

奥行手がかりの統合

テクスチャと視差による奥行知覚

脳が曖昧な視覚入力から三次元 (3D) 情報を導き出すかをモデル化するために、ベイズ推論理論が広く用いられてきた。特に最尤推定 (MLE) モデル (MJE) は、複数の奥行き手がかりからの推定値の相対的信頼性から最も可能性の高い 3D 解釈を生成する。しかしここでは、別の考え方 intrinsic constraint 理論、すなわち視覚系は視覚入力の最も安定した解釈を導き出すが、最も確率的な解釈は導き出さないと仮定している。

3D 視覚では次の 2 つの仮説が研究を導いてきた：(1) 個々の独立モジュールが平均的でノイズの多い推定値を導き出す、(2) 視覚メカニズムは個々のモジュールは確率分布をあらゆる感覚レベルのノイズを出力。確率分布をあきらかにすることでベイズ統合の枠組みで提案された独立した推定値の最適な組み合わせを可能にする (Landy et al. 2011)。もっと一般的なベイズの組み合わせの実装があるが、線形最尤推定 (MLE) モデルにしたがったこれまでの研究である両眼視差とテクスチャ、あるいは両方で行われた対象を観察する時の無視できるほどの影響を事前に仮定していた (Chen & Saunders, 2020)。

両眼視差とテクスチャを統合した情報の線形 MLE モデルの予測は以下の 2 つの公式にまとめられる。第一に、両眼視差とテクスチャのモジュールからの奥行評価値である \hat{z}_t と \hat{z}_d およびこれらの要素のノイズの標準偏差である σ_t と σ_d とすれば、統合した評価値 \hat{z}^c は評価値の重みづけに比例した信頼度 $r_i = 1/\sigma_i^2$ となる。

$$\hat{z}^c = w_t \hat{z}_t + w_d \hat{z}_d \quad (1)$$

ここで、 $w_t = r_t / (r_t + r_d)$ 、 $w_d = r_d / (r_t + r_d)$

第二に、知覚を弁別より他の方法 (独立したプローブなど) で測定したら、統合した推定値の分散は以下の関係に示すように単一の手がかりの推定値はどちらよりも小さい。

MLE モデルを知覚過程の説明にあて、直ぐにはその中心となる仮定は人間の知覚システムの説明を果たさない。第 1 に、これまでの実験は奥行手がかりが一般的に正確な知覚を生み出すのに失敗し、それどころか真実な仮定に反していた (Campagnoli et al. 2017、他)。第 2 に、奥行知覚の弁別法以外の測定法で測定する場合、知覚された奥行の測定のバラツキは奥行手がかりの相対的重みづけを予測していない、どころか奥行手がかりの評定は確率分布として表現されるという仮定に反すると考えられる (Todd et al. 2010)。これらのことを考えると、MLE モデルを広範囲に人間の知覚に適応することは適切ではないこと、さらに奥行手がかりの統合する実験は、別の説明で再検討する必要がある。

そのため 3D 手がかりの統合の理論的枠組みを提起し、上述した主流となっている MLE の仮定を必要としないことを考え、その代わりに 3D の特性を引き出す枠組みを以下のように仮定した。すなわち (1) 一般的には観察条件にはバイアスがあるが、しかし一定の観察条件では真実である可能性がある、(2) 確率的ではなく決定論的である。したがって、手がかり推定内におけるバラツキはごくわずかであるため、複数の信号の組み合わせルールに考慮されない。その代わりに、このプロセスは観察条件および表面の特性 (表面テクスチャ

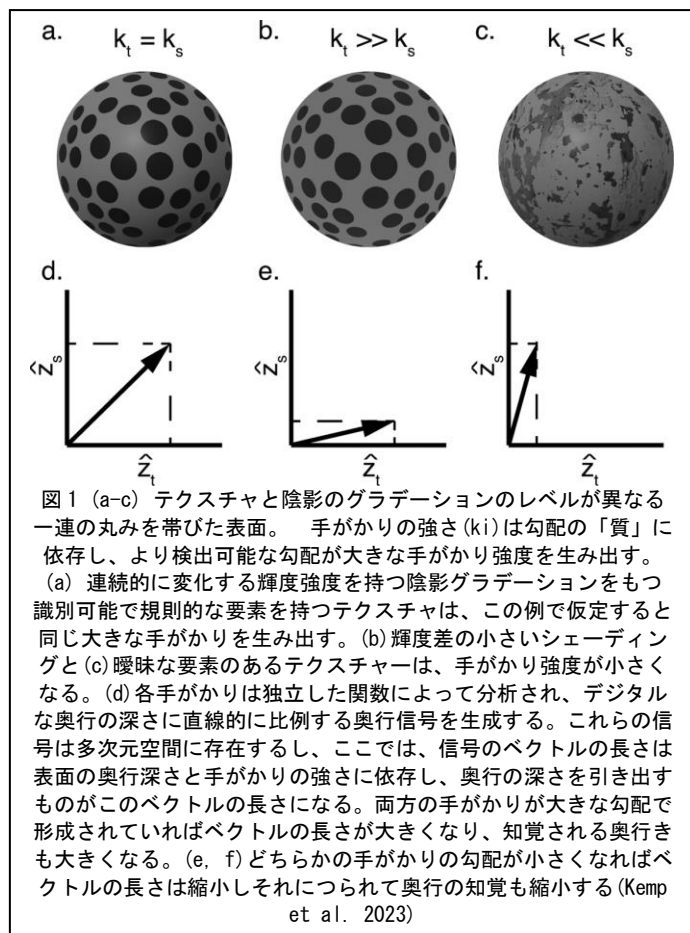
ア、照明など)の自然な変化に対して知覚的安定性を達成するために最適化されている。

Kemp et al. (2023)提案する規範的モデルは、内在的制約 (IC: intrinsic constraint theory) 理論と呼ばれ、IC 理論とは全く異なる仮定をもつ。第 1 に、別々の視覚モジュールがそれぞれ独立にイメージの規則性を処理すると仮定し、それら別々のモジュールがディスタル (distal) の 3D の特性および内的 3D の評価値は、 $\hat{z}_i = k_i z_i$ となる \hat{z}_i である。勾配は線形をとる k_i で奥行きがかり内の i の視覚情報の質に依存する。例えば、図 1 の 1a から 1c のイメージからモジュールは評価値 ($\hat{z}_t = k_t z_t$) に導かれる形と空間周波数テクスチャの要素に体系的変化を出す。とくに、MLE モデルに対して知覚機能は本当の、あるいはなんら関係する明確な感覚ノイズは仮定してはいない。それゆえに本当らしい仮説を ICE は MLE のモデルから取り除く。IC 理論は勾配 k 、これは知覚関数から独立して変わる変数で 3D 形状に無関係な独立する視覚モジュールと定義される。モジュールの出力を奥行評価値と単純に考えるが、形式的にはそれらはベクトルが直交する多次元空間に直交信号を生成するものとしてモデル化され、これらのベクトル信号の大きさは知覚される奥行の大きさを示す。

図 1 のテクスチャ要素から出現するイメージシグナルにそって奥行に関連する独立したイメージを生み出す陰影の勾配 (gradient) ができる。しかし、対象の表面は照明条件が異なるので、独立した厄介な変数は表面特性と照明光源に関連する。同様に、異なる厄介な変数は他のイメージシグナル (運動視差における速度、両眼視差での対象と観察者の間の距離)

にも影響を与える。知覚関数 k の勾配は個々の手がかりからのディスタルな奥行に対する反応を示すので手がかりの強度反応を示すので、それらを手がかり強度とする。独立した手がかり強度は与えられた刺激から得られる手がかり推定値の全体で、それは各手がかりからの情報をあらゆる各直交軸の多次元ベクトルとして所与の刺激から引き出される。

これは図 1d から 1f に示されていて、図 1a から 1c の刺激に含まれるテクスチャーと陰影情報を表す 2D ベクトルの具体例を示している。図 1d は、図 1a の刺激によって伝達される奥行き情報を出している。この場合、テクスチャーとシェーディングは (説明のために) 同じ強さ ($k_t = k_s$) とし、図 1e で



は、シェーディングが除去されているため、シェーディングの強さよりもテクスチャの強さの方がはるかに大きい ($k_t > k_s$)。図 1f では、テクスチャは非常に不規則である ($k_t < k_s$) のでテクスチャの強さよりも陰影の強さの方が大きい。 \hat{z}_t と \hat{z}_s はともに 3 次元特性 z に比例するため、組み合わされたベクトル (\hat{z}_t, \hat{z}_s) の長さも z に比例する。しかし、複合されたベクトルの長さはテクスチャとシェーディング(表面特性と環境照明)に影響する厄介な変数で個々の手がかりの強さに依存するため変動する。この理論の中心的な主張は、視覚システムの目標が 3D 情報に対する感度を最大にする一方で厄介な変数を最小化することである。

実験 1 では、手がかりが複合された時に予測される単一手がかりによる推定値の不正確さと、予想される系統的バイアスを検証した。これらのバイアスはベクトル和のモデルによって、自由パラメータなしで予測することができる。

実験 2 では、識別の閾値が単一手がかり刺激に対して複合手がかり刺激では単一手がかり刺激に比べて弁別閾が低下するといこれまでの知見を繰り返している。ベクトル和のモデルと MLE モデルが、弁別閾値に関して類似した予測を行うが理由は異なり、線形手がかり強度の特性と確率分布の特性に関連する。そこで、まずこれらの閾値は、実験 1 で測定された手がかりの強さによって予測できることを示す。実験 3 では、平らな奥行手がかりの信頼性を高めても知覚された奥行をベクトル和のモデルの予測に従えば縮小されない。その結果、刺激が平板の手がかりでは MLE モデルが実験 1 で説明したバイアスを説明できる可能性は低い。

実験 1 では、被験者に 3 条件 (視差のみ、テクスチャのみ、複合した奥行手がかり) の面の奥行を判断させた。図 2 には、手がかり条件(列)、各シミュレートされた深さの大きさ(行)、40cm の観察距離を示し、最後の列はプローブは正しい振幅に設定されている。テクスチャのみの条件のサーフェスは、ボリユーメトリック・テクスチャリングによって構築、視角 0.55° の半径を持つ球の中心を、シミュレートした 3 次元波形の上に置いた。奥行方向の曖昧さをなくすため、シェーディング情報を生成し、x 軸の周りに 45° の角度をつけた単一指向性光源を表面の上に配置し、さらに実験にある複合手がかり条件と混同を避けるためにテクスチャ手がかり条件のみとした。さらに、他のテクスチャー・パターンと比較して識別課題において最高レベルの精度を引き出す水玉模様のテクスチャを使用することにした。この方法は奥行知覚のために MLE モデルが最も高い信頼性の影響ポテンシャルがあるのか、あるいは他のテク

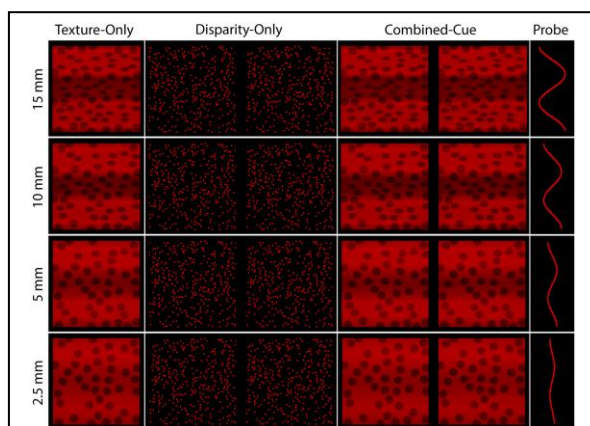


図 2 観察距離 40cm における各奥行での各手がかりタイプによる刺激の例。テクスチャのみの条件、3D 構造は水玉で定義され、奥行の順序のあいまいさをなくすために陰影が加えられた。視差のみで定義された表面はランダムドットステレオグラム(視差交差可能)。複合手がかり条件では、両方の手がかりが存在した(視差交差可能)。参加者は 2D のプローブの振幅を調整し、知覚された奥行を報告(Kemp et al. 2023)

スチャに比べて最大の手がかり強度を持つ（ベクトル和のモデル）を意味する。

実験では、被験者は手がかり条件(視差のみ、テクスチャのみ、複合した奥行き手がかり)の一つから4つの異なるピークからトラフまで(2.5, 5, 10, or 15 mm)の深さが異なる正弦波サーフェスをそれぞれ7回観察させた。被験者はプローブの振幅を調整し、自由に視線を3D表面と2Dプローブの間を行ったり来たりさせ、判断したらボタンを押させた。

実験1の結果、ひとつの同じ刺激から測定された弁別閾値は、複合手がかり刺激の弁別閾値よりも式2で予測されるように、複合手がかり刺激の閾値よりも小さい。ベクトル和のモデルもまったく同じ方

$$\sigma_c^2 = \frac{\sigma_d^2 \sigma_t^2}{\sigma_d^2 + \sigma_t^2} = \frac{1}{\frac{1}{\sigma_d^2} + \frac{1}{\sigma_t^2}}. \quad (2)$$

方で同じ予測したが、これはまったく異なるメカニズムを示した。

ベクトル和のモデルを正しく予測できることをさらに実証して、JNDの大きさと独立に測定された大きさと、実験1で得られた手がかりの強さとの間に密接な関係があることを示した。JNDの解釈が、IC理論による経験的結果と非常に整合的であることが示されたので、実験2では、IC理論によるJNDの解釈が奥行き弁別課題の経験的結果と非常に整合的であることを検討した。

実験2では刺激は実験1と同じで、しかし絶対的な奥行き深さの知覚の判定は行わず、代わりに2IFC奥行き弁別課題を実施した。ベクトル和のモデルからJNDを定量的に予測するためには、知覚された奥行きが単一手がかりと複合手がかり条件で観察しても、測定せずに各観察条件の各被験者について3つの奥行き深さを推定させて、そのかわりに、被験者には2IFC選択法で奥行き識別課題を行った。ベクトル和のモデルからJNDを定量的に予測する

ために知覚される深さは、単一手がかり基準と複合手がかり基準で一致しなければならない。そのため、知覚の深さに関するにノイズは一定に保たれた。このために実験1からのデータを3組(視差のみ、テクスチャのみ、複合した奥行き手がかり)のシミュレートした奥行きを知覚された2回の奥行きとしてもちいて検討した。被験者は2IFC課題を行い、知覚された深さが固定された標準刺激と、シミュレートした奥行き深さを段階的に変化させた比較刺激の知覚的深さと、どちらが深いかを比較した。JNDは、PSEと被験者が反応したシミュレートした奥行き深さとの差と定義された。

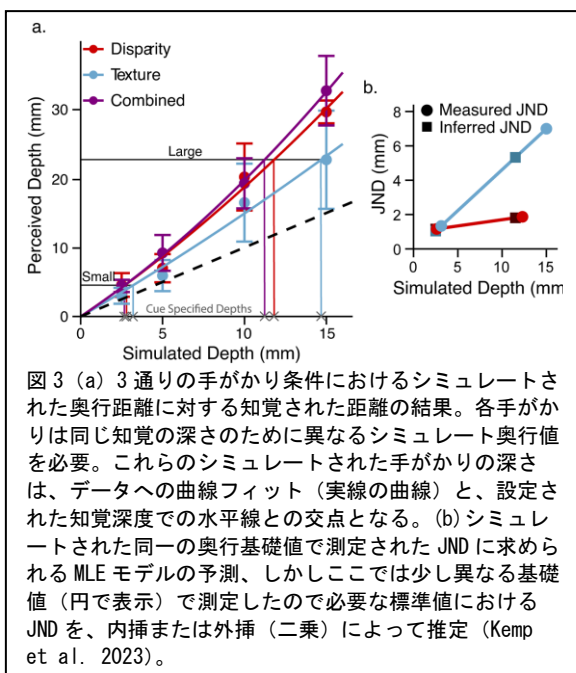


図3 (a) 3通りの手がかり条件におけるシミュレートされた奥行き距離に対する知覚された距離の結果。各手がかりは同じ知覚の深さのために異なるシミュレート奥行き値を必要とする。これらのシミュレートされた手がかりの深さは、データへの曲線フィット(実線の曲線)と、設定された知覚深度での水平線との交点となる。(b)シミュレートされた同一の奥行き基礎値で測定されたJNDに求められるMLEモデルの予測、しかしここでは少し異なる基礎値(円で表示)で測定したので必要な標準値におけるJNDを、内挿または外挿(二乗)によって推定(Kemp et al. 2023)。

実験2の結果、標準刺激の知覚された奥行の深さは、図3(a)にあるようにJNDが3通りの条件 disparity (赤色)、texture (青色)、combined cue (パープル色) ごとに示された。濃い灰色はベクトル和のモデルの予測値を表し、薄い灰色はMLE 予測値を表す。図(b)に示すように、測定されたJNDの線形グラフ(丸印)から、MLEモデルの適切な値(四角)を推定すると、2つのモデルの予測は予想通り非常によく似て精度に有意差はない。MLEモデルとベクトル和のモデルの両方にとって重要なのは、複合した手がかり条件は単一の手がかり条件よりJNDが小さいことであった。IC理論では、JNDは、推定ノイズの指標ではなく、むしろタスクに関連した要求や比較刺激の手がかり強度から生じるノイズである。

これらの結果は、JNDは奥行の深さ推定のノイズではなく、手がかりの強さによって説明できることを示し、デジタルの3D特性とモジュール出力との間の奥行深度モジュールの決定論的マッピングを仮定するIC理論の予測と一致する。

実験3では、2つの目的、すなわち一つ目は、MLEモデルとベクトル和のモデルの予測値の違いを検証、フラット面を指定する手がかり(フラット面の手がかり)があるかどうか、二つ目は、テクスチャのみの刺激の両眼視の使用を正当化することである。MLEモデルは、複合的な手がかりを持つ刺激がフラット面性の手がかり刺激よりも深く感じられるという発見に対応することができ、MLEモデル複合手がかり刺激が単一手がかり刺激よりも深く知覚されるか否かである。

実験3Aでは、刺激はテクスチャーとシェーディングで構成された同じ画像を左右の眼にシャッターグラスに提示、両眼視差によって前額平行面を提示し、単眼視と両眼視をランダムに混在させて観察させた。実験3Bでは、被験者には2条件(RDSテクスチャーなし、フラットテクスチャ)の視差条件で知覚された奥行の判断を実施させた。

実験3Aの結果、平均知覚奥行値は観察距離2条件および単眼と両眼条件ではシミュレートした奥行に有意な影響が示され、その平均値は実験1より大きかった(図4)。多分、実験1では複数の手がかりを含む試行と交互に行われることはなかったため被験者は判断基準を単眼視手がかり条件に調整した。フラットなテクスチャとフラットRDS条件の知覚された奥行はシミュレーションされた奥行、さらにそれと凝視距離との間に有意な効果があることを明らかにした。これらの実験結果では、奥行手がかりの深さをゼロに設定すると、それを完全に削除するのとでの間に違いはないのでベクトル和のモデルの予測を裏付けている。

結論として、IC理論の方が3Dの手がかり統合実験を説明するには、IC理論の方が次の点で適している。(1) ベイズモデルでこれまでデータを予測できる、(2) これまでのモデルでは新たな結果は予測できない、(3) このモデルではベイズ推定が機能するため

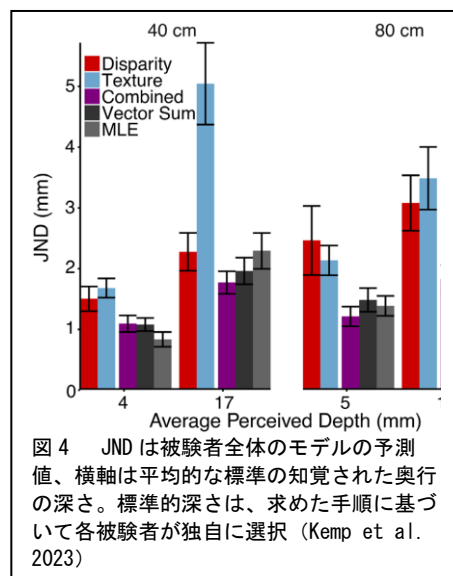


図4 JNDは被験者全体のモデルの予測値、横軸は平均的な標準的知覚された奥行の深さ。標準的深さは、求めた手順に基づいて各被験者が独自に選択 (Kemp et al. 2023)

に必要な、知覚の真偽や手がかりの信頼性を推定する必要がない簡潔なモデルである。

このベクトル和のモデル (vector sum model) は、個々の手がかりがノルムの合成推定値を決定する多次元の構成要素のベクトルであり、そのノルムが合成推定値を決定するのを評価して目標を達成する規範的アプローチを提供する。個々の手がかりの評価は正確ではないがデジタルな 3D の特性を決定的なマッピングを通して関係する。