brain areas without the need for a central integration unit. We show that this decentralized system can integrate information optimally in a biologically realistic setting, and is in good agreement with anatomical constraints and the experimental observations.

1. Introduction

In our daily life, we sense the world through multiple sensory systems. For instance, while walking, we perceive heading direction through either the visual cue (optic flow), or the vestibular cue generated by body movement, or both of them [1–3]. Because of ubiquitous noise in the nature, our perception of the input information is often uncertain. In order to achieve an accurate representation of the input, it is important for the brain to integrate cues which all convey information of the same feature. A number of elegant psychophysical experiments have demonstrated that humans and animals integrate multisensory information in an optimal Bayesian way. For instances, the integration of visual and auditory cues to infer object location [4], combining visual and proprioceptive cues to hand position [5], visual and haptic integration of object height [6], and texture and motion integration [7].

Mathematically, Bayesian inference provides an optimal way to estimate the stimulus value based on multiple uncertain information resources. Consider that stimulus s generates cue 1 c_1 and cue 2 c_2 (internal representation in the brain). Under the assumption that the noise processes of two cues given stimulus are conditionally independent with each other, the posterior distribution $p(s|c_1, c_2)$, satisfies [8]

$$p(s|c_1, c_2) \propto p(c_1|s)p(c_2|s)p(s), \quad (1)$$

where $p(s)$ is the prior distribution of stimulus, which is modelled as a uniform distribution in this study. $p(c_l|s)$ ($l = 1$ or 2) is the likelihood function, which is modelled as a Gaussian distribution with mean μ_l and variance σ_l^2 .

Because of flat prior, the posterior of stimulus when giving two cues can be factorized as the products of the posterior given each single cue

$$p(s|c_1, c_2) \propto p(s|c_1)p(s|c_2). \quad (2)$$

And the posterior under single cue, $p(s|c_l)$ satisfies a same Gaussian distribution as the likelihood function $p(c_l|s)$. Eq. (2) indicates that the mean and variance of posterior given two cues can be predicted from the posteriors under either single cues

$$V(s|c_1, c_2)^{-1} = V(s|c_1)^{-1} + V(s|c_2)^{-1}, \quad (3)$$

$$\frac{\langle s|c_1, c_2 \rangle}{V(s|c_1, c_2)} = \frac{\langle s|c_1 \rangle}{V(s|c_1)} + \frac{\langle s|c_2 \rangle}{V(s|c_2)}, \quad (4)$$

where $\langle s|c_l \rangle = \mu_l$, $V(s|c_l) = \sigma_l^2$. Eqs. (3 and 4) together are used as a criteria to check whether Bayesian is achieved in experiment (e.g., [6]), which will be also used in this study.

Nevertheless, the detailed neural mechanism underlying Bayesian integration remains largely unclear. Previous modelling studies all used a single dedicated network to receive feedforward inputs and then integrate information [9, 10]. However, in the visual and vestibular integration for heading direction, recent electrophysiological experiments found that instead of single multisensory brain area, but more than one brain area, including dorsal medial superior temporal (MSTd) area and ventral intraparietal (VIP) area, can optimally integrate visual and vestibular inputs to infer the heading direction [2, 3]. Moreover, in anatomy, MSTd and VIP are reciprocally connected [11, 12].

Based on these experimental evidence, we propose that the MSTd and VIP may constitute a decentralized system in the brain to integrate visual and vestibular information. To verify this possibility, we construct a model of decentralized system which is composed of two reciprocally connected networks. Because of the wide tuning function of MSTd and VIP neurons with respect to heading direction, each network is chosen as a continuous attractor neural network (CANN). And each network receives a cue as feedforward input, which conveys the information of corresponding stimulus. By using theoretical analyses and numerical simulations, we demonstrated that each network in the decentralized system can integrate information optimally in

the sense of Bayesian inference. Our result suggests that the brain may implement Bayesian information integration distributively through reciprocal connections between cortical regions.

2. The decentralized network model

We consider a decentralized system consist of two reciprocally connected networks, with each receives a feedforward inputs representing the information of a stimulus (see Fig. 1A). Each network is modelled as a CANN, due to the wide tuning functions of MSTd and VIP neurons. CANNs, also known as neural field model, have been successfully applied to describe the encoding of head-direction in neural systems [13]. And it has been demonstrated that the CANN can infer the underlying stimulus value from noisy inputs through a template-matching operation, and behave in a manner near a maximal likelihood estimator [14, 15].

The dynamics of a neuron in the decentralized system is determined by the recurrent inputs from other neurons in the same network, the reciprocal inputs from neurons in the other network, the feedforward input $I_l(\theta, t)$, and its own relaxation. Mathematically, it can be written as [16]

$$\tau \frac{\partial u_l(\theta, t)}{\partial t} = -u_l(\theta, t) + \sum_{m=1}^2 \rho \int_{-\pi}^{\pi} W_{lm}(\theta, \theta') r_m(\theta', t) d\theta' + I_l(\theta, t), \quad (5)$$

where θ denotes the stimulus value (i.e. the heading direction) encoded by both networks, and the neuronal preferred stimuli are in the range of $-\pi < \theta \leq \pi$ with periodic boundary condition. And $u_l(\theta, t)$, for $l = 1, 2$, represents the synaptic input at time t of the neuron preferring stimulus θ in the l -th network. $r_l(\theta, t)$ is the firing rate of neurons, which is a function of $u_l(\theta, t)$ with a divisive normalization form [14, 17],

$$r_l(\theta, t) = \frac{[U_l(\theta, t)]_+^2}{1 + k\rho \int_{\theta'} [U_l(\theta', t)]_+^2 d\theta'}, \quad (6)$$

where the symbol $[x]_+$ denotes a half-rectifying function, i.e., $[x]_+ = 0$, for $x \leq 0$ and $[x]_+ = x$, for $x > 0$, and k reflects the strength of global inhibition.

$W_{lm}(\theta, \theta')$ denotes the connection from the neurons θ' in the network m to the neurons θ in the network l . Thus, $W_{ll}(\theta, \theta')$ denotes the recurrent connection within the same network, and $W_{lm}(\theta, \theta')$ ($l \neq m$) represents the reciprocal connection between the networks. We assume they are of the Gaussian form, i.e.,

$$W_{lm}(\theta, \theta') = \frac{J_{lm}}{\sqrt{2\pi}a} \exp\left[-\frac{(\theta - \theta')^2}{2a^2}\right], \quad (7)$$

where a determines the neuronal interaction range. We choose $J_{lm} > 0$, for $l, m = 1, 2$, implying excitatory recurrent and reciprocal neuronal interactions. And reciprocal connection strength J_{lm} ($l \neq m$) is always smaller than

recurrent connection strength J_{ll} . The contribution of inhibitory neurons is implicitly included in the divisive normalization (Eq. 6).

When cue l is presented, network l receives a feedforward input $I_l(\theta, t)$, which can be regarded as the inputs from unisensory cortical area,

$$I_l(\theta, t) = \alpha_l \exp\left[-\frac{(\theta - \mu_l)^2}{4a^2}\right] + \gamma_l \xi_l(\theta, t) + \eta_l \epsilon_l(\theta, t), \quad (8)$$

where α_l is the input intensity and μ_l denotes the stimulus value conveyed to the network l by the corresponding sensory cue. This can be understood as I_l drives the network l to be stable at μ_l when no reciprocal interaction and noise exist. The term $\gamma_l \xi_l(\theta, t) + \eta_l \epsilon_l(\theta, t)$ denotes the input noise, with $\xi_l(\theta, t)$ and $\epsilon_l(\theta, t)$ are two independent Gaussian white noises of zero mean and unit variance. The input noise consists of two parts. The first part $\gamma_l \xi_l$ is stimulus-independent, and the second one $\eta_l \epsilon_l$ depends on the signal strength α_l , implying that $\eta_l = 0$ when $\alpha_l = 0$. The noise term causes the uncertainty of the input information, which induces fluctuations of the network state. The exact form of I_l is not critical here, as long as it has an unimodal form. In this study, to simplify the analyses, we take a symmetric parameter setting for two networks, meaning that $J_{11} = J_{22} \equiv J_{rc}$, $J_{21} = J_{12} \equiv J_{rp}$.

2.1. Simply the network dynamics

It has been demonstrated that the dynamics of CANN is dominated by position changes [18, 16]. In this case, the network dynamics can be solved by using the following Gaussian ansatz

$$u_l(\theta, t) \approx U_l \exp\left[-\frac{(\theta - z_l(t))^2}{4a^2}\right], \quad (9)$$

$$r_l(\theta, t) \approx R_l \exp\left[-\frac{(\theta - z_l(t))^2}{2a^2}\right], \quad (10)$$

where U_l and R_l represent the mean values of bump height, which are treated unchanged in the statistically stationary state. $z_l(t)$ denotes the position of bump at time t , which is interpreted as the network estimates for stimulus [14]. The dominant mode corresponding to bump position change is found as [18]

$$\phi(\theta|z) = \frac{\theta - z}{a} \exp\left[-\frac{(\theta - z)^2}{4a^2}\right]. \quad (11)$$

By projecting the full network dynamics (Eq. 5) onto the dominant mode of bump position (Eq. 11), we get the dynamics of the bump position for both networks (projecting a function $f(\theta)$ onto mode $\phi(\theta|z)$ is to compute $\int_{\theta} f(\theta) \phi(\theta|z) d\theta / \int_{\theta} \phi(\theta|z)^2 d\theta$).

$$\frac{dz_1}{dt} = g_{12}(z_2 - z_1) + h_1(\mu_1 - z_1) + \beta_1 \xi_1(t), \quad (12)$$

$$\frac{dz_2}{dt} = g_{12}(z_1 - z_2) + h_1(\mu_2 - z_2) + \beta_2 \xi_2(t), \quad (13)$$

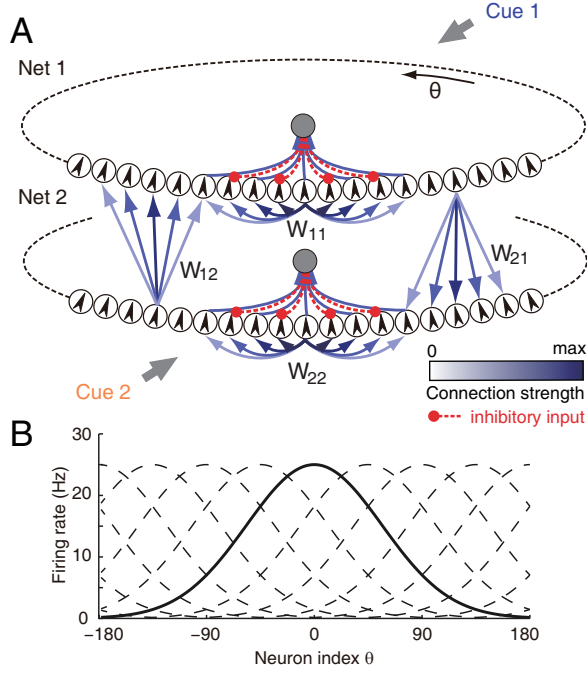


Figure 1: Network structure and stationary state. (A) The two networks are reciprocally connected and each of them forms a CANN. Each circle represents an excitatory neuron with its preferred heading direction indicated by the arrow inside. The red circle represents the inhibitory neuron pool which sums the total activities of excitatory neurons and generates divisive normalization (Eq.6). The solid line with arrow is excitatory connection with the color indicating the strength. The red dashed line with dots represents inhibitory connection. (B) The stationary states of network 1, whose position in the perceptual space is determined by the position of feedforward inputs. Parameters: $N = 100$, $k = 5 \times 10^{-4}$, $a = 40^\circ$, $J_{rc} = 0.5J_c$, $J_{rp} = 0.5J_{rc}$. $J_c = 2\sqrt{2}(2\pi)^{1/4}\sqrt{k\alpha\rho}$ is the minimal recurrent connection strength for holding persistent activity. And $\alpha = 0.6U^0$, where $U^0 = J_c/4ak\sqrt{\pi}$ is the synaptic bump height the network hold when $J_{rc} = J_c$ and without reciprocal connection. $\gamma = 0.5$ and $\eta = 0.3$.

where $g_{lm} = \rho J_{lm} R_m / (\sqrt{2}\tau U_l)$ denotes the effective strength of reciprocal connections, $h_l = \alpha_l / (\tau U_l)$ the effective strength of feedforward inputs, and $\beta_l = 2\sqrt{a(\gamma_l^2 + \eta_l^2)} / ((2\pi)^{1/4}\tau U_l)$ the effective strength of noise.

3. Optimal integration in decentralized network

Let's verify whether the network estimates satisfy the Bayesian predictions. Similar with typical cue integration experiments (e.g., [6]), we first applied two single cues individually, and then applied both of them simultaneously (Fig. 2A). Then we check, for each network, whether the estimates under combined cues can be predicted from the ones under two single cues by using Bayesian inference (E-

qs. 3 and 4).

In theoretical analysis, we assume that the effective strength g_{lm} and h_l (when cue l is presented in Eqs. (12, 13) are approximately unchanged with stimulus conditions and abbreviated to $g_{lm} \equiv g_{rp}$ and $h_l \equiv h$ ($\alpha_l \neq 0$) under symmetric parameter setting, respectively. Moreover, we further assume that the effective noise strength β , a ratio of noise variance over bump height and reflect the signal to noise ratio of the network, is also approximately unchanged across different stimulus conditions. This is supported by the experimental evidence that the Fano factor of neuron changes insignificantly with stimulus conditions [2]. With these approximations, the estimation mean and variance of network 1 in statistically stationary state under three stimulus conditions can be calculated from Eqs. (12 and 13) (the results of network 2 can be obtained by interchanging the indices of 1 and 2)

- Only cue 1 is presented ($h_1 = h; h_2 = 0$)

$$\langle z_1 | c_1 \rangle = \mu_1, \quad V(z_1 | c_1) = \frac{\beta^2}{2} h^{-1}; \quad (14)$$

- Only cue 2 is presented ($h_1 = 0; h_2 = h$)

$$\langle z_1 | c_2 \rangle = \mu_2, \quad V(z_1 | c_2) = \frac{\beta^2}{2} (h^{-1} + g_{rp}^{-1}); \quad (15)$$

- Combined cues ($h_1 = h_2 = h$)

$$\langle z_1 | c_1, c_2 \rangle = \frac{(h^{-1} + g_{rp}^{-1})\mu_1 + h^{-1}\mu_2}{2h^{-1} + g_{rp}^{-1}}, \quad (16)$$

$$V(z_1 | c_2) = \frac{\beta^2 h^{-1} (h^{-1} + g_{rp}^{-1})}{2(2h^{-1} + g_{rp}^{-1})}. \quad (17)$$

It can be checked that the estimates of both network satisfy the Bayesian predictions (Eqs. 3 and 4), meaning that the estimates of each network individually satisfy

$$V(z_l | c_1, c_2)^{-1} = V(z_l | c_1)^{-1} + V(z_l | c_2)^{-1}, \quad (18)$$

$$\frac{\langle z_l | c_1, c_2 \rangle}{V(z_l | c_1, c_2)} = \frac{\langle z_l | c_1 \rangle}{V(z_l | c_1)} + \frac{\langle z_l | c_2 \rangle}{V(z_l | c_2)}, \quad l = 1 \text{ or } 2. \quad (19)$$

While the optimal integration can be achieved in our theoretical analysis of decentralized network, the effective strength g_{rp} , h and β will generally change nonlinearly with network parameters and stimulus conditions, possibly making the network estimations deviate from the Bayesian observer. Hence, we performed numerical simulation to check whether the network can integrate two cues optimally. Fig. 2B plots the population activities of two networks in response to two congruent cues both centered at 0° . When two cues are simultaneously presented, the network responses are increased compared with single cue condition. Fig. 2C is a snapshot of the population activity of network 1, which is a noisy bump. And the bump position

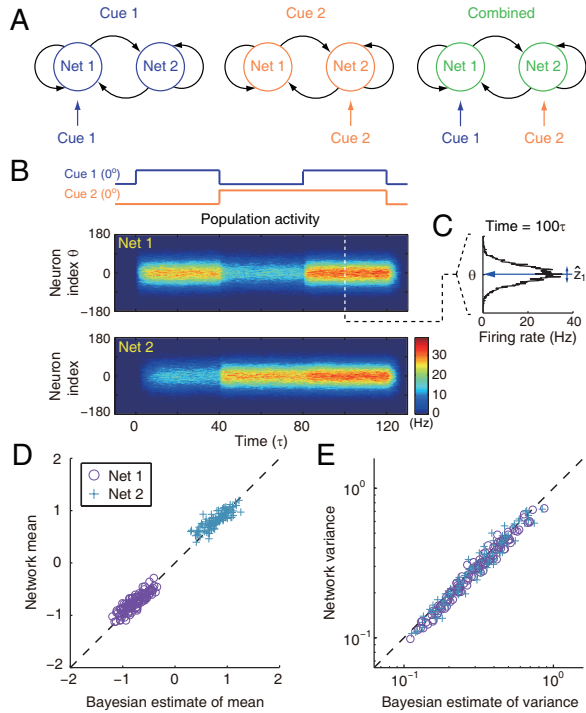


Figure 2: Optimal information integration in decentralized system. (A) Illustration of the three stimulus conditions applied to the system. (B) Population activities of two coupled networks in response to stimulus condition shown in (A). (C) A snapshot of the population activity in network 1. The position of the activity bump is considered as the current estimate of the network (z_1). (D) and (E). Comparisons between network's estimation mean and variance with Bayesian predictions (Eq. 3 and 4) under different parameters. Parameters: $\alpha \in [0.5, 1.2]U^0$, $\gamma = 0.5$ and $\eta \in [0.3, 0.6]$, and $J_{rc} = 0.5J_c$, $J_{tp} \in [0.6, 0.9]J_{rc}$.

z_1 , can be estimated by population vector numerically, represents the network estimates of stimulus [14, 17]. Finally, we compare the estimation means and variances of both networks versus Bayesian predictions under different combinations of parameters. For each network, the Bayesian prediction is substituting the estimation mean and variance under single cue conditions of the same network into Eqs. (3 and 4). Fig. 2D and 2E compares the network estimation results with Bayesian predictions, indicating that the decentralized system actually integrate two cues optimally.

4. Conclusions and Discussions

In the present study, we have proposed a novel decentralized architecture to implement Bayesian information integration. The decentralized system is composed of two coupled networks, and in general can be extended into multiple connected ones. Each network in the decentralized system is modeled as a CANN and individually receives the stimulus information from an independent cue. We found

the information integration is achieved by the reciprocal connections between networks. Our study may have a far-reaching implication on neural information processing. It suggests that the brain can implement efficient information integration in a distributive manner through reciprocal connections between cortical regions. Compared to centralized information integration, distributive processing is more robust to local failures and facilitates parallel computation.

Acknowledgments

This work was supported by the National Basic Research Program of China (2014CB846101 (SW, MJR), the National Natural Science Foundation of China (31371109 (MJR), 31261160495 (SW)), and the Research Grants Council of Hong Kong (N_HKUST606/12).

References

- [1] R. Bertin, A. Berthoz, *Experimental Brain Research* **154**, 11 (2004).
- [2] Y. Gu, D. E. Angelaki, G. C. DeAngelis, *Nature Neuroscience* **11**, 1201 (2008).
- [3] A. Chen, G. C. DeAngelis, D. E. Angelaki, *The Journal of Neuroscience* **33**, 3567 (2013).
- [4] B. E. Stein, T. R. Stanford, *Nature Reviews Neuroscience* **9**, 255 (2008).
- [5] R. J. van Beers, A. C. Sittig, J. J. D. van der Gon, *Journal of Neurophysiology* **81**, 1355 (1999).
- [6] M. O. Ernst, M. S. Banks, *Nature* **415**, 429 (2002).
- [7] R. A. Jacobs, *Vision Research* **39**, 3621 (1999).
- [8] J. J. Clark, A. L. Yuille, *Data fusion for sensory information processing systems*, vol. 105 (Springer Science & Business Media, 2013).
- [9] W. J. Ma, J. M. Beck, P. E. Latham, A. Pouget, *Nature Neuroscience* **9**, 1432 (2006).
- [10] T. Ohshiro, D. E. Angelaki, G. C. DeAngelis, *Nature neuroscience* **14**, 775 (2011).
- [11] D. Boussaoud, L. G. Ungerleider, R. Desimone, *Journal of Comparative Neurology* **296**, 462 (1990).
- [12] J. S. Baizer, L. G. Ungerleider, R. Desimone, *The Journal of Neuroscience* **11**, 168 (1991).
- [13] K. Zhang, *The Journal of Neuroscience* **16**, 2112 (1996).
- [14] S. Deneve, P. Latham, A. Pouget, *Nature Neuroscience* **2**, 740 (1999).
- [15] S. Wu, S.-I. Amari, H. Nakahara, *Neural Computation* **14**, 999 (2002).
- [16] W. Zhang, S. Wu, *Neural Computation* **24**, 1695 (2012).
- [17] S. Wu, K. Hamaguchi, S.-i. Amari, *Neural Computation* **20**, 994 (2008).
- [18] C. A. Fung, K. M. Wong, S. Wu, *Neural Computation* **22**, 752 (2010).