皮質における物体認識の階層モデル

Hierarchical models of object recognition in cortex

Maximilian Riesenhuber and Tomaso Poggio (1999)

大脳皮質における視覚処理は，Hubel&Wiesel の単純な細胞から複雑な細胞へのモデルを自然に拡張して，ますます洗練された表現の階層として古典的にモデル化される。 しかし驚くべきことに，物体認識などの高次視覚処理を説明するために，このクラスのモデルが生物学的に実現可能かどうかを検討するための定量的なモデル化はほとんど行われていない。 我々は，この複雑な視覚課題を説明し，検証可能な予測を行う，下頭側頭皮質の生理学的データと一致する新しい階層的モデルを説明する。 このモデルは，特定の皮質ニューロンへの入力に適用される 最大値をとるような演算に基づいており，皮質の機能において一般的な役割を果たしていると考えられる。

視覚的対象物の認識は，不変性と特異性という 2 つの必須要件を持つ，頻繁に行われる基本的な認知課題である。 例えば，我々は，視点やスケール，照明や表情が変わっても，多くの顔の中から特定の顔を認識することができる。 脳は，このような物体の認識や検出課題を，素早く (1) かつ，うまく実行する。 だが，どうやって？

マカクサルの内側側頭葉皮質 (IT) (2) にある細胞は，物体認識に重要な役割を果たしていると考えられており (3)，顔のような複雑な物体の見え方に同調する。 これらの細胞は，顔に対して強く放電するが，他の物体に対してはほとんど放電しないか， 全く放電しない。 これら細胞の特徴は，スケールや位置の変更などの刺激変換に対する反応が頑強であることである。 この発見は，なぜこれらの細胞が，網膜の光受容体を同様に活性化する類似の刺激 (例えば 2 つの異なる顔) には異なる反応を示すのに，網膜上でまったく異なる活性化パターンを生じる好ましい刺激の拡大・縮小 や 変形版には一貫して反応するのか，という興味深い問題を提起している。

このパズルは，ネコの縞模様の皮質で記録された単純細胞と複合細胞が，はるかに小さなスケールで示したものと似ている (4)。 しかし，単純細胞が小さな受容野で強い位相依存性を持つのに対し，複雑細胞は大きな受容野で位相依存性を持たない。 そこで Hubel&Wiesel は，受容野が隣り合った単純細胞が同じ複合細胞に入力することで，その複合細胞に位相不変の応答を与えるというモデルを提案した。 この方式を単純に (しかし非常に理想的に) 拡張すると，単純な細胞から「高次の超複雑な細胞」へとつながる(5)。

並進不変の物体認識のための Neocognitron(6) を皮切りに，複雑な物体に同調する変形不変の細胞が単純な細胞入力から生じることを説明するため，視覚系における形状処理のいくつかの階層的モデルが提案されてきた (7, 8)。 しかし，これらのモデルは，定量的に規定されていなかったり，特定の実験データとの比較がなされていなかったりする。 並進不変やスケール不変の物体認識の代替モデルは，「シフター」回路 (9) やその拡張版 (10) のように，入力信号を適切に迂回させる制御信号に基づいているか，あるいは不変認識の「ゲインフィールド」モデルのように，ニューロンの応答を調節する信号に基づいている (11,12)。 マカクサルの視覚野 V4 の細胞は，注意によって受容野の空間的な移動や変調を示すことができるが (13, 14)，この機構が変換不変の物体認識に使われていることや，同様の機構が他の変換 (スケーリングなど) にも適用されることを示す証拠はまだほとんどない。

Perrett&Oram (7) が描いた階層モデルの基本的な考え方は，あらゆる変換 (Neocognitron (6) の場合のような画像平面の変換だけでなく) に対する不変性は，同じ刺激の様々な変換バージョンに同調した求心性をプールすることで構築できるというものであった。 実際，このようなプーリングメカニズムを用いて，視点不変の物体認識が可能であることが以前に示されている (15)。 学習ネットワーク (ガウス型 RBF) は，複雑な紙芝居のような物体を 3 次元空間で 1 軸周りに回転させた個々のビューを用いて学習され，奥行き方向に回転させてもこの物体を不変に認識できるようになっている。 このネットワークでは，視点調整されたユニットは， 視点不変ユニットに供給され， これらのユニットは， 学習ネットワークが視点不変を達成するために補間するプロトタイプを効果的に表している。

現在，定量的な心理物理学的証拠 (16-18) と生理学的証拠 (19-21) が得られており，全景または部分景に同調するユニットは，おそらく学習処理過程によって作られ，視野不変の出力は少数の個々のニューロンによって明示的に表現されているの

ではないかという仮説が提唱されている (19,21,22)。 紙ハサミに似た見慣れない標的刺激の限られた視野で訓練を受け，その後，多数の類似した「妨害」物体の視野中で，奥行き方向に回転した「目標」の新しい視野を認識することを要求されたサルでは，前部 IT ニューロンは，訓練中に見られた物体の視野に選択的に反応する (17,21)。 この研究では，これまでの研究で問題となっていた 2 つの問題を解決した。 まず，サルがよく知っている物体 (顔など) ではなく，新規の刺激を認識するように訓練することで，物体の 1 つの見えから得られる視野不変性の度合いを推定することができる。 さらに，多数の妨害物体を用いることで，妨害物体に関する視野不変性を定義することができる。

これは重要な点で，VTU (視点調合されたユニット) の不変範囲は，ニューロンの選好刺激の変換版に対する反応と，一連の (類似した) 妨害物体に対する反応を比較することによってのみ決定することができまる，単に調合曲線を測定するだけでは十分ではない。

これは，1 つの物体視点で訓練後，訓練視点の周りの 3 次元回転に対して限定的な不変性を示す細胞であり (図 1)(21)，視点補間モデル (15) と一致する。 さらに，この細胞は，物体が以前に 1 つの縮尺と位置で提示されていたにもかかわらず，並進と縮尺の変化に対しても不変であることがある。

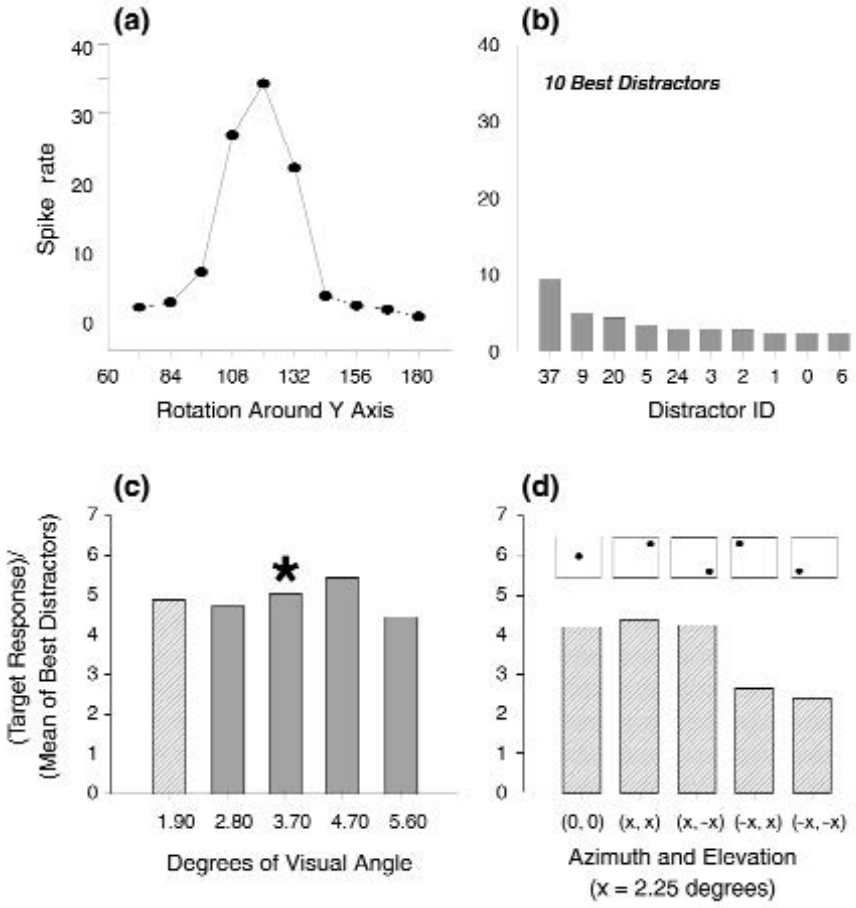

図 1. 1 つのニューロンの不変性特性 (Logothetis+ より改変(21))。
この図は， サルに紙クリップのような物体を認識するように訓練後， 前部 IT に見られる 1 つの細胞の反応を示している。 この細胞は紙クリップの 1 つの見えに選択的に反応し， 訓練中にサルが 1 つの位置と縮尺でしか刺激を見ていなかったにもかかわらず，奥行きの回転に対する訓練ビュー周辺での限定的な不変性を示し，平行移動と大きさの変化に対しても有意な不変性を示した。
(a) 優先的見えを中心とした奥行き方向の回転に対するセルの応答。
(b) 最も強い反応を誘発した 10 個の妨害物体 (他の紙クリップ) に対するセルの反応。
下のプロット (c, d) は，刺激の大きさ (アスタリスクは訓練中の見えの大きさを示す) と 位置 (1.9° の大きさを使用) の変化に対する細胞の反応を， 10 個の最適な妨害刺激の平均値と比較したもの。 「不変性」とは， 好ましい刺激の変換された見えに対して， 妨害物体よりも高い反応を示すことと定義し，ニューロンは平均 42° の回転不変性を示した (訓練中，サルに完全な 3D 情報を提供するために，刺激は実際に奥行き方向に ±15°回転させられていた。 また，並進不変量と尺度不変量は，それぞれ訓練時の視野に対して ±2° と ±1 オクターブ の オーダーであった (J. Pauls, 私信)。

これらのデータ，視野調整された細胞の特性の基礎となる回路の問題を，定量的な観点から明確に示している。 当初のモデルでは，VTU を使って視野不変ユニットを構築する方法が説明されていたが(15)，視野同調ユニットがどのようにして生じるのかは明示されていなかった。 したがって，重要な問題は，VTU が 1 つの物体の見えから得られる並進および縮尺に対して不変であることを，生物学的に妥当な機構で説明することである。 この不変性は，特定の物体に対する選択性と，位置や縮尺の変化に対する相対的な耐性 (発火の頑健性) との間のトレードオフに相当する。 本研究では，解剖学的および生理学的な制約に準拠したモデルを作成し，上述の不変性データを再現し，IT 細胞の見えに合調された亜集団の実験を予測した。 興味深いことに，このモデルは，文脈の中での認識 (23) や，細胞の受容野に複数の物体が存在する場合 (24) の実験データとも一致した。

結果

このモデルは，単純な階層型のフィードフォワードアーキテクチャに基づいている (図 2)。 その構造は，位置や縮尺に対する不変性と，特徴の特異性が別々の機構で構築されなければならないという仮定を反映したものである。 より単純な特徴を符号化する求心性神経の加重和，すなわちテンプレートマッチは， 特徴の複雑さを増すのに適した神経伝達関数である。 しかし，異なる重みの求心性の和をとることで，不変性も高まるのだろうか？

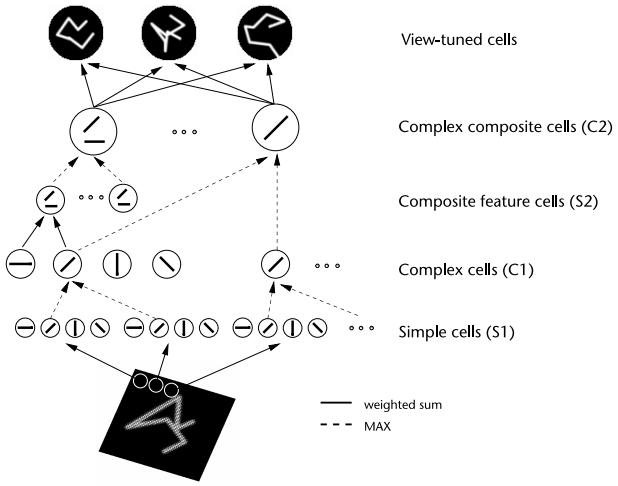

図 2. モデルのスケッチ。 このモデルは，単純細胞から作られた複雑細胞の古典的なモデル (4) を拡張したもので，線形演算 (Fukushima (6) の表記法では S ユニット，テンプレートマッチングを行う，実線) と非線形演算 (C プーリングユニット (6)，MAX 演算，破線) を持つ層の階層で構成されている。 この非線形の MAX 演算は，セルの入力の最大値を選択し，それを用いてセルを駆動するもので，複雑細胞に通常想定される基本的に線形の入力の合計とは異なり，モデルの特性を決定する鍵となっている。 この 2 種類の操作により，パターンの特異性と，異なる位置に合調された求心性をプールすることによる並進に対する不変性，および異なる縮尺に合調された求心性をプールすることによる縮尺に対する不変性 (図示せず) が得られた。

計算上の観点からは，プーリング機構は，頑健な特徴検出器を生成する必要がある。 つまり，受容野のクラッターや文脈に惑わされることなく，特定の特徴を検出できるようにする必要がある。 一次視覚野に見られる，ある方向の棒に位相不変で優先的に反応する複雑細胞を考えてみよう (4)。 オリジナルの複合細胞モデル (4) によると，複雑細胞は，異なる位置にある単純細胞の配列からの入力をプールして，位置不変の応答を生成していると考えられる。

理想化されたプーリング機構には，等方的な応答を実現するために重みを等しくした線形加算 (SUM) と，最も強い求心性がシナプス後の応答を決定する非線形最大演算 (MAX) の 2 つの選択肢がある。 いずれの場合も，モデル複合細胞の受容野内の 1 本のバーに対する応答は，位置不変である。 反応レベルは，刺激が求心性の好ましい特徴に類似していることを示している。 ここで，視覚野に紙クリップのような複雑な刺激がある場合を考えてみよう。 線形和の場合，刺激が細胞の受容野にある限り，複合細胞の応答は不変だが，出力信号がすべての求心性の合計であるため，応答レベルによって，複合細胞の受容野のどこかに好ましい方向の棒が実際にあったかどうかを推測することはできない。 つまり，特徴の特異性が失われてしまう。 しかし，MAX の場合，応答は最も活発な求心性によって決定され，したがって，刺激のどの部分も求心性の好みの

特徴に最もよく一致することを示すことになる。 この理想的な例は，乱雑な場所での認識や，受容野に複数の刺激がある場合に，MAX 機構がより強固な応答を提供することを示唆している (以下参照)。 なお，入力に飽和非線形性を持たせた SUM 応答は，求心性神経の活動レベルに応じてパラメータをケースバイケースで調整する必要があるため「脆い」と思われる。

同様に重要なのは，SUM 機構がサイズ不変性を実現できないことである。 例えば，ある「複雑」細胞 (V4 や IT にある細胞) への求心性が，ある程度の大きさや位置の不変性を示しているとする。 この「複雑」細胞を同じ物体で刺激し，そのサイズを徐々に大きくしていくと，より多くの求心性神経が刺激によって興奮することになる (求心性神経が空間や縮尺において重複していない場合)。 その結果，求心性神経がサイズ不変であるにもかかわらず，「複雑」細胞の興奮は刺激の大きさとともに増加することになる。 (このことは，単純化した 2 層モデルを用いたシミュレーションでも証明されている (25))。 しかし，MAX 機構では，細胞の反応は，刺激の大きさが大きくなっても，ほとんど変化しない。

これらの考察 (後述するモデルの定量的なシミュレーションによる裏付け) は，非線形 MAX 関数が不変性を達成するために応答をプールするための賢明な方法であることを示唆している。 これは，応答が不変であるべき変換のパラメータ (例えば，縮尺不変の場合は特徴量) が異なる同じタイプの求心性を暗黙のうちにスキャンし (考察を参照)，最適にマッチした求心性を選択するというものである。 これらの考慮事項は，プーリングセルの異なる求心性 (例えば，空間の異なる部分を見ているもの) が，視野内の異なる物体 (または同じ物体の異なる部分) に反応する可能性がある場合に適用されることに注意。 (これは，低次視覚野の細胞が幅広い形状のチューニングを持っていることと同様である) ここで，求心性神経を組み合わせてプーリングすると，異なる刺激による信号が混ざってしまう。 しかし，モデルの最終段階で予想されるように，求心性神経が 1 つのパターンにしか反応しないような特異性を持っている場合は，RBF ネットワーク (15) のように，異なる視点に同調した VTU を組み合わせて，保存されている見えを補間するように，加重和を用いてプールするのが有利である。

回路のいくつかの段階における MAX のような機構は，神経生理学的なデータと一致している。 例えば，ある IT ニューロンの受容野に 2 つの刺激を提示すると，そのニューロンの反応は，そのニューロンに単独で提示された場合に高い発火率をもたらす刺激に支配されるようだ (24)。 これは，このニューロンやその求心性のレベルで MAX 的な操作が行われた場合に予想されることである。 V1 複雑細胞のプーリング機構の可能性についての理論的な研究でも，MAX のようなプーリング機構が支持されている (Sakai&Tanaka, Soc. Neurosci. Abstr. 23, 453, 1997)。

MAX 機構を間接的に裏付けるものとして，「単純化手順」(26) や「複雑性の低減」(27) を用いて，IT 細胞の好ましい特徴，すなわち，細胞を駆動する原因となる刺激成分を決定する研究が挙げられる。 これらの研究では，IT 細胞の高度に非線形なチューニングが一般的に見られる (図 3a)。 このようなチューニングは，MAX 応答関数と一致する (図 3b 黒棒)。 なお，線形モデル (図 3b 灰色棒) では，入力画像のわずかな変化に対するこの強い応答の変化を再現できなかった。

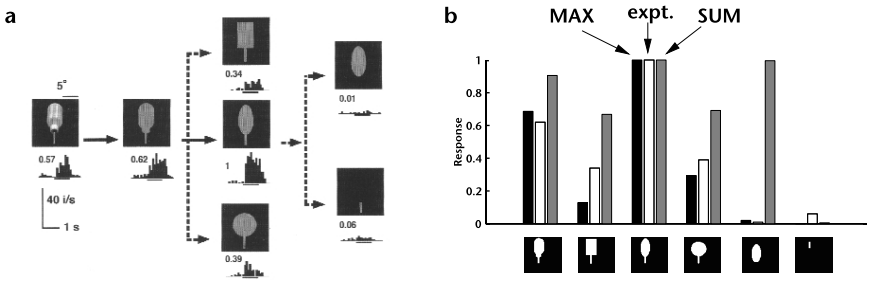

図 3. MAX 機序の高度に非線形な形状調整特性。
(a) 「最適な」特徴を決定するために考案された「単純化手順」(26)を用いて得られた，IT 細胞の実験的に観察された反応 (選好刺激に対する反応が 1 になるように正規化された反応)。 この実験では，IT 細胞はもともと「水のボトル」 (左端の物体) の画像に非常に強い反応を示していた。 次に，この刺激を単色の輪郭に単純化すると，細胞の発火が増加し，さらに，楕円を支える棒からなるパドルのような物体に変更した。 この物体が強い反応を引き起こすのに対し，棒や楕円だけではほとんど反応しなかった （図は許可を得て使用)。
(b) 実験とモデルの比較。 白い棒は (a) の実験用ニューロンの反応を示す。 黒と灰色の棒は，選好刺激の幹ー楕円の基部の遷移に同調したモデルニューロンの反応を示している。 モデルニューロンは，図 2 に示したモデルを簡略化したもので，受容野の各位置には，それぞれ遷移領域の左側または右側に同調した 2 種類の S1 特徴のみが存在し，それらを MAX 関数 (黒棒) または SUM 関数 (灰色棒) を用いてプールする C1 ユニットに供給されている。 モデルニューロンは，実験ニューロンが好む刺激が受容野にあるときに反応が最大になるように，これらの C1 ユニットに接続されていた。

我々が開発した見え合調ユニットモデル (図 2) では，スキャンとテンプレートマッチングという 2 種類の操作を階層的に組み合わせて，モデル「網膜」からの入力を受ける最下層の小さな局所的な単純細胞のような受容野から，複雑で不変的な特徴検出器を構築した。 モデルの「網膜」からの入力を受ける最下層の小さな局所的な単純細胞のような受容野から，複雑で不変的な特徴検出器を構築する。 この 2 つの操作は厳密に交互に行う必要はない。 図 2 のモデルの C1-C2 の直接接続のように，接続が階層のレベルを飛び越えても構わない。

問題は，提案モデルが実際に生理学からの結果と互換性のある応答選択性と不変性を達成できるかどうかであった。 この疑問を解決するため，実験で使用されたように，ランダムに選択された異なる紙クリップの見え方にそれぞれが同調するモデルの 21 個のユニットの不変性を調べ(21)。

図 4 は，モデルビューにチューニングされた 1 つのユニットの，選好の見えを中心とした 3 次元の回転，拡大縮小，移動に対する反応を示したものである (方法を参照)。 このユニットは，訓練中の見えに対して最大の応答を示し，刺激が訓練中の見えから離れて変換されると， 応答は徐々に落ちていった。 実験と同様に，選好刺激に対する応答と 60 個の妨害刺激に対する応答を比較することで，VTU の不変範囲を決定することができる。 VTU の不変範囲は，モデル VTU の反応がどの妨害刺激に対するものよりも大きくなる範囲と定義される。 その結果，モデル VTU は，回転不変 24°，縮尺不変 2.6 オクターブ，並進不変 4.7° の視角を示した (図 4)。 21 個のユニットを平均すると，平均回転不変度は 30.9° 以上，縮尺不変度は 2.1 オクターブ以上，並進不変度は 4.6°以上となった。

訓練中の見えの周辺では，ユニットは不変性を示し，その範囲は実験的に観測された値とよく一致した。 また， 一部のユニット (21 個中 5 個，図4d の例) は，実験的に観察されたように，疑似鏡像に対してもチューニングを示した (紙クリップの自己閉塞感が少ないためで，優先する紙クリップを奥行き方向に 180° 回転させることで得られる)(21)。

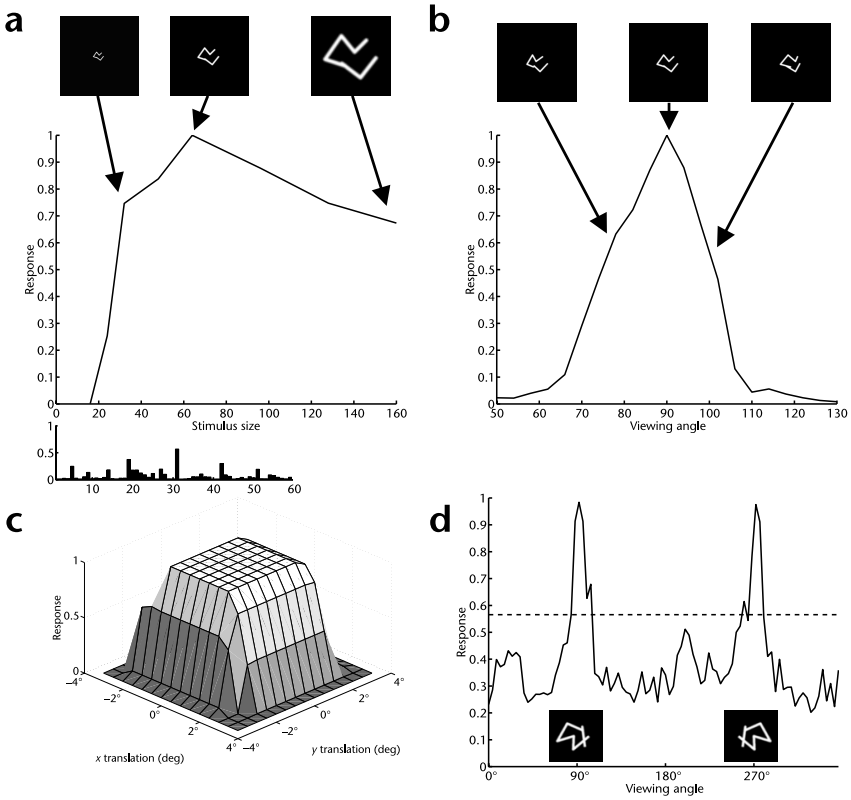

図 4. 好みの刺激をさまざまに変化させたときのモデルニューロンの反応。 パネルは (a) 刺激の大きさを変えたときの同じニューロンの反応を示している (挿入図は， 生理学実験で使用したクリップからランダムに選んだ 60 個の気晴らし物体に対する反応を示している(21) (b) 奥行き方向の回転。 (c) 平行移動。 訓練刺激の大きさは 64×64 画素ルで， 視角 2°に 相当する。 (d) 擬似鏡像に対する別のニューロンの反応 (本文参照。 破線は「最適な」妨害刺激に対するニューロンの反応を示す。

ソフトマックス近似 このモデルの簡略化された 2 層バージョン (25) では， MAX 演算の近似値が認識性能に与える影響を調べた。 このモデルには， プーリングステージ C1 のみが含まれており， プーリングの非線形性の強さを パラメータ p で制御することができた。 ここでは，求心性 s_j を持つ C1 セルの出力 c_i^1 は:

$$C_i^1 = \sum_j \frac{\exp(p \cdot |s_j|)}{\sum_k \exp(p \cdot |s_k|)} s_j$$

ここで，$p = 0$ では線形加算 (求心性の数でスケールアップ)，$p \to \infty$ では MAX 演算を行う。