Binding Problem for Input vs. Output Representations and the Role of the Thalamus in Its Solution

As presented at the Computational Neurobiology Lab at Salk: June 14, 2005

Yoonsuck Choe
Department of Computer Science
Texas A&M University
http://faculty.cs.tamu.edu/choe

1

Motivation: The Binding Problem

- Distributed representations lead to the **superposition catastrophe** (von der Malsburg 1986).
- How does the brain piece together partial representations to form a whole?
- Which feature should go along with which?

2

Potential Solution to the Binding Problem

- Timing may be important in solving the problem.
- **Interleave** the activity pattern *over time* (von der Malsburg 1986).

3

Evidence for Temporal Coding

- Gray et al. (1989) and Eckhorn et al. (1988) (and many thereafter) showed that neural representations of coherent object features are **synchronized**.
- **But, that may not be the end of the story!**
**The Main Research Question**

How does the brain **distinguish** between cortical activities that represent:

1. **Questions** posed to the cortex, and
2. **Answers** to those questions?

That is, how can the input and the output of cortical computation be distinguished?

**Input–Output Binding Problem (IOBP)**

Similar to the original binding problem, but not between input representations, **but between input and output representations.**

**Why Is That a Problem at All?**

The problem is nontrivial because:

- The same representation can serve as **both question and answer** at different times, under different contexts.
- The source and the target cortical region will maintain almost **simultaneous activation** while the source region is active.

**Possible Answer: Simply Promote the Output**

Promote (or propagate) cortical activity that are:

1. **Not input-driven,** or
2. **Relatively less input-driven.**

But, **how (and where)** does the brain achieve this?
Possible Neural Basis: The Thalamus

Some clues:

- **Heavy feedback** from the cortex.
- Covered by an inhibitory shell, the **Thalamic Reticular Nucleus (TRN)**.

Image Source: http://mail.biofarm.unibo.it/aunsnc/3dobjb.html

Related Work on the Thalamus

- Sensory relay (see Sherman and Guillery 2001 for a review).
- Sleep rhythms (Destexhe and Sejnowski 2001; Steriade and McCormick 1993; McCormick and Bal 1997) / Epilepsy.
- Synchrony (Llinás and Ribary 1994; Sillito et al. 1994).
- Mediating cortical communication (Guillery and Sherman 2002).
- Cross-modality switching (Crabtree and Isaac 2002).
- Attention (LaBerge 1995; Crick 1984).
- Active blackboard (Mumford 1995; Harth et al. 1987)
- Consciousness (Crick 1984; Taylor 1998).

Dorsal Thalamus-TRN-Cortex Network

- A candidate circuit can be found in the dorsal thalamus-TRN-cortex circuit: **TRN plays a key role**.

Activation Sequence (1/6)

Initially, only $T_1$ receives an afferent sensory input.
The cortical neuron $C_1$, through fast connections, invokes another cortical neuron $C_2$. $C_1$ also sends out feedback to $R_1$ and $T_1$, but these connections are slow. $R_1$ retains the level of excitation in the meanwhile.

Cortical feedback from both $C_1$ and $C_2$ arrives at the TRN, and adds to the existing activity at TRN. Reticular neurons $R_1$ and $R_2$ inhibit each other through fast connections.

The reticular neurons exert inhibition on the thalamic relays. Feedback from $C_1$ is canceled out, while that from $C_2$ is not.
Finally, only $T_2$ is allowed to fire again, reactivating $C_2$ for the second time.

1. TRN neurons must have slow a dynamic ($b$–$d$).
2. Inhibition between reticular neurons must be strong ($e$).
3. Either the cortico-cortical connections must be very fast or the corticothalamic feedback connections are unmyelinated (i.e., very slow; Tsumoto et al. 1978).
4. Interaction between reticular neurons must be fast ($d$).

For each neuron $i$, the membrane potential $V_i$ evolved according to the following dynamic equation:

$$C_i \frac{dV_i}{dt} = I_i(t) - \frac{V_i}{R_i},$$

where $C_i$ is the membrane capacitance, $R_i$ the resistance, and $I_i(t)$ the input contribution to neuron $i$ at time $t$. When $V_i$ reaches a threshold value $\theta_i$, a spike is generated and $V_i$ is reset to 0.0.
A spike generated by a presynaptic neuron $j$ results in a postsynaptic potential (PSP) $s_{ij}$ at a target neuron $i$, which is set to 1.0 at the moment the spike is received and is decayed over time as follows:

$$\frac{ds_{ij}}{dt} = -\frac{s_{ij}}{\tau}$$

(2)

where $\tau$ is the time constant of the PSP.

### Model Parameters

**Table 1: Neuron Parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Thal. Relay ($T_i$)</th>
<th>TRN ($R_i$)</th>
<th>Cortex ($C_i$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacitance $C_i$</td>
<td>0.3</td>
<td>0.6</td>
<td>0.3</td>
</tr>
<tr>
<td>Resistance $R_i$</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Threshold $\theta_i$</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>PSP time constant $\tau_i$</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>

**Table 2: Connection Parameters**

<table>
<thead>
<tr>
<th>Weight $w_{ij}$</th>
<th>$T_j$</th>
<th>$R_j$</th>
<th>$C_j$</th>
<th>$T_i$</th>
<th>$R_i$</th>
<th>$C_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_j$</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_j$</td>
<td>2.0</td>
<td>10.0</td>
<td>0.9</td>
<td>2.0</td>
<td>2.0</td>
<td>0.2</td>
</tr>
<tr>
<td>$C_j$</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Overview of Results

**Core results:**
- Experiment 1: **Direct stimulation** of thalamus or cortex.
- Experiment 2: Selecting **not input-driven** cortical activity.
- Experiment 3: Selecting **less input-driven** cortical activity.

**Predictions under disruptions:**
- Experiment 4: When TRN is fast.
- Experiment 5: When R→T inhibition is weak.
- Experiment 6: When C→C is slow.
- Experiment 7: When R→R is slow.
Exp 1: Thalamic vs. Cortical Stim.

- Thalamic stimulation: No reactivation of the cortex.
- Cortical stimulation: Cortical reactivation through the thalamo-cortical loop.

Exp 2: Input vs. No-Input

- Input-driven cortical activity does not reactivate.
- Cortically induced cortical activity reactivates through the cortex-thalamus-cortex loop.

Exp 3: Strong vs. Weak Input

- Strongly input-driven cortical activity does not reactivate.
- Weakly input-driven cortical activity reactivates through the cortex-thalamus-cortex loop.

Exp 4: Fast TRN dynamics

- With faster TRN dynamics ($C_i = 0.5$), the reticular neurons fail to integrate the thalamic and cortical contributions, and thus timely inhibition is interrupted.
Exp 5: Weak TRN to Thalamus Inhibition

- With lowered R→T weight (2.0), due to the weaker disinhibition effect, loop2 reticular neuron generates more activity to suppress the thalamic relay. As a result, loop2 fails to reactivate the cortex.

Exp 6: Slow Corticocortical Connections

- With longer C→C connection delay, the phases of loop1 and loop2 activities start to drift and become irregular.

Exp 7: Slow intra-TRN connections

- With longer R→R connection delay (1.5), the disinhibition effect did not happen in time to allow loop2 to reactivate the cortex.

Summary of Results

- A thalamocortical model was implemented with parameters derived from functional, anatomical, and physiological considerations.
- The model was successful in detecting and promoting (1) non-input-driven, and (2) less input-driven cortical activity.
Discussion

• How particular answers are generated from the questions?
  – Analogy, inference, association, etc.
• Why need such a round-about? Why not do it in the cortex?
• What about primitive animals without the thalamus?

Discussion (cont’d)
The model does not account for the following:
• Drivers vs. modulators innervating thalamic relays.
• Slowness of TRN is in $I_T$.
• Low-threshold firing in thalamic relay and TRN (burst, as opposed to tonic firing).
• Role of the interneurons in dorsal thalamic nuclei.
• Other inputs to TRN and dorsal thalamus (parabrachial region, brain stem, etc.).
• Higher-order relays: feedback is from layer V, not layer VI.
• Intricate circuitry in the cortex (layers IV, II/III, etc.).

Predictions

• Results from as Exp 1 to Exp 3 would be replicable in in vivo experiments.
• Not just $I_T$ but other currents in TRN may turn out to have a slow dynamic.
• Intra-TRN connectivity will reflect that of its cortical counterpart (majorly in its extent, but maybe also in its broader pattern).
• The time-course of a unit of computation $T_u$ in the cortex would follow:

\[ T_u = T \rightarrow C + C \rightarrow C + C \rightarrow T + T \rightarrow C. \]

  feedforward computation feedback reactivation

Conclusion

• Input–output binding problem (IOBP) may need more attention.
• The thalamo-cortical loop may be able to solve the IOBP.
• It may be important to look at how pieces of circuit properties fall into place in the puzzle.
Acknowledgments

- Insightful comments by James A. Bednar, Bruce H. McCormick, Risto Miikkulainen, Lokendra Shastri, S. Murray Sherman, Ray Guillery, Jason Traesger, Dennis Glanzman, and Yingwei Yu helped clarify several points.

- The simulation results presented here are based on earlier results in (Choe 2002, 2004). All simulations were implemented in XPPAUT by Ermentrout (2002).

- This research was supported in part by Texas A&M University, by the Texas Higher Education Coordinating Board grant ATP#000512-0217-2001, and by the National Institute of Mental Health Human Brain Project grant #1R01-MH66991.

References


