

総説論文「注意の 正規化モデル」の紹介

(所内ゼミでの発表資料の修正版)

産業技術総合研究所

脳神経情報研究部門

一杉裕志

2009-04-08

概要

- 下記の総説論文を紹介する。

The normalization model of attention

(Reynolds JH, Heeger DJ, Neuron. 2009 Jan 29;61(2):168-85)

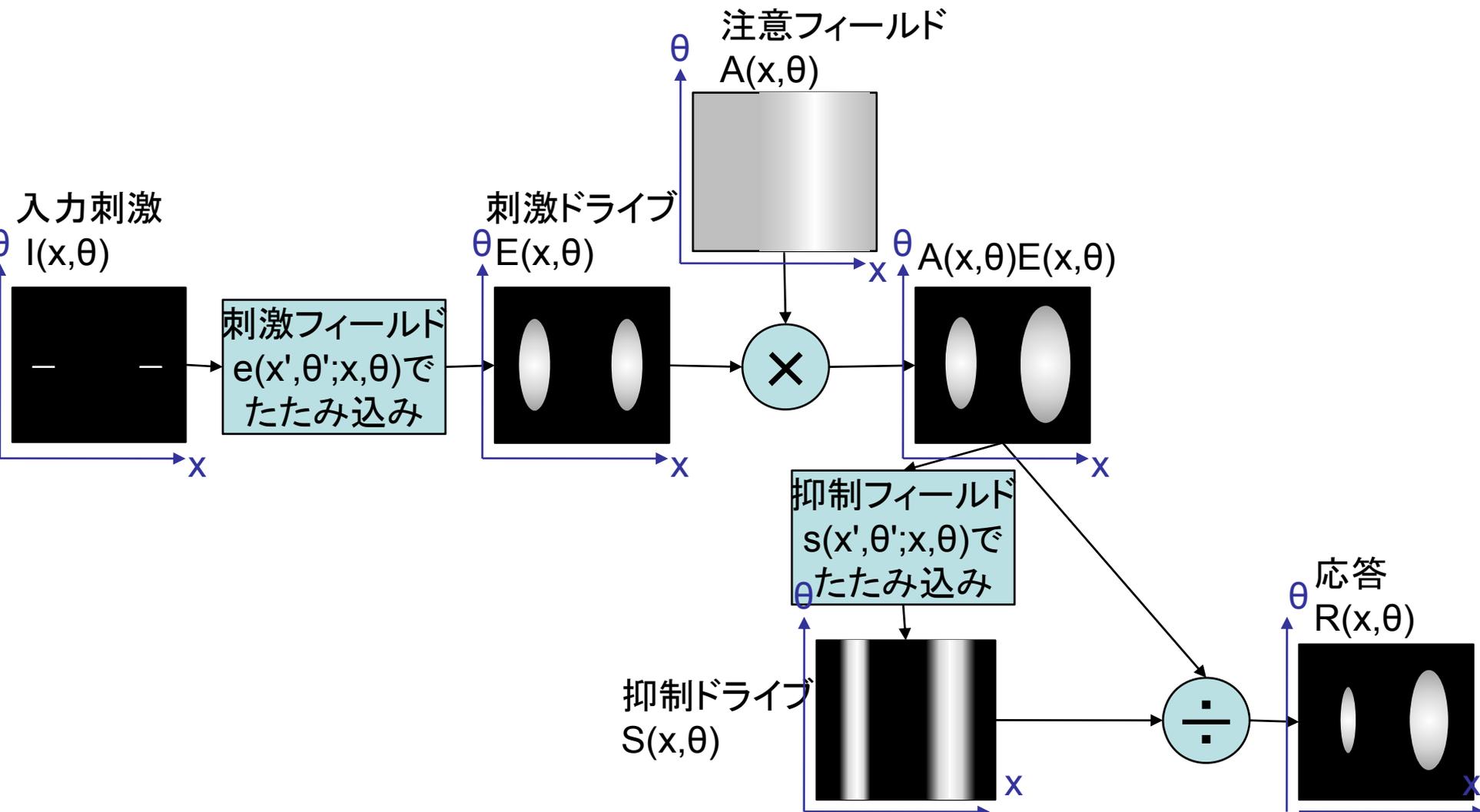
<http://www.ncbi.nlm.nih.gov/pubmed/19186161>

- 視覚野のニューロン応答に対する注意の様々な影響を統一的に説明する、「注意の正規化モデル」という計算論的モデル。
- 論文中の数式は一部分かりにくいので補って説明。
- BESOM モデル(大脳皮質の神経回路モデル)が「注意の正規化モデル」をほぼ包含していることを述べる。

よいモデル(理論)とは？

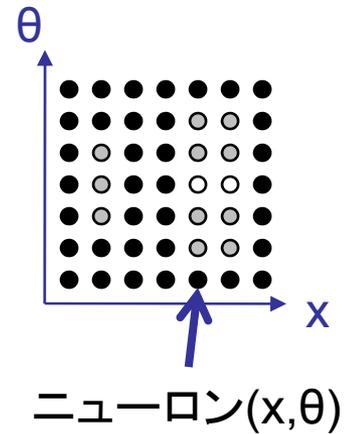
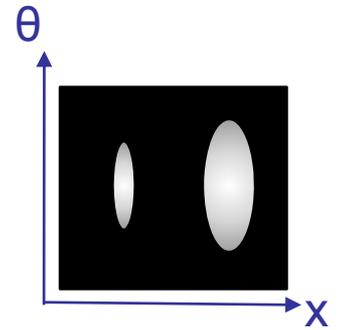
- 少ないパラメタ。
 - 適用範囲が広い。
 - 予言能力が高い。
 - より強力な理論への洞察を与える。
-
- 今回紹介する「注意の正規化モデル」はかなりよいモデルに見える。

刺激からニューロン応答を計算する過程



"Activity Map" の見方

- 受容野の中心の座標が x 、preferred direction が θ であるような、各ニューロンの応答を明暗で示して、2次元平面にならべたもの。

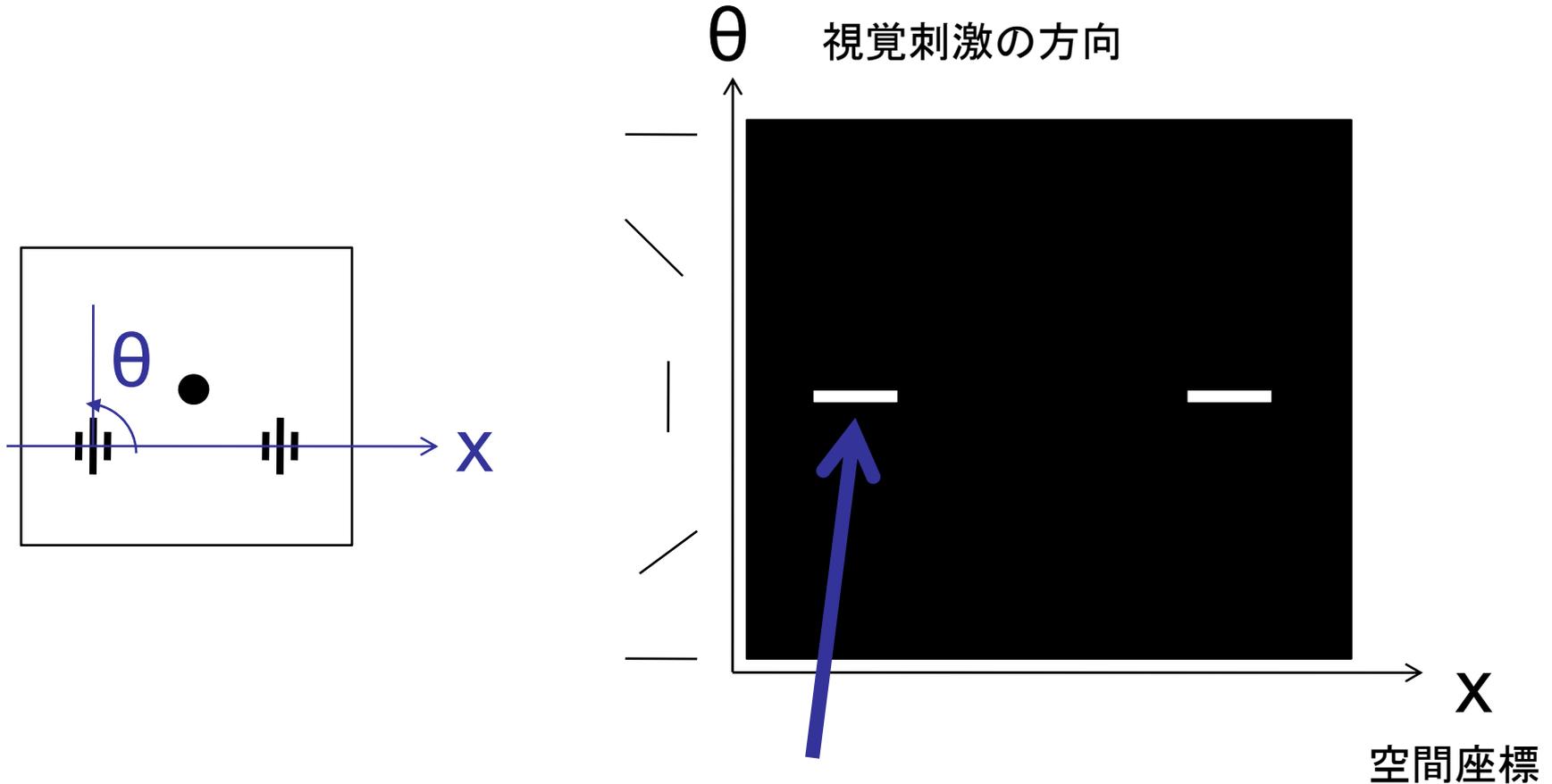


注

- 説明を分かりやすくするため、論文とは異なる以下の変数を用いる。
 - 入力刺激 $I(x, \theta)$
 - 刺激フィールド $e(x', \theta'; x, \theta)$
 - 抑制フィールド $s(x', \theta'; x, \theta)$

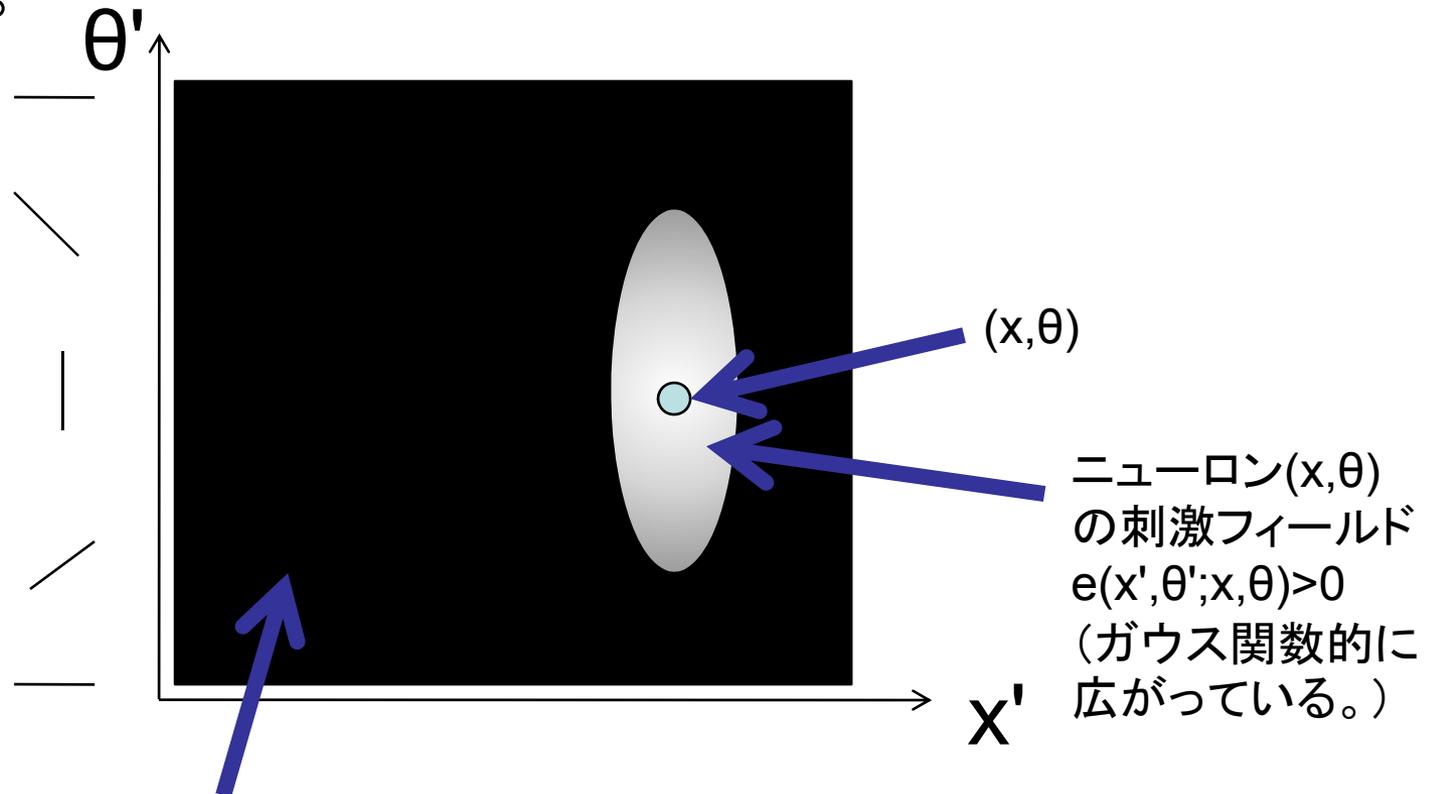
入力刺激(stimulus)

入力刺激の強さを、 x と θ の関数 $I(x, \theta)$ と書くことにする。



刺激フィールド(stimulus field)

受容野の中心の座標が x 、preferred direction の方向が θ であるようなニューロン (x, θ) の、刺激 (x', θ') への応答のしやすさを $e(x', \theta'; x, \theta)$ と書くことにする。

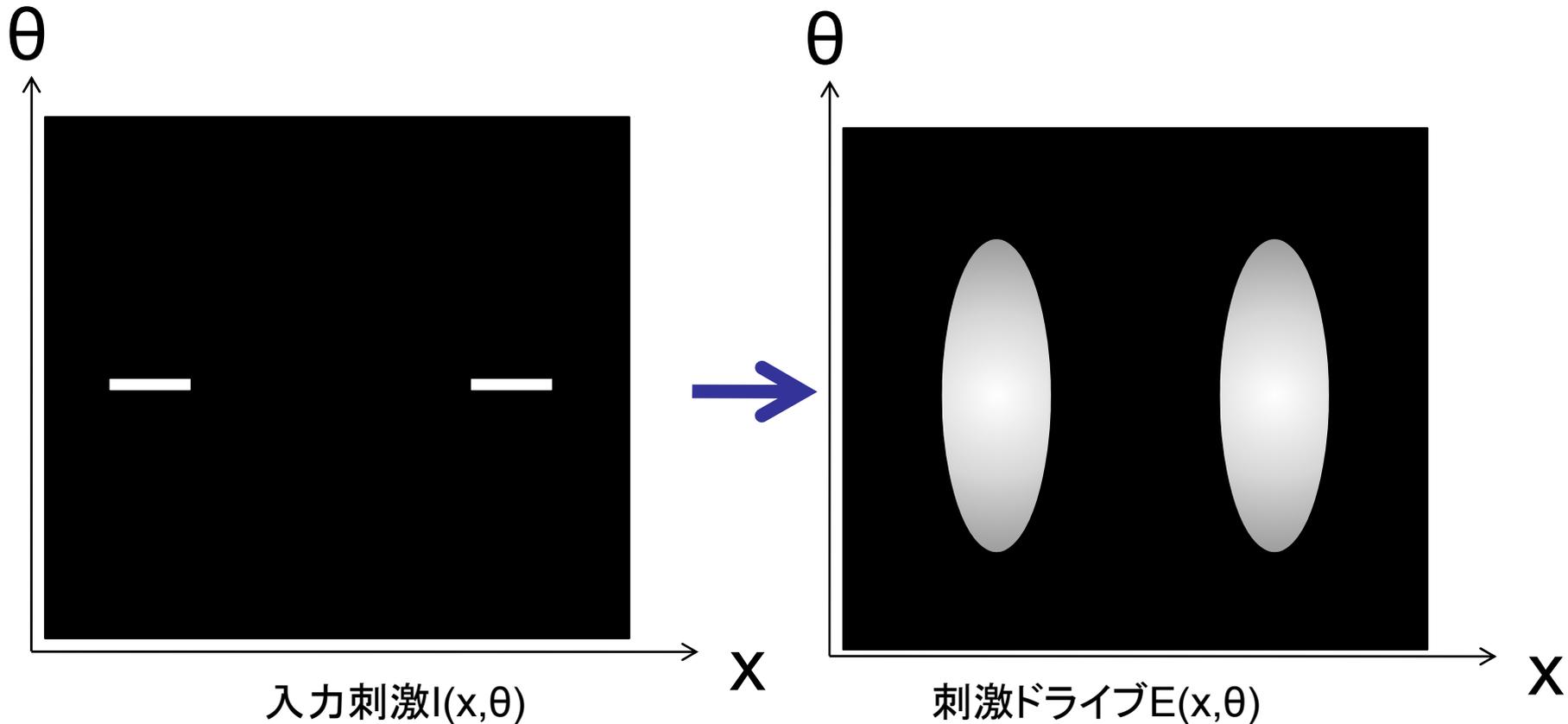


ニューロン (x, θ) の刺激フィールド以外の領域。
 $e(x', \theta'; x, \theta) = 0$

刺激ドライブ(stimulus drive)

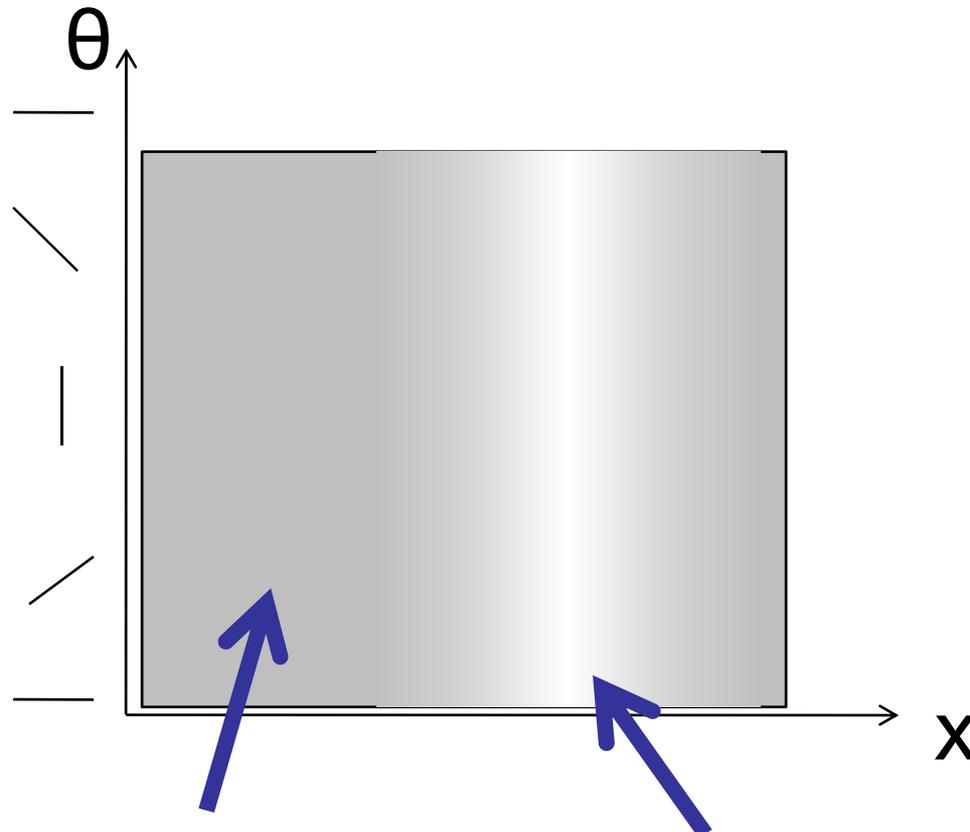
注意と正規化の効果を入れる前の、刺激のみで決まる値。
(入力刺激をガウシアンカーネルでぼかしたものになる。)

$$E(x, \theta) = \iint I(x', \theta') e(x', \theta'; x, \theta) dx' d\theta'$$



注意フィールド(attention field)

座標 x における方向 θ の刺激に対する注意の強さを $A(x, \theta)$ で表す。
サルが自分の意思 (attention strategy) で決める値。

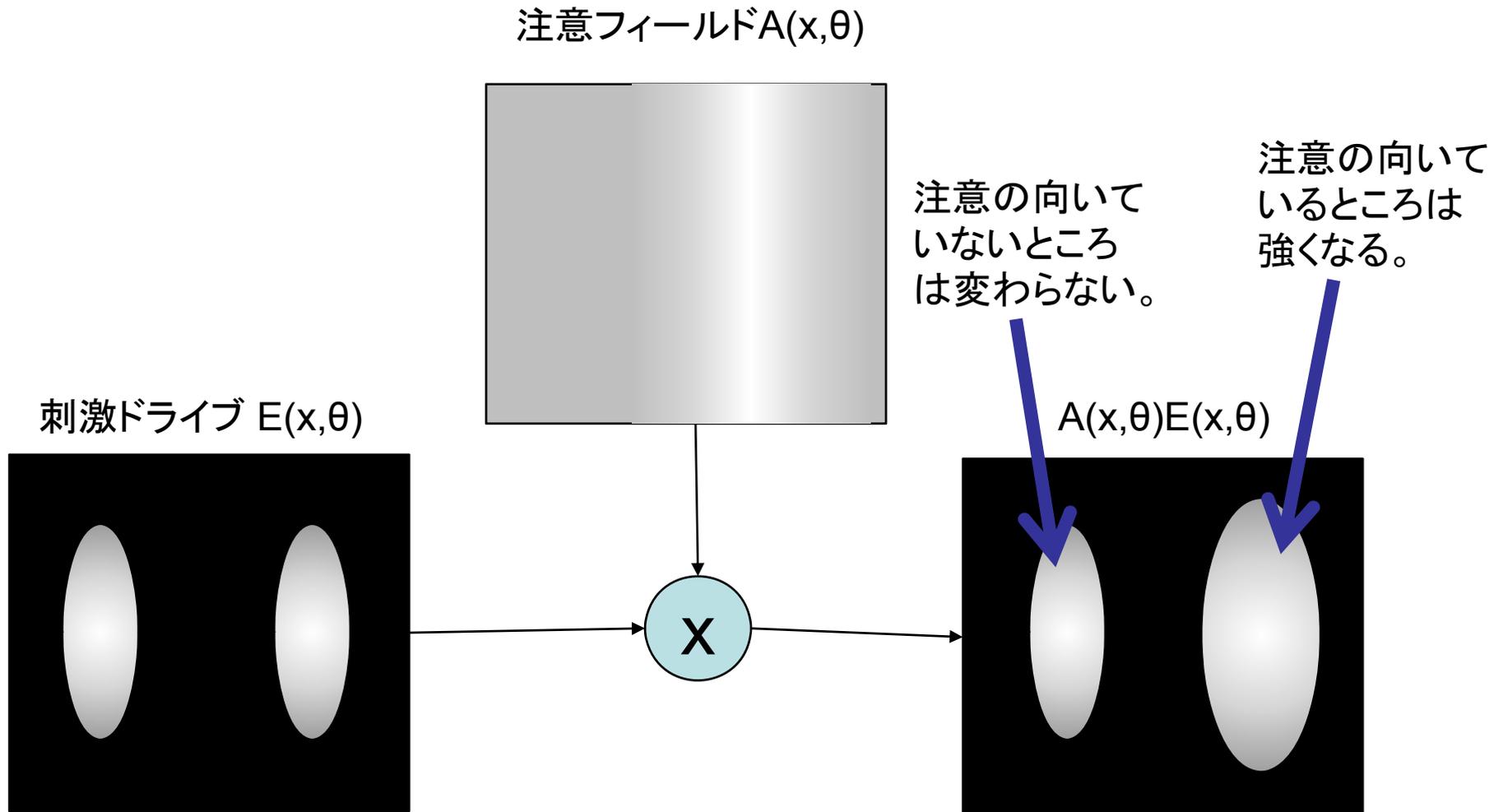


注意を向けていない領域
では $A(x, \theta) = 1$

注意を向けている領域
 $A(x, \theta) > 1$

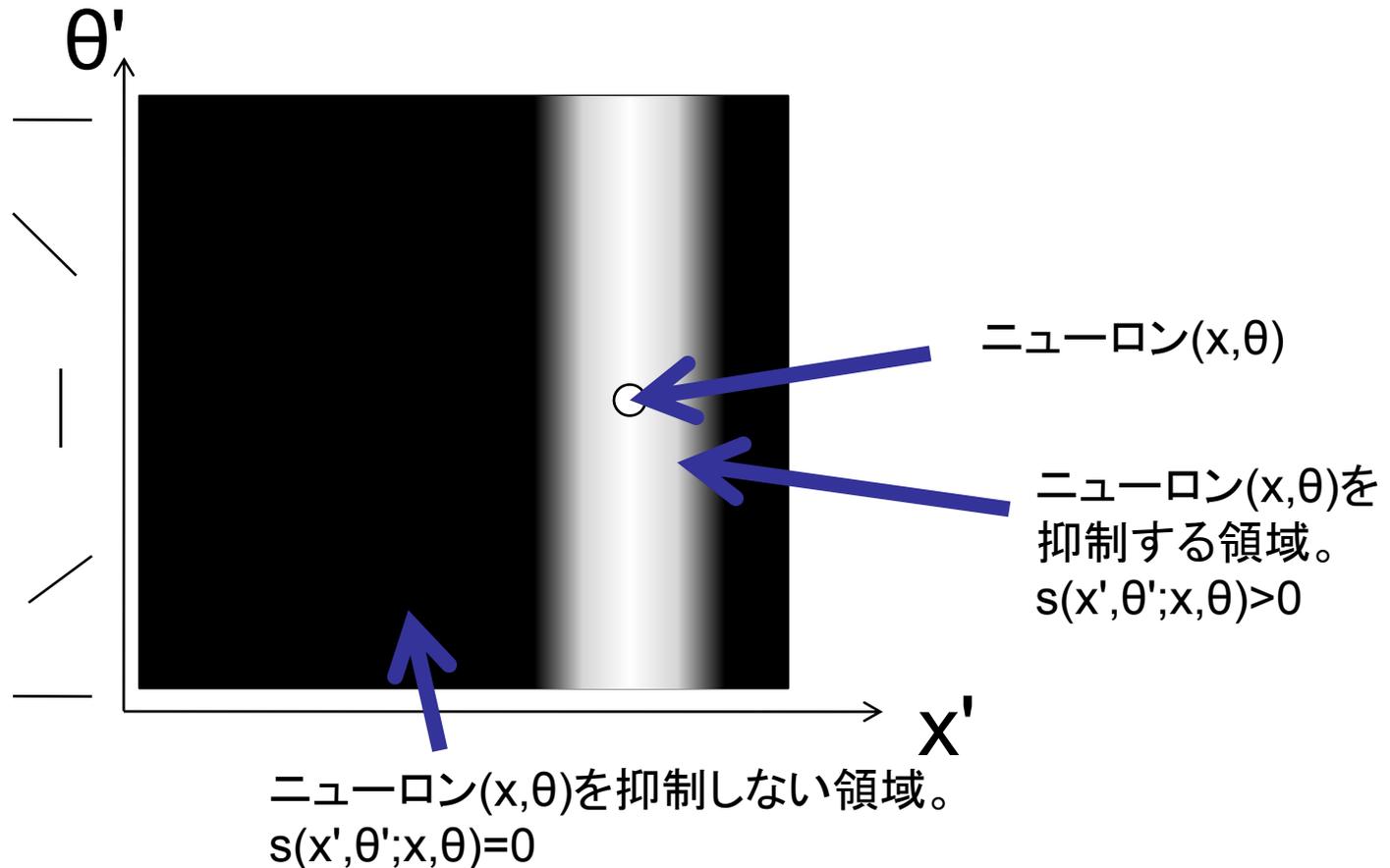
注意フィールドと刺激ドライブの積

刺激ドライブの値と注意フィールドの値を、座標ごとに掛け算。



抑制フィールド(suppressive field)

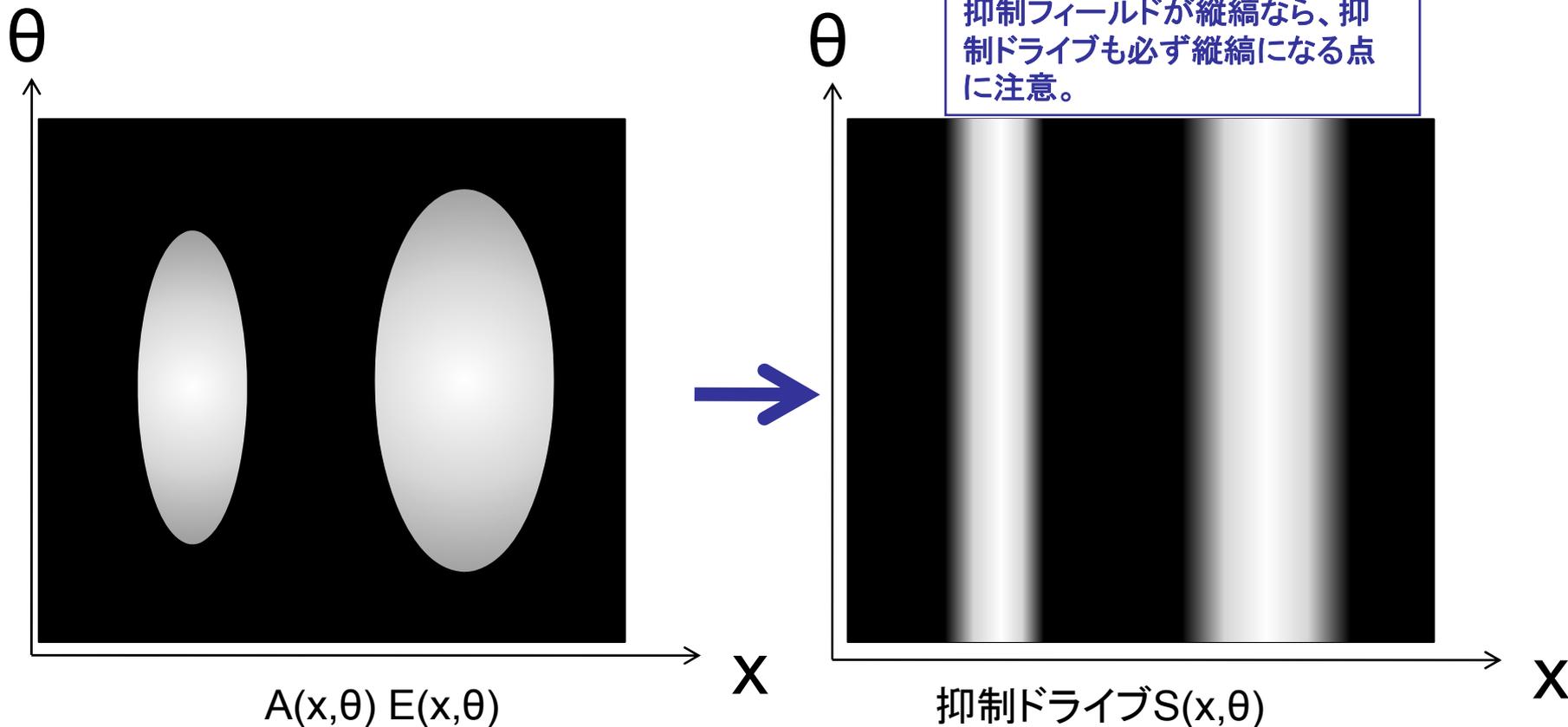
受容野の中心の座標が x 、preferred direction の方向が θ であるようなニューロン(x, θ)の、ニューロン (x', θ') からの抑制の度合いを $s(x', \theta'; x, \theta)$ と書くことにする。



抑制ドライブ(suppressive drive)

ニューロンの応答が周辺のニューロンによってどのくらい抑制されるかを表す値。

$$S(x, \theta) = \iint A(x', \theta') E(x', \theta') s(x', \theta'; x, \theta) dx' d\theta'$$

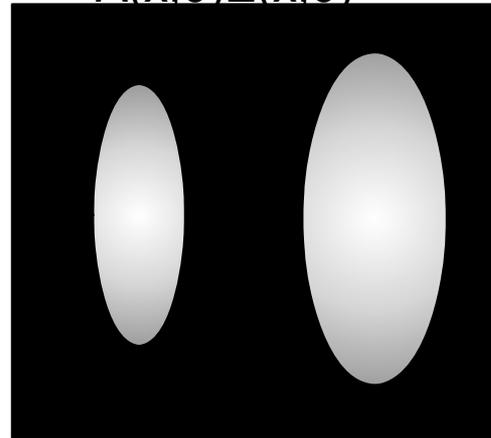


ポピュレーションの応答 (population response)

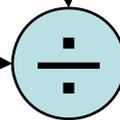
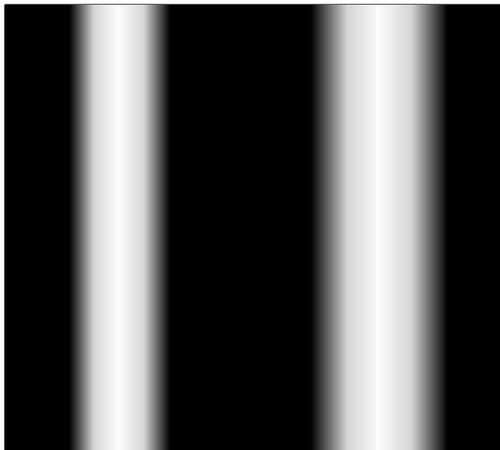
注意フィールドと刺激ドライブの積を座標ごとに抑制ドライブの値で割ったものが、ニューロン応答。

$$R(x, \theta) = | A(x, \theta)E(x, \theta) / (S(x, \theta) + \sigma) |_T$$

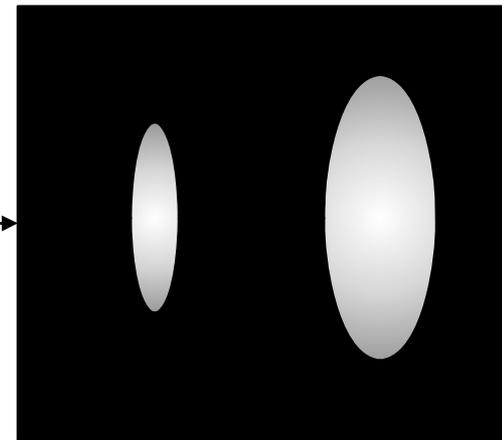
$A(x, \theta)E(x, \theta)$



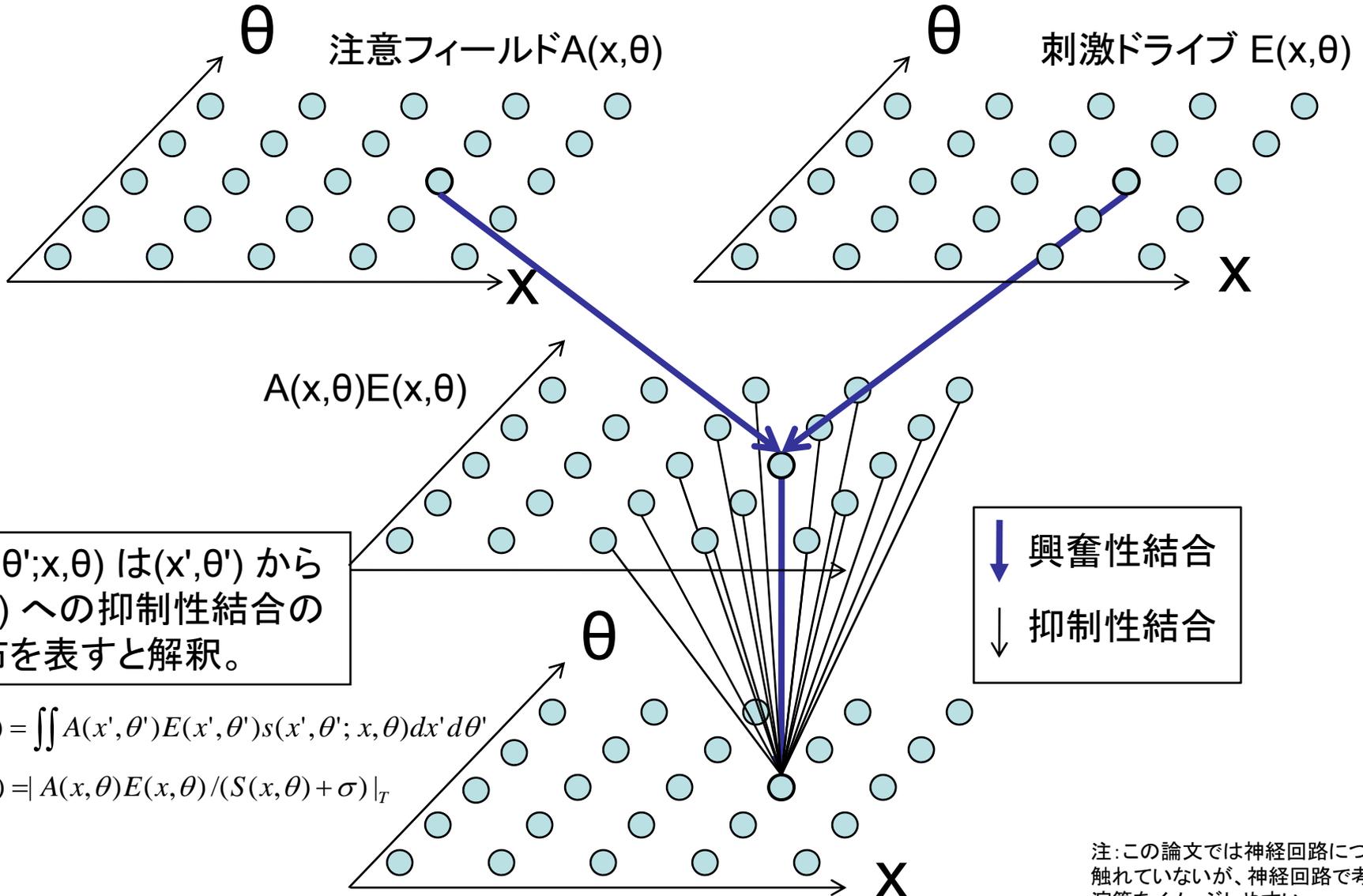
抑制ドライブ $S(x, \theta)$



応答 $R(x, \theta)$



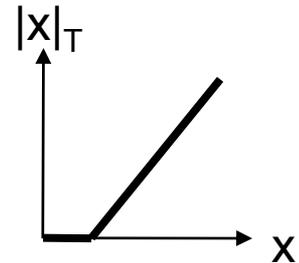
応答を計算する神経回路の例



注: この論文では神経回路については触れていないが、神経回路で考えると演算をイメージしやすい。

補足

- T, σ は非負の定数。
- $|\cdot|_T$ は閾値 T に関する rectification 。 matlab のコードには出ておらず詳細不明。こんな関数を想定？

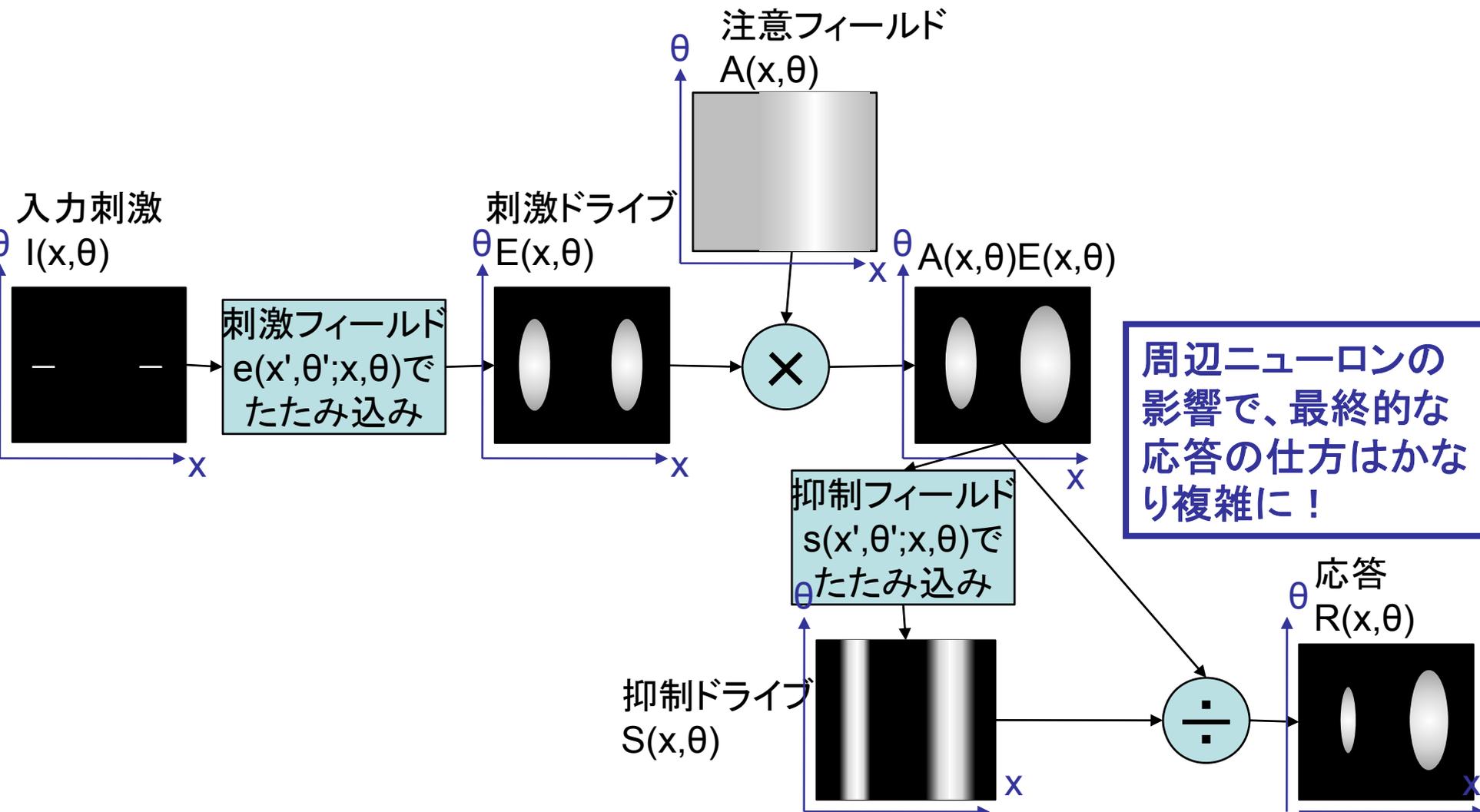


- シミュレーションによっては、ベースラインの活動を表す2つの定数パラメタを追加。

$$E(x, \theta) = \iint I(x', \theta') e(x', \theta'; x, \theta) dx' d\theta' + \text{baselineMod}$$

$$R(x, \theta) = A(x, \theta) E(x, \theta) / (S(x, \theta) + \sigma) + \text{baselineUnmod}$$

正規化モデルの全体像



シミュレーション

- matlab のコードが公開されている。
 - (動かしてはいないが、ソースは大体読んだ。)
- 過去の**多様かつ複雑**な実験結果が定性的に再現されている。

コントラストー応答関数の注意による 変化 [Reynolds et al., 2000][Williford and Maunsell, 2006]

Figure 3.

ポイント

- 受容野、視覚刺激、空間的注意のサイズによってコントラスト-応答関数は複雑に変化するが、モデルは過去の実験結果を定性的に再現できる。
- V4(Fig.3)もMT(Fig.4)も同じモデルで説明できる。
- 受容野、視覚刺激、空間的注意のサイズのすべての組み合わせが過去に実験されているわけではない。
→ 今後実験でモデルの予言を検証可能。

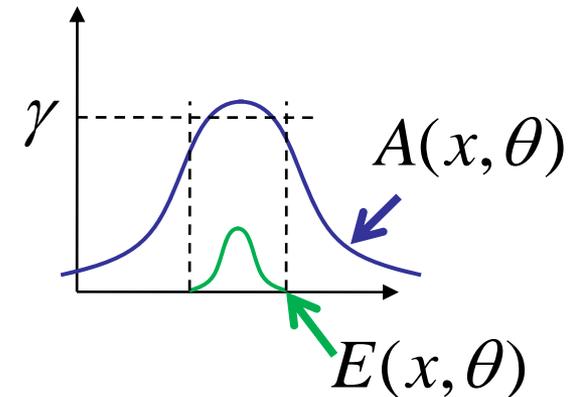
刺激のサイズが小さい場合(1/3)

(論文に書かれている説明では理解不能なので、独自に補足した説明を試みる。)

- 仮定:

- 注意フィールドは刺激ドライブを覆うほど十分に広い、すなわち、 $E(x, \theta) > 0$ ならば $A(x, \theta) \approx \gamma$ とする。この時、すべての x, θ について下記の近似が成り立つ。

$$A(x, \theta)E(x, \theta) \approx \gamma E(x, \theta)$$



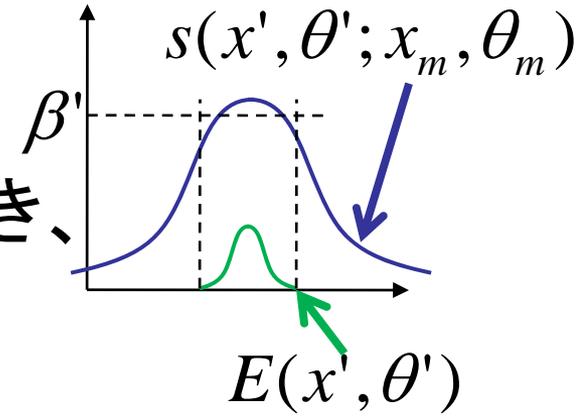
- 仮定:

- 刺激ドライブはコントラスト c に比例 $E(x, \theta) \propto c$

刺激のサイズが小さい場合(2/3)

- 仮定:

- 計測するニューロンが (x_m, θ_m) のとき、抑制フィールドも十分広いとすると下記の近似が成り立つ。



$$S(x_m, \theta_m) = \iint A(x', \theta') E(x', \theta') s(x', \theta'; x_m, \theta_m) dx' d\theta'$$

$$\approx \iint \gamma E(x', \theta') \beta' dx' d\theta'$$

$$= \gamma \beta' \iint E(x', \theta') dx' d\theta'$$

$$\approx \beta \gamma c \quad (\beta \text{ は比例定数})$$

(論文の式(7)ではなぜか β がない。)

刺激のサイズが小さい場合(3/3)

$$R(x_m, \theta_m) = |A(x_m, \theta_m)E(x_m, \theta_m) / (S(x_m, \theta_m) + \sigma)|_T$$

$$\approx A(x_m, \theta_m)E(x_m, \theta_m) / (S(x_m, \theta_m) + \sigma)$$

$$= \gamma\alpha c / (\beta\gamma c + \sigma) = \alpha c / (\beta c + \sigma / \gamma) \quad (E(x_m, \theta_m) = \alpha c \text{ とした})$$

$\beta c \ll \sigma / \gamma$ のとき $R(x_m, \theta_m) \approx \alpha c / (\sigma / \gamma) = \alpha\gamma c / \sigma$

注意の量とコントラストに比例

$\beta c \gg \sigma / \gamma$ のとき $R(x_m, \theta_m) \approx \alpha c / \beta c \approx \alpha / \beta$

注意の量とコントラストによらず一定

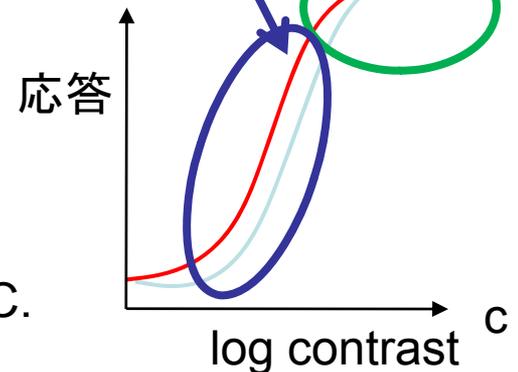


Figure 3C.

刺激のサイズが大きく注意のサイズが小さい場合(1/2)

- 抑制ドライブは、注意に比例する量と、注意によらない量に分解できる。どちらもコントラストに比例すると仮定。

$$S(x_m, \theta_m) \approx \beta_1 \gamma c + \beta_2 c \quad (\text{論文の式(8)ではなぜか}\beta_1\text{がない。})$$

$\beta_1 \gamma c$ は、注意の向いている、測定ニューロンの近傍ニューロンからの抑制。

$\beta_2 c$ は、注意の向いていない、測定ニューロンの周辺ニューロンからの抑制。

刺激のサイズが大きく注意のサイズが小さい場合(2/2)

$$R(x_m, \theta_m) \approx A(x_m, \theta_m)E(x_m, \theta_m)/(S(x_m, \theta_m) + \sigma)$$
$$= \alpha\gamma c / (\beta_1\gamma c + \beta_2 c + \sigma)$$

$c \ll \sigma$ のとき $R(x_m, \theta_m) \approx \alpha\gamma c / \sigma$

注意の量とコントラストに比例

$c \gg \sigma$ のとき $R(x_m, \theta_m) \approx \alpha\gamma c / (\beta_1\gamma c + \beta_2 c) \approx \alpha / (\beta_1 + \beta_2 / \gamma)$

注意の影響はあるが、コントラストによらず一定

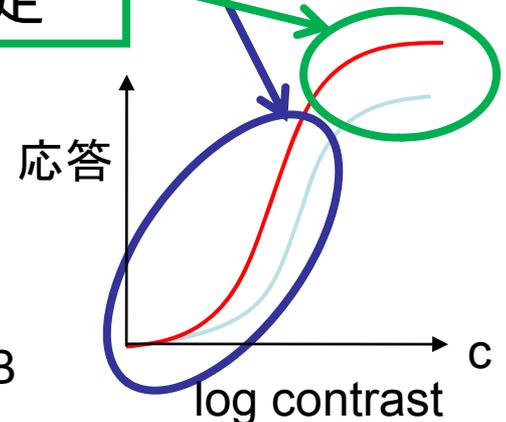
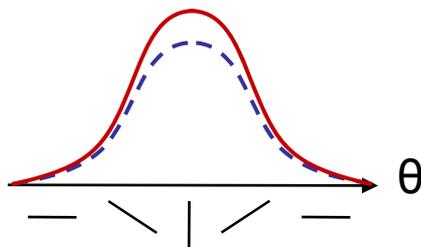


Figure 2B

空間的注意はチューニングカーブを等比率で増加させる [McAdams and Maunsell, 1999]

Figure 5.



数式による説明(1/2)

- 仮定:

- 注意フィールドの値は方位に依存しない。

$$A(x, \theta) = A(x)$$

- 抑制フィールドの値は方位に依存しない。

$$s(x', \theta'; x, \theta) = s(x'; x)$$

数式による説明(2/2)

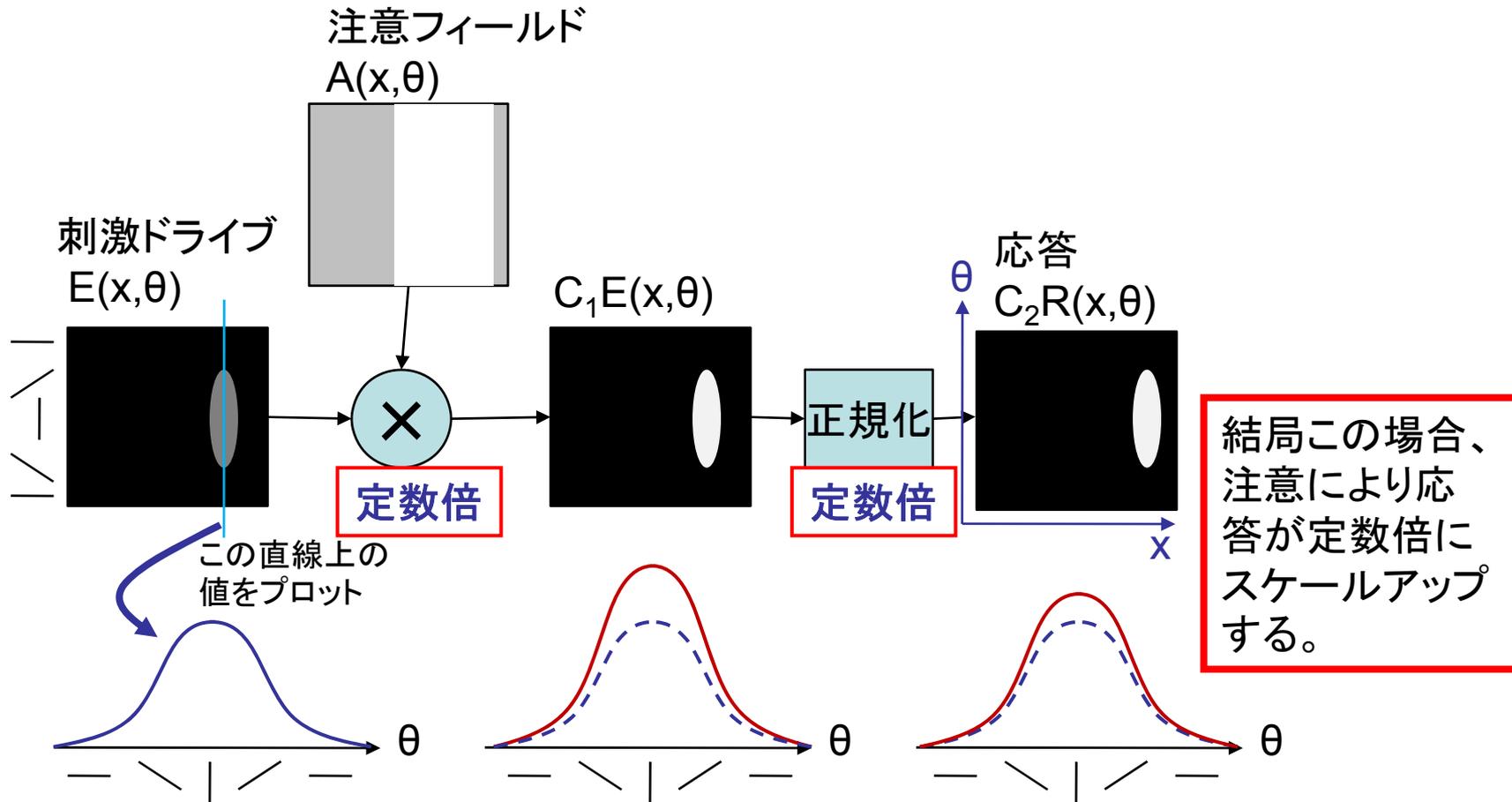
$$\begin{aligned} S(x, \theta) &= \iint A(x', \theta') E(x', \theta') s(x', \theta'; x, \theta) dx' d\theta' \\ &= \iint A(x') E(x', \theta') s(x'; x) dx' d\theta' \\ &= C(x) \end{aligned}$$

ただし、 $C(x)$ は θ に依存しない関数。 $A(x)$ には依存。

$$\begin{aligned} R(x, \theta) &= | A(x, \theta) E(x, \theta) / (S(x, \theta) + \sigma) |_T \\ &\approx A(x, \theta) E(x, \theta) / (S(x, \theta) + \sigma) \\ &= A(x) E(x, \theta) / (C(x) + \sigma) \\ &= \underline{\underline{(A(x) / (C(x) + \sigma)) E(x, \theta)}} \end{aligned}$$

$A(x)$ には依存するが θ には依存しない値

Activity map を使った説明

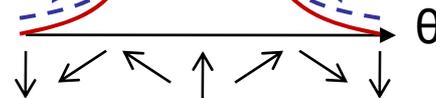


この曲線は、1個のニューロンの tuning curve と一致する点に注意。(理由はわかりにくいけど。)

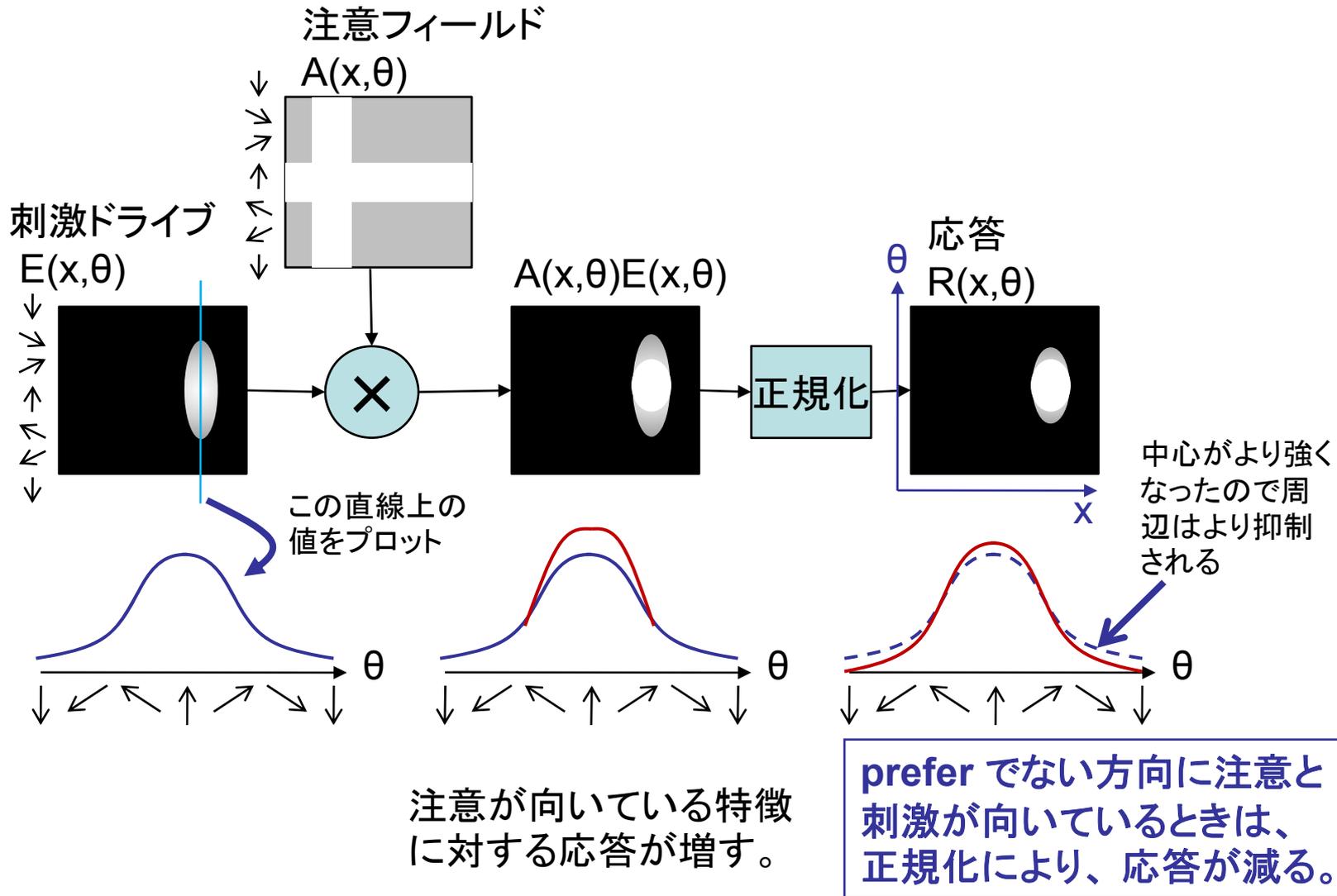
特徴ベースの注意はチューニングカーブを鋭くする [Martinez-Trujillo and Treue 2004]

- prefer な方位刺激の時、その方位刺激に対する注意があると応答はより増える。
- prefer でない方位刺激の時、その方位刺激への注意があると応答はより減る。

Figure 6.

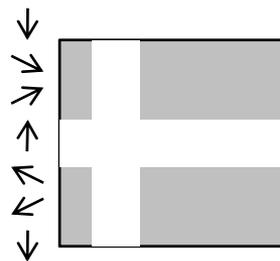


Activity map を使った説明

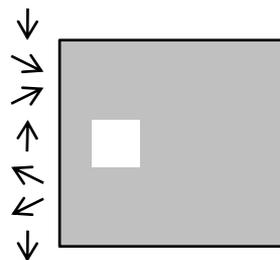


注意点

- 実験では視野の左位置の刺激の方向に注意を向けるようサルに仕向けている。しかしシミュレーションでは、下図のような注意フィールドを仮定している。

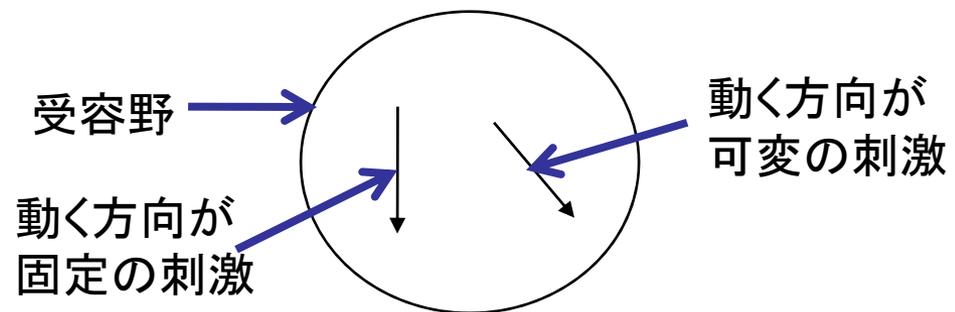


- 下図のような注意フィールドでは実験結果を再現できない。



受容野内に2つの刺激がある場合 (Fig. 7)

- 実験結果とシミュレーションとは異なっている。
 - 著者らは本文中で「表面上の矛盾」の理由の推測を書いているが、意味がよくわからない。
 - (原論文^[Treue and Martinez-Trujillo, 1999]を見てみると、2つの刺激の位置が違う。サルが特徴ベースの注意ではなく空間的注意を向けていたとすれば、実験結果は素直に説明がつくのでは?)



正規化の意味は？

- 正規化の合理的理由については、理論家によってすでにいくつか提案されている。
 - ニューロン間の相対的な発火率に意味がある？
その場合正規化すれば必要なダイナミックレンジを減らせる。 (Heeger, 1992b)
 - 表現の効率を上げる。 (Schwartz and Simoncelli, 2001, Simoncelli and Olshausen, 2001, Wainwright et al., 2002, Wainwright and Simoncelli, 2000),
 - 確率分布の正規化と同じで、ポピュレーションから意味を読みだすのを容易にする。 (Heeger and Simoncelli, 1993, Simoncelli, 2003, Simoncelli and Heeger, 1998)
 - 1つ以上の刺激の次元に対して応答を不変にする。 (Heeger et al., 1996, Kouh and Poggio, 2008, Simoncelli and Heeger, 1998).

感想

- モデルはシンプルだが、適用範囲が広い。
- 「注意フィールド」の値は直接観測されたものではなく、著者らによる「想像」。しかし今後実験によって存在が証明されるだろう。
- 他にもパラメタがけっこうあるが、シミュレーションで使われているパラメタ値はまあ妥当。
- 予言能力が高く、今後様々な実験で検証されていくだろう。

Q&A

- このモデルはパラメタがずいぶん多いのでは？
 - 刺激フィールド、抑制フィールド、注意フィールドの設定値はまあ妥当だと思う。実験でも決められる。
- σ って何？
 - 「コントラスト c が小さい時、応答は c に比例する」、という実験事実を説明するために必要。
- そもそも固視点以外の場所に注意を向ける機能を生物が持つ意味は？
 - 後述。

正規化モデルと BESOM モデルの関係

BESOM [Ichisugi 2007]との共通点1

- コラムごとにトップダウンの信号とボトムアップの信号が掛け算され、ハイパーコラム内で正規化されて(5層から)出力される。

$$l_{XY}^{t+1} = z_Y^t + W_{XY} o_Y^t$$

$$o_X^{t+1} = \prod_{Y \in \text{children}(X)} l_{XY}^{t+1}$$

$$k_{UX}^{t+1} = W_{UX}^T b_U^t$$

$$p_X^{t+1} = \sum_{U \in \text{parents}(X)} k_{UX}^{t+1}$$

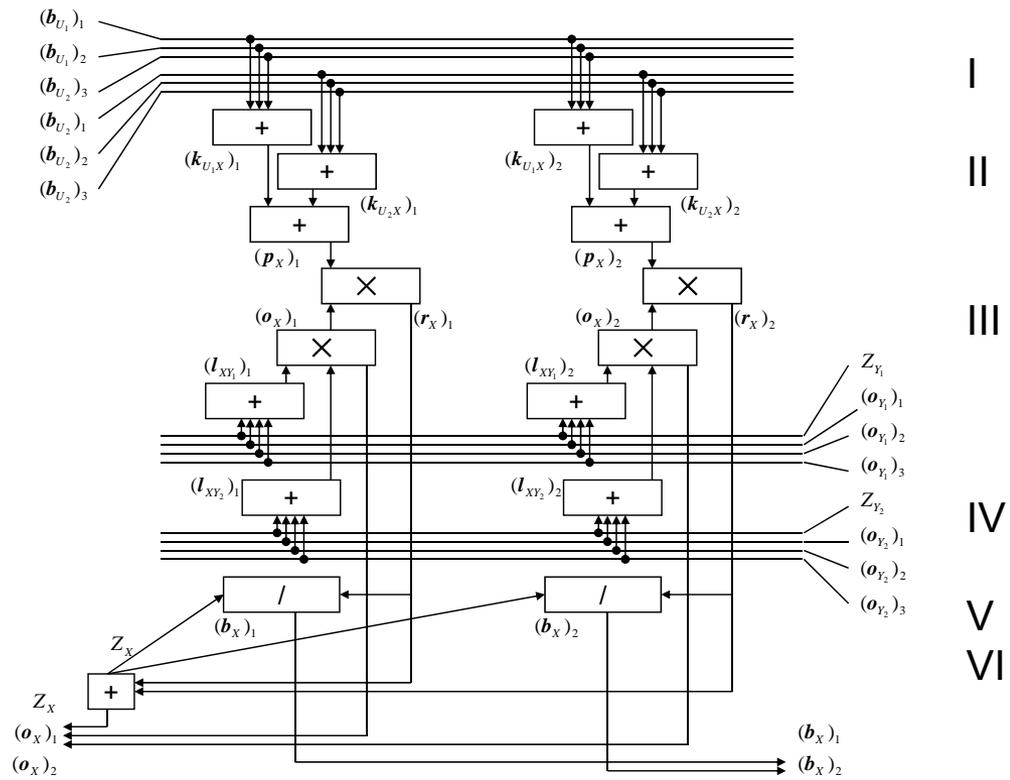
$$r_X^{t+1} = o_X^{t+1} \otimes p_X^{t+1}$$

$$Z_X^{t+1} = \sum_i (r_X^{t+1})_i \quad (= \|r_X^{t+1}\|_1 = o_X^{t+1} \cdot p_X^{t+1})$$

$$z_X^{t+1} = (Z_X^{t+1}, Z_X^{t+1}, \dots, Z_X^{t+1})^T$$

$$b_X^{t+1} = (1/Z_X^{t+1}) r_X^{t+1}$$

ただし、 $\mathbf{x} \otimes \mathbf{y} = (x_1 y_1, x_2 y_2, \dots, x_n y_n)^T$



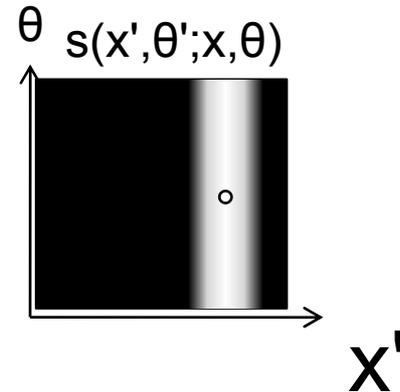
2つのモデルの対応

正規化モデル	意味	BESOMモデル	意味
$l(x, \theta)$	入力刺激	o_Y	ボトムアップの信号
$e(x', \theta'; x, \theta)$	刺激フィールド	W_{XY}	ユニット間の結合行列
$E(x, \theta)$	刺激ドライブ	o_X	観測
$A(x, \theta)$	注意フィールド	p_X	予測
$A(x, \theta)E(x, \theta)$	刺激と注意の積	r_X	観測と予測の積
$s(x', \theta'; x, \theta)$	抑制フィールド	固定	-
$S(x, \theta)$	抑制ドライブ	Z_X	正規化定数
$R(x, \theta)$	ニューロン応答	b_X	出力(事後確率)
x	受容野の中心座標	X	ノードの名前
θ	preferred dir.	i	ベクトルのインデックス

2つのモデルが行う計算は、かなり精密に一致している。

BESOM [Ichisugi 2007]との共通点2

- 特徴に対し対称に正規化。
 - 「注意の正規化モデル」では全シミュレーションで、抑制フィールドの値は方位によらないものを使用。



- BESOM モデルでは、ノード内の各ユニット(=特徴)の出力の総和が1になるように正規化。
 - 3層バスケット細胞の側抑制の幅がハイパーコラム程度、という解剖学的知見とも一致。

2つのモデルの構成要素の起源

構成要素	正規化モデル	BESOM モデル
トップダウンの信号とボトムアップの信号の掛け算	電気生理の実験事実	ベイズの定理 $P(x y) \propto P(y x)P(x)$
正規化	電気生理の実験事実	計算の効率化の要請を満たすために導入 $\sum \pi_{Y_i}(x) = 1$
特徴に対し対称に正規化	電気生理の実験事実 (ハイパーコラムサイズの側抑制の存在も意識?)	ハイパーコラムとコラムをノードとユニットに対応付けた結果

正規化モデルの中核的要素は、BESOM モデルとまったく異なる起源を持つのに、結果的に一致している！

BESOM との違い

- コントラストの扱い

- BESOM の発火率は確率分布。刺激の強さではない。
 - 自然画像をBESOMで非線形スパース符号化すると、コントラストの低い刺激はどういう表現になるのか、現時点では不明。
 - 正規化モデルと同じ説明能力があるだろうか？

- 近傍の空間で正規化。

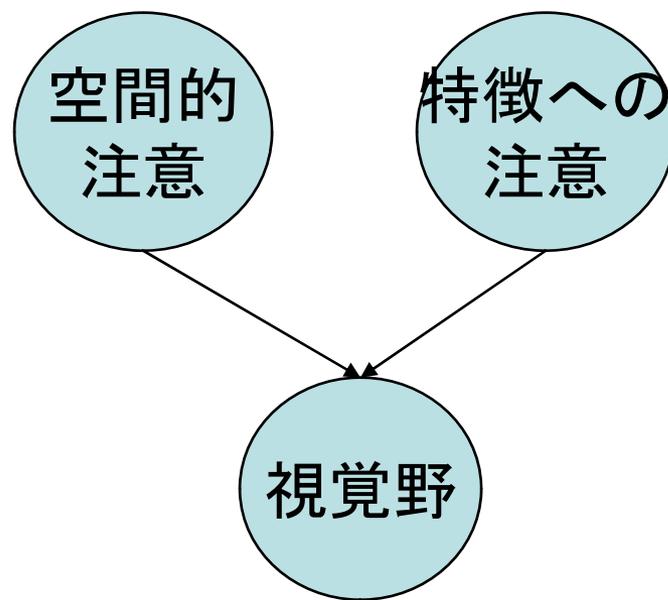
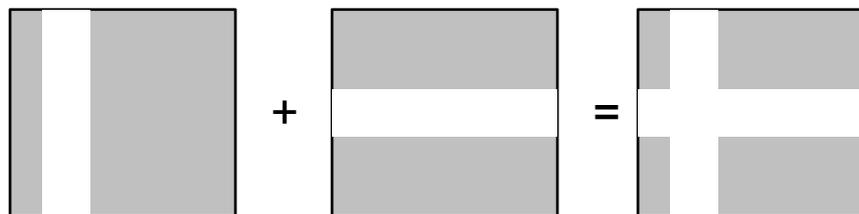
- 「注意の正規化モデル」では近傍の受容野を持つニューロンからも抑制を受ける。
 - 周辺抑制を説明できる。
 - でも、周辺抑制を説明するモデルは他にもある。
- BESOM モデルではノード間の競合(未発表)に相当？

BESOM モデルにしかない点

- BESOM モデルははるかに広範囲の事象を説明するモデル。
 - 学習。
 - 多層化。
 - ベイズに基づく理論的基盤。
 - コラム構造・6層構造と対応する詳細な神経回路モデル。
 - パターン認識、時系列学習、運動制御などの情報処理の機能。
 - etc.

十字型の注意フィールドになる理由

- BESOM モデルで十字型になる理由が説明できる。
 - BESOM モデルでは、トップダウン信号は複数の親ノード（独立成分）からの信号の線形和。
 - サルが空間的注意と特徴ベースの注意のそれぞれを独立成分として学習済みだとしたら、十字型の注意フィールドしか作れない。



注意の機能的意味は？

- そもそも視線とは違う場所に注意を向けることは日常生活であまりない。そのような機能を脳が持っている意味は？
 - BESOM モデルによれば、トップダウンの信号の意味は文脈からの予測。「こういう刺激が来るはずだ」という予測情報とベイズの定理を使うことでロバストな認識ができる。
 - したがって、固視点以外の場所に「注意」を向けられるのは、「予測機構」の副産物であって、何か意味があってそのような機能が作り込まれているわけではない、というのが BESOM モデルによる答え。

BESOM との比較の今後

- 全く違う方向から導かれた2つのモデルが強い共通点を持つのは、**2つが正しいモデルである証拠**。
- BESOM で「注意の正規化モデル」と同じシミュレーションをやってみるべき。
- BESOM と正規化モデルで異なる予言をする実験をデザインして、実験で検証できるとよいのだが。コントラスト関係？

最後に、BESOM モデルの現状報告

- 過適合を避けるため4種類の正則化(≠正規化)の機構を盛り込んだ、理論的に妥当な学習則を設計し、基本的な動作を確認。
- ノード数 n に対して計算量・メモリ量 $O(n)$ で動作するアルゴリズムが完成の見込み。
 - 現状の実装は計算量 $O(2^n)$ 。
- 2つの1次元確率分布SOMによる非線形IC Aの動作を確認。

2次元のデータから49次元の特徴ベクトルを作って学習させた例。

黒の格子はMPEの受容野重心をつないだもの。

赤と青はユニットの受容野重心をつないだもの。



参考

- 一杉裕志、「脳の情報処理原理の解明状況」
産業技術総合研究所テクニカルレポート
AIST07-J00012, Mar 2008.
<http://staff.aist.go.jp/y-ichisugi/besom/AIST07-J00012.pdf>
- Yuuji ICHISUGI, "The cerebral cortex model that self-organizes conditional probability tables and executes belief propagation",
In proc. of International Joint Conference on Neural Networks ([IJCNN2007](#)), Aug 2007.
<http://staff.aist.go.jp/y-ichisugi/besom/20070509ijcnn-paper.pdf>