A Neural Implementation of Optimal Cue Integration

Wei Ji Ma, Jeff Beck, and Alexandre Pouget

Presented by Min-Je Kim
College of Liberal Studies, Major Biological Sciences
2011. 10. 11
Problem of Former Assumptions

- Former Assumptions
  - Neurons fire with Poisson statistics
  - Neuronal noise is independent

- These assumptions are not always satisfied in real, considering that neurons are dependent to each other.

- So we need more general likelihood function which can stand alone without the assumptions above.
Poisson Like Variability

\[ p(r|s) = \frac{\Phi(r)}{\eta(s)} e^{h(s) \cdot r} \]

\((-\cdot):\) Arbitrary function of \(r\)

\((-\cdot):\) Normalization factor

\((-\cdot):\) Matrix on tuning function

When independent,

\[ p(r|s) = \prod_{i=1}^{N} p(r_i|s) = \prod_{i=1}^{N} \frac{e^{-gf_i(s)} (gf_i(s))^{r_i}}{r_i!} \]

\[ \Phi(r) = \frac{1}{\prod_i r_i!} \]

\[ \eta(s) = e^{g \sum_i f_i(s)} \]

\[ h_i(s) = \log f_i(s) \]

\[ h_i(s) \cdot r_i = r_i \cdot \log f_i(s) \]

\[ e^{r_i \log f_i(s)} = (f_i(s))^{r_i} \]

\[ h'(s) = \Sigma^{-1}(s) f'(s) \]

\[ \Sigma(s) = \begin{pmatrix}
  f_1(s) & 0 & \cdots & 0 \\
  0 & f_2(s) & \cdots & \vdots \\
  \vdots & \vdots & \ddots & 0 \\
  0 & \cdots & 0 & f_N(s)
\end{pmatrix} \]
Dealing with “Irrelevant” Variables

• Nuisance parameters
  – There almost always exist parameters not of interest to the observer for estimation of stimulus.
  – These parameters just add complexity of task, meaninglessly.

• The problem raises when nuisance parameters affect the gain, which is unknown.

• There are two ways to approach this problem
  – Estimate the gain with external operation (which is burdensome)
  – Bayes-optimal one
    • Average out the influence of unknown parameter, done as follows.

\[
p(s|r) \propto p(r|s) = \int p(r|s, g) p(g) \, dg, \quad \text{when } g \text{ and } s \text{ are independent.}
\]  
\[
(21.17)
\]
Dealing with "Irrelevant" Variables

\[
p(s|r) \propto p(r|s) = \int p(r|s, g) p(g) \, dg, \quad (21.17)
\]

\[
p(r|s) = \frac{\Phi(r)}{\eta(s)} e^{h(s) \cdot r} \quad (21.13)
\]

when \( g \) and \( s \) are independent.

\[
p(r|s, g) = \frac{\Phi(r, g)}{\eta(s)} e^{h(s) \cdot r} \quad (21.18)
\]

\[
p(s|r) \propto \int \frac{\Phi(r, g)}{\eta(s)} e^{h(s) \cdot r} p(g) \, dg
\]

\[
= \frac{e^{h(s) \cdot r}}{\eta(s)} \int \Phi(r, g) p(g) \, dg
\]

\[
\propto \frac{e^{h(s) \cdot r}}{\eta(s)}, \quad (21.19)
\]

\[
\therefore p(s|r) \propto \frac{e^{h(s) \cdot r}}{\eta(s)}
\]
Dealing with “Irrelevant” Variables

$$p(s|r) \propto \frac{e^{h(s) \cdot r}}{\eta(s)}$$

$$\frac{d}{ds} \log \eta(s) = \frac{1}{\eta(s)} \frac{d}{ds} \int \Phi(r, g) e^{h(s) \cdot r} \, dr$$

$$= h'(s) \cdot \int \frac{r \Phi(r, g) e^{h(s) \cdot r}}{\eta(s)} \, dr$$

$$= h'(s) \cdot \langle r \rangle = h'(s) \cdot gf(s).$$

(21.20)

$$\frac{d}{ds} \log \eta(s) = h'(s) \cdot gf(s)$$

Differentiate with respect to $g$
(Note that $g$ and $s$ are independent.)

$$0 = h'(s) \cdot f(s).$$

∴ $\eta(s) = \text{const. for } s \Rightarrow \text{can be absorbed to } \Phi(r, g)$

Also, $h(s)$ does not depend on $g$.

∴ $\Sigma^{-1}(s)f'(s) = h'(s) = h'(s, g) = \Sigma^{-1}(s, g)f'(s, g)$

here, as $f(s, g) = gf(s),$

$$\Sigma^{-1}(s)f'(s) = \Sigma^{-1}(s, g)f'(s, g) = \Sigma^{-1}(s, g)gf'(s)$$

$$\Sigma^{-1}(s) = g \cdot \Sigma^{-1}(s, g)$$

∴ $\Sigma(s, g) = g \cdot \Sigma(s)$

$$p(r|s, g) = \frac{\Phi(r, g) e^{h(s) \cdot r}}{\eta(s)} \quad (21.18)$$
Dealing with “Irrelevant” Variables

- So we can say that neural variability and posterior distribution should be in the form below, respectively.

\[ p(r|s) = \Phi(r, g) e^{h(s) \cdot r} \quad (21.21) \]

\[ p(s|r) \propto e^{h(s) \cdot r} \quad (21.22) \]

- We can also say that the overall process is not complicated by the existence of nuisance parameter \( c \).

- Although whether cortical variability follows Poisson-like distribution is not demonstrated, such assumptions is reported to explain the phenomenon well.

\[ p(s|r) \propto \frac{e^{h(s) \cdot r}}{\eta(s)} \]

\[ \eta(s) = \text{const. for } s \]
Optimal Cue Integration with Poisson-like Populations

- \( r_{AV} = r_A + r_V \) (21.11) can also be used for Poisson-like case.
  - NOTA BENE. This is actually one of main claims for this paper.
  - Assumption: \( h(s) \) is same for auditory and visual inputs.

\[
p (r_{AV} | s, g_V, g_A) \\
= \int p (r_V | s, g_V) \times p (r_A = r_{AV} - r_V | s, g_A) \\
= \int \Phi_V (r_V, g_V) e^{h(s) \cdot r_V} \times \Phi_A (r_{AV} - r_V, g_A) e^{h(s) \cdot (r_{AV} - r_V)} dr_V \\
= \int \Phi_V (r_V, g_V) \times \Phi_A (r_{AV} - r_V, g_A) e^{h(s) \cdot r_{AV}} dr_V \\
= e^{h(s) \cdot r_{AV}} \int \Phi_V (r_V, g_V) \times \Phi_A (r_{AV} - r_V, g_A) dr_V. \quad (21.23)
\]

\[
p (s | r_{AV}) \propto e^{h(s) \cdot r_{AV}} \quad (21.24)
\]

\[
p (s | r_{AV} = r_A + r_V) \propto p (r_A | s) p (r_V | s), \quad (21.25)
\]

\[
p (s | r) \propto e^{h(s) \cdot r} \quad (21.22)
\]
Optimal Cue Integration with Poisson-like Populations

- \( r_{AV} = r_A + r_V \) (21.11) can also be used for Poisson-like case.
  - NOTA BENE. This is actually one of main claims for this paper.
  - Assumption : \( h(s) \) is same for auditory and visual inputs.
  - Exception : When tuning curve differ between auditory and visual inputs.
  - However, though \( h_A \) and \( h_V \) differ, these can be converted to same matrix by linear combination.

\[
\begin{align*}
    r_{AV} &= W_A^T r_A + W_V^T r_V \quad (21.26) \\
    h_A(s) &= W_A H(s) \\
    h_V(s) &= W_V H(s)
\end{align*}
\]

- Adjusting these weights may be onerous, but this cannot be the problem as the weights are learned once and for all.
Optimal cue integration can be performed and guaranteed by simply combining two cues linearly, if neural variability is Poisson-like.

Figure 21.2 Optimal cue integration with probabilistic population codes. The cues elicit activity in input populations $r_A$ and $r_V$, indicated by green and blue dots. The dialogue boxes show the probability distributions over the stimulus encoded in each population on a single trial. A simple linear combination of the population patterns of activity, $r_{AV} = W_A r_A + W_V r_V$, guarantees optimal cue integration, if neural variability is Poisson-like. Optimal cue integration means that the probability distribution over the stimulus encoded in the multisensory population is a product of the distributions encoded in the unisensory populations, i.e. $p(s|r_{AV}) \propto p(s|r_A)p(s|r_V)$. The synaptic weight matrices $W_A$ and $W_V$ depend on the tuning curves and covariance matrices of the input populations, but they do not have to be adjusted over trials.
A Physiological Prediction

• **Superadditivity**
  – Earlier work had claimed that multisensory activity evoked by both cues is *more than just sum of the activities* evoked by individual cues.

• **Additivity**
  – Recent work claims that multisensory activity involved with optimal cue integration is same as a *linear combination of input cues*.

\[
r_{AV} = W_A^T \cdot r_A + W_A^T \cdot r_V
\]
Relating Back To Behavior

- It was earlier assumed that $p(s|r)$ follows Gaussian distribution.

$$p(s|r) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} \propto e^{-\frac{1}{2\sigma^2} s^2 + \frac{\mu}{\sigma^2} s} = e^{-\frac{1}{2} s^2 \cdot a(r) + s \cdot b(r)}$$

$$= e^{\left(-\frac{1}{2} s^2 \cdot a + s \cdot b\right) \cdot r} = e^{h(s) \cdot r}$$

while $a(r) = a \cdot r = \frac{1}{\sigma^2}$

$b(r) = b \cdot r = \frac{\mu}{\sigma^2}$

as $p(s|r) \propto e^{h(s) \cdot r}$
Relating Back To Behavior

\[ r_{AV} = r_A + r_V \quad a \quad a \cdot r_{AV} = a \cdot r_A + a \cdot r_V \]

\[ \frac{1}{\sigma^2_{AV}} = \frac{1}{\sigma^2_A} + \frac{1}{\sigma^2_V} \]  (21.30)

\[ a \cdot r = \frac{1}{\sigma^2} \]

\[ r_{AV} = r_A + r_V \quad b \quad b \cdot r_{AV} = b \cdot r_A + b \cdot r_V \]

\[ \frac{\mu_{AV}}{\sigma^2_{AV}} = \frac{\mu_A}{\sigma^2_A} + \frac{\mu_V}{\sigma^2_V} \]  (21.31)

\[ b \cdot r = \frac{\mu}{\sigma^2} \]
Relating Back To Behavior

\[
\frac{1}{\sigma_{AV}^2} = \frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2} \quad (21.30)
\]

\[
\frac{\mu_{AV}}{\sigma_{AV}^2} = \frac{\mu_A}{\sigma_A^2} + \frac{\mu_V}{\sigma_V^2} \quad (21.31)
\]

- These are the single-trial version of the optimal cue combination. For many trials, single tried value \( p(s|r) \) should be turned into a form, \( \hat{s}(r) \).

- Assumption:
  - The optimal estimator is the maximum-likelihood estimator.
    - This is satisfied when \( p(s|r) \) is maximized.
  - Prior \( p(s) \) is uniform.
  - \( p(s|r) \) follows Gaussian distribution.

- Result:
  - The optimal estimator is same as the mean of \( p(s|r) \).
Relating Back To Behavior

\[
\frac{1}{\sigma_{AV}^2} = \frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2} \quad \text{(21.30)}
\]

\[
\frac{\mu_{AV}}{\sigma_{AV}^2} = \frac{\mu_A}{\sigma_A^2} + \frac{\mu_V}{\sigma_V^2} \quad \text{(21.31)}
\]

- **Result:**
  - The optimal estimator is same as the mean of \( p(s|r) \).

This offers a particularly easy way to check optimality of \( r_{AV} = r_A + r_V \). Taking the average on both sides, we find \( g_{AV} = g_A + g_V \), and since \( I(s) \) is proportional to \( g \) according to Eq. 21.33, it follows from Eq. 21.32 the estimate's inverse variances sum:

\[
\frac{1}{\sigma_{\text{estimator,AV}}^2} = \frac{1}{\sigma_{\text{estimator,A}}^2} + \frac{1}{\sigma_{\text{estimator,V}}^2} \quad \text{(21.34)}
\]

\[
\langle \tilde{s} \rangle = \sigma^2 b \cdot \langle r \rangle = \sigma_{\text{estimator}}^2 b \cdot g f(s) \quad \text{(21.35)}
\]

\[
\frac{\langle \tilde{s} \rangle_{AV}}{\sigma_{\text{estimator,AV}}^2} = \frac{\langle \tilde{s} \rangle_A}{\sigma_{\text{estimator,AV}}^2} + \frac{\langle \tilde{s} \rangle_V}{\sigma_{\text{estimator,AV}}^2} \quad \text{(21.36)}
\]

where the average \( \langle \cdot \rangle \) is over \( r \) drawn from \( p(r|s) \). Fisher information and the Cramér-Rao can be applied to any distribution. For Poisson-like variability, there are several ways to express Fisher information:

\[
I(s) = -\sum h''(s) \cdot g f(s) = g h'(s) \cdot \Sigma(s) h'(s)
\]

\[
= g f'(s) \cdot \Sigma^{-1}(s) f'(s) \quad \text{(21.33)}
\]

This is the relationship found in behavior. Similarly, for the mean estimate, we have from Eq. 21.29:
Optimal Cue Integration with Biophysical Populations

• The model used for the paper is completely characterized by neuron’s firing rates and therefore without dynamics of neurons’ population. That is, the model was top-down driven, while the real neurons do have their own complex network.

• However, this approach is fundamental for describing neuronal patterns of firing. So the authors expect this Poisson-like model can also be applied for realistic modeling of the brain.
The Big Picture

1. **Stimuli s** → **Input population activities r** → **Neural operations** → **Output population activity** → **Motor action/judgment**

   - **Encoding**
   - **Bayes**

2. **Assuming a form of neural variability**

3. **Probability distributions ρ(s|r)** → **Probabilistic computations** → **Output probability distribution**

   - **neural activity space**
   - **probability space**
Summary

• Optimal cue integration can be implemented by population coding.

• Brain is likely to perform Bayes optimized operation which can average out irrelevant variables.

• Optimal cue integration can be performed and is guaranteed by simply combining two independent cues linearly, even if neurons are dependent to each other and neuronal variability follows Poisson-like distribution.