講　座《新連載》

映像情報×認知科学〔全6回〕開講にあたって

編集幹事 寺岡丈博，張　英夏，金成　慧

近年，大規模言語モデル（LLM）やDiffusionモデルの進展にともない，ChatGPTやGemini，Stable Diffusionなど，言語や画像の生成AIが広く注目されています．LLMだけを見ても，言語処理のみならず，意思決定や因果推論などに関して私たち人間に比肩するくらいの性能を実現しています．しかし，これらに関する人間の認知機能が解明された上で発展を遂げたわけではなく，そのような意味で認知科学の知見が充分に活かされているとはいえません．

一方で，人間の認知機能について解明が進み，その知見を基にしたアプローチ，例えば視覚などの感覚処理を模倣した手法が注目を集めています．また，現在のAIでは処理できない言語・行動・コミュニケーションに関する認知機能について，映像情報を利用して解明されつつあります．

本講座では，こうした人間の認知に関する研究動向と応用の可能性を探求します．第1回から第4回にかけては視覚に関する研究や映像情報への応用について，第5回と第6回は映像情報に基づいた人間の言語獲得やマルチモーダルインタラクションについて，執筆者の方々に取り上げていただきます．本講座を通して，映像情報と認知科学の可能性について考える契機になることを願ってやみません．

予定目次

講　座　映像情報×認知科学〔第1回〕

瞳　孔

金成　慧†

1．まえがき

瞳孔は，虹彩（目にある模様のような部分）に囲まれた孔であり，目に入る光の量を調整する機能をもつ．しかしながら，古代ローマ帝国の時代から言われてきたように（目は魂の鏡（キケロ：106-47 BCE）；目に心は宿る（プリニウス：23-79 CE）），瞳孔はさまざまな認知処理の影響を受けることが多くの研究で示されている．本稿では，瞳孔の基本的な光学的機能に加え，瞳孔と関連する認知処理のうち，情動と視覚的注意に関する研究について取り上げる．そして，それらの知見を利用したヒューマンマシンインタフェースなどの応用的な展開についても紹介する．

2．光学的機能

瞳孔から入った光は網膜に到達し，そこからの反射光は光源の位置にしか戻らないため，瞳孔は私たちには黒く見える．虹彩にあるメラニン色素は太陽からの紫外線を吸収することで細胞を守る役割があり，色素の量によって虹彩の色が変わり，その重なり具合が虹彩の模様となる．日本など太陽光の強い地域の人の色素量は多く，虹彩は黒や茶色に見える．一方，北ヨーロッパなど太陽光の弱い地域の人の色素量は少なく，虹彩は青や灰色に見える．

虹彩には瞳孔括約筋と瞳孔散大筋の二つの平滑筋がある．明るいところでは，輪状の瞳孔括約筋が収縮することで瞳孔の大きさが小さくなり（縮瞳），目に入る光を減らす．暗いところでは，放射線状の瞳孔散大筋が収縮することで瞳孔が大きくなり（散瞳），目に入る光を増やす．瞳孔の反応には個人差があるが，瞳孔の直径は1.5～9 mmの範囲で変化し，約200 msで刺激に反応する[1]．標準的な照明環境下では，瞳孔の大きさは約3 mmである[2]．

瞳孔には光量の調整以外にも，網膜に映る像のボケを減少させる機能もある．一つは，近い対象を見るときに瞳孔が収縮し，被写界深度を深めることで網膜像のボケを減少させる機能である（近見反応）．もう一つは，錐体が機能するのに充分明るいレベル（明所視）において，瞳孔が収縮することで，色収差（光の波長ごとに屈折率が異なるため，焦点位置が変化する現象収差）や球面収差（光が光軸に平行に入射したとき，光軸から離れるほど焦点からずれる現象）によるボケを減少させる．

3．瞳孔と視覚的注意

瞳孔は視線位置の輝度だけでなく，視覚的に内因性注意（意識的に制御できるトップダウン注意）を向けた位置の輝度に対応して変化することが示されている．Bindaら[3]の実験では，画面中央に固視点と注意を向ける方向を示す手がかりを始めに呈示し，その左右に白と黒の円を呈示した．被験者は固視点を見たまま，教示された方向の円に注意を向け，注意タスクとして，その円の中心のドットの色が変化した回数を応答した．その結果，白い円に注意を向けたときの瞳孔は黒い円よりも収縮した．Mathôtら[4]は，注意タスクとして，ガボールパッチの傾きを応答してもらうことで同様の結果を示した．また，刺激輝度の周波数変化[5]，特徴ベース[6]および空間ベース[7]の注意において，刺激輝度と対応した瞳孔変化が生じることが明らかになっている．さらに，近見反応と関連して，注意を向けた刺激の空間周波数によって瞳孔が調整されることも示されている[8]．

外因性注意（フラッシュのような手がかり刺激に対して空間的注意が無意識的に向くこと）に関しても，注意が向いた位置の輝度と対応して瞳孔が変化する[9]．Mathôtらは，手がかり刺激としてガボールパッチの位相を変化させることで外因性注意を向けさせた．また，手がかり刺激呈示からターゲット呈示までの時間間隔（SOA: Stimulus Onset Asynchrony）を変化させることで，瞳孔が外因性注意と復帰抑制（ターゲットが手がかり刺激から300 msほど後に呈示されると，反応時間が遅れる現象）によって調整されるか検討した．その結果，外因性注意が向いた位置の輝度と対応して瞳孔が変化した．さらに，IORが1000 msのとき，ターゲットに対する復帰抑制が生じたと同時に瞳孔の変化の方向が逆転した．このことから，瞳孔は注意によって調

† 東北大学
"Visual Information × Cognitive Science (1): Pupillometry" by Kei Kanari (Tohoku University, Miyagi)

映像情報メディア学会誌 Vol. 79, No. 2, pp. 168～171（2025）

整されることで，視覚入力を安定させ，色収差や球面収差によるボケを減少させることで視力を最適化している可能性が示唆される．

4．瞳孔と情動

瞳孔は自律神経に支配されているため，交感神経系が優位なときには散瞳し，副交感神経系が優位なときには縮瞳する．そのため，瞳孔反応は自律神経系と関連すると考えられる情動反応の指標として用いられる．

情動を喚起させる画像で構成されたデータセットであるInternational Affective Picture System（IAPS）を用いて瞳孔反応を検討した研究では，pleasant/unpleasant(快/不快)な評価の画像は，中立評価の画像より瞳孔が散大することが示されている[10]．また，快・不快に関わらず，Arousal(興奮度評価である覚醒度)が高い方がより瞳孔が散大した．この傾向は絵画鑑賞時の瞳孔反応でも報告されている[11]．絵画を鑑賞する際，さまざまな情動が生じるが，この研究では快・不快評価である感情価（Valence），興奮度評価である覚醒度（Arousal），絵画の好み（Liking）の三つの評価と瞳孔反応との関係を検討した．刺激には，絵画データセットであるThe Vienna Art Picture System（VAPS）から選ばれた50枚の絵画を用い，絵画の種類はScene, Portrait, Landscape, Still life, Abstractの5種類であった．実験の結果，絵画呈示後1～2秒の区間において，Liking・Valenceが高いと瞳孔が収縮し，Arousalが高いと瞳孔が散大した．ValenceとArousalの間に負の相関が見られたため，鑑賞した絵画への情動が快・不快であるかに関係なく，興奮によって瞳孔が散大したと考えられる．

このように，瞳孔は人の内部状態を反映するため，それをコミュニケーションのツールとして用いる社会的機能も担っている可能性が示唆されている．例えば，男性に瞳孔の大きさだけが異なる同じ女性の写真をペアで呈示した．その結果，写真の瞳孔の大きさの違いに気づいていないにもかかわらず，瞳孔の大きい女性を“soft", "more feminine", "pretty”と評価した[12]．また，同一男性・同一女性の瞳孔の大きさがそれぞれ異なる写真4枚に対する瞳孔反応を測定した．その結果，男性被験者の瞳孔は瞳孔の大きい女性の写真に対してより散大し，女性被験者の瞳孔も瞳孔の大きい男性の写真に対して，より散大した[13]．瞳孔が大きいことはその人と対面している人に対する好意(興味)を示していると解釈されるため，この結果は，人は好意を抱いてくれる人に対して好意を返すという「好意の返報性」が関連していると考えられる．

興味や性的魅力に対しては瞳孔の散大が生じるが，人の顔の魅力に関しては異なる結果が示されている[14][15]．かわいさと瞳孔反応の関係を検討した研究では，動物や食べ物など人以外の画像に対するかわいさの評価（かわいさ評価）が高いほど，観察者の瞳孔が散大する結果が見られた．しかしながら，成人女性の顔を観察した際，そのかわいさ評価が高いほど観察者の瞳孔が縮小することが示された．人の顔とその他の対象で反対の瞳孔反応が生じた理由の一つとして，かわいさ評価が低い顔に対する嫌悪感の影響が考えられる．魅力的な顔の特徴の一つは対称性であり，これは配偶者の健康や質の指標である．したがって，かわいさ評価の低い顔は非対称であり，これは生理的健康状態の悪さを反映し，嫌悪感が生じたことにより瞳孔が縮小した可能性がある．人の顔とその他の対象で反対の瞳孔反応が生じた別の理由として，この研究では，刺激の顔と観察者が成人女性のみであったため，同性のかわいさに対する嫉妬と関与していた可能性がある．女性は他の女性の身体的魅力に特別な注意を払い，嫉妬を経験することが知られている．そのため，同性の人のかわいらしさを評価する際に，嫉妬の影響により，よりかわいい顔に対する瞳孔の散大が抑制された可能性がある．このように，瞳孔と魅力の評価には観察者と観察される顔の性別や情動が関連するため，今後は異なる性別間での検討が必要である．

瞳孔は輝度やコントラストなど視覚的な特性の影響を受けるため，刺激として音を使用することで，それらの要因を排除できる．情動音として，拡張版国際情動音刺激データベース（IADS-E: Expanded Version of the International Affective Digitized Sounds）を用いた研究[16]では，Arousal刺激として20個，Valence刺激として20個の計40個に対する瞳孔反応を測定した．その結果，Arousalと瞳孔反応の間に正の相関が見られ，Valenceと瞳孔反応の間に負の相関が見られた．ArousalとValenceには負の相関が見られたため，これらの結果は画像を用いた先行研究と同様に，情動的な興奮（Arousal）によって交感神経系が優位になり，瞳孔が散大したことを示唆する．さらに，この研究では，情動音が呈示されると同時に，異なる輝度の対象に視覚的注意を切り替えている．そのため，瞳孔が視覚的注意における輝度の変化と情動からどのような影響を受けるか検討している．その結果，注意を白から黒のドットパターンに切り替えた後の瞳孔は，黒から白に切り替えた後よりも散大した．また，どちらの条件においても，Arousalと瞳孔反応の間に正の相関が見られた．したがって，瞳孔は視覚的な注意における輝度の変化と情動の両方から影響を受けるが，その効果は分離できることが示唆される．

5．瞳孔反応を用いた障害診断

瞳孔が注意を向けている位置や対象の輝度に応じて変化することは，注意機能障害をもつ自閉症スペクトラム障害（ASD：Autism-spectrum Disorder）などの診断や重症度の評価に応用できる可能性がある．ASDの特徴として，パターン内に隠された対象を見つける能力は高いものの，コヒーレント運動や顔認識が苦手であることが知られている．これは，ASDのグローバル処理や運動処理に障害があることを示しており，注意機能障害との関連が考えられる．この特性を利用し，瞳孔反応を測定して注意障害の度合い

を診断する研究が進んでいる．

Turiらの実験[17]では，反対方向に運動するランダムドットパターンを重ねて呈示し，それぞれのパターンは白と黒で明るさが異なっていた．この刺激は回転する円柱のように知覚され，回転方向は双安定性で，時計回りと反時計回りが交互に切り替わる（https://doi.org/10.7554/eLife.32399.005）．白のパターンが右方向，黒のパターンが左方向に動くため，円柱が時計回りに見えるときは白が手前に，反時計回りに見えるときは黒が手前に見える．定型発達の被験者は円柱の回転方向を持続的に応答し，時計回りから反時計回りに切り替わった後の瞳孔は，反時計回りから時計回りに切り替わった後の瞳孔よりも拡大した．このことから，被験者が円柱の手前のパターンに注意を向けていたことが示唆される．また，瞳孔変化の差は，質問紙で得られたASDの傾向の度合い（AQ: Autism-Spectrum Quotient）が高いほど大きかった．つまり，AQが高い被験者はローカルな処理，AQが低い被験者は円柱全体に注意を向けるグローバルな処理を行っていたと考えられる．これにより，瞳孔反応の測定が個人の注意に関連する障害の診断に役立つ可能性が示された．ただし，ASDはADHD（Attention Deficit Hyperactivity Disorder：注意欠陥/多動性障害）や不安症などの合併が多いため，これらと共通するメカニズムの関連について検討が必要である．

瞳孔対光反射における瞳孔パラメータ（初期瞳孔径，平均瞳孔径，潜時，振幅，縮瞳速度，相対縮瞳率など）は自律神経系の機能を反映する指標として用いられている．ASDやADHDなどの発達障害において，自律神経系の機能障害が示されおり，子どもの障害児と定型発達の2グループ間での瞳孔反応が異なることが報告されている[18]．しかしながら，ADHDにおける瞳孔対光反射の非定型性については結果が一致していない[19]．ASDに現れる発達特性は健常者においても連続的に現れるという自閉症スペクトラム仮説が提唱されている．また，ADHDの症状は一般集団に連続的に分布していると考えられる．そこで，成人健常者の発達障害傾向と瞳孔対光反射を用いた自律神経系の機能が関連する可能性がある．しかしながら，感覚処理傾向（AASP: Adolescent/Adult Sensory Profile）との相関を示すものはあるが[20][21]，ASDやADHDとの関連は報告されていない．これは，ASD, ADHDの傾向の高い被験者が少なくサンプルの分布に偏りがあったことや，成長するにつれて瞳孔調整機能が改善されている可能性が考えられる．今後は子どもから成人まで幅広い年代で多くのサンプルを対象としたさらなる検討が必要である．

6．瞳孔反応を用いた情報入力

近年，手足の動きや発話での意思表示ができない重度の障害を持つ人のための意思伝達手法として，視線移動の代わりに，瞳孔と視覚的注意の関係を活用した非接触型の文字入力装置の開発が進められている．例えば，同心円上に文字を配置し，文字の背景輝度を周期的に変化させたり[22]，テンキーのように文字を配置し，文字背景の輝度変動を異なる周波数で変化させ[23]，瞳孔反応から注意を向けた文字を推定するというものである．しかし，瞳孔反応という一つの指標のみで入力文字を推定しているため，入力時間が長くなるといった課題がある．ここでは，瞳孔反応に加えて視運動性眼振（Optokinetic Nystagmus: OKN）と呼ばれる眼球運動の二つの指標を用いた情報入力手法[24]について紹介する．

OKNは視野内に持続的に動く物体が現れたときに引き起こされる眼球運動で，緩やかに物体を追う緩徐相と，反対方向に素早く戻る急速相が交互に繰り返される．OKNも瞳孔と同様に，視線位置だけではなく，注意を向けた対象に対応したOKNが発生する．この知見をもとに，運動方向と輝度が異なるパターンを呈示し，注意を切り替えるタイミングをOKNと瞳孔反応から推定することで，入力文字を決定する手法を提案している．実験では，円内に運動方向（右と左）と輝度（黒と白）が異なる2種類のドットを配置し，円内上部に文字を呈示させた．文字の輝度・運動方向は円内を運動しているどちらかのドットと同じだった．被験者は始めにターゲット文字と異なる明るさのドットに注意し，ターゲット文字が円内に現れている間，文字と同じ輝度のドットに注意した（図1）．

実験の結果，ターゲット文字が円内に現れて注意を切り替えたときに，注意対象の輝度・運動方向と対応した瞳孔・OKNの変化が見られた．瞳孔の場合，輝度の条件に関わらず，ターゲットが呈示されると瞳孔が散大する傾向が見られた（図2（左））．非ターゲット文字が円内に呈示された場合，瞳孔は変化することはなかった．したがって，この散瞳は注意を切り替える精神的努力[25]によるものと考えられる．その後，黒から白ドットに注意を切り替える条件では，ターゲット文字が呈示された約0.68 s後，瞳孔が収縮し始めたが，白から黒の条件では，瞳孔は散大が継続した．OKNの場合，運動方向の条件に関わらず，ターゲット文字が呈示された約0.2～0.3後に緩徐相の速度が0を通過し，速度の方向が切り替わった（図2（右））．このことから，入力文字を推定できることが示唆される．

上記の先行研究から，入力候補の文字数を46文字に増やし，OKNの急速相の速度に基づいてリアルタイムで文字入力を推定するシステムの開発も行われている[26]．このシステムでは，同心円上を文字が回転し，被験者はターゲット文字が特定の位置に来たときに，固視点から運動パターンに注意を切り替える（図3）．そのとき発生するOKNの急速相の速度から入力文字を推定する手法となる．実験の結果，4.42 msで入力文字を推定できたが，これは眼球運動装置やプログラムソフトのサンプリングレートに依存する．文字の推定は60.2%であったが，エラーの原因は瞬きによる影響がほとんどであるため，開眼率などを用いて瞬きのデータを除外することでさらに精度が上がることが期待される．今後は既存の入力システム（キーボードなど）と組み

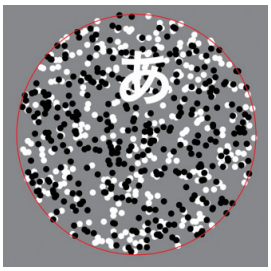


図1　瞳孔・OKNを用いた文字入力画面

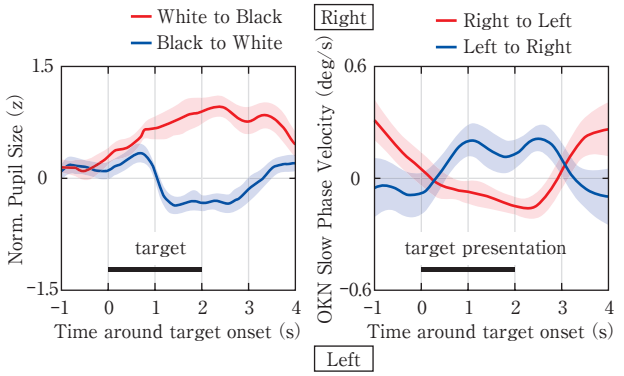


図2　瞳孔とOKNの結果

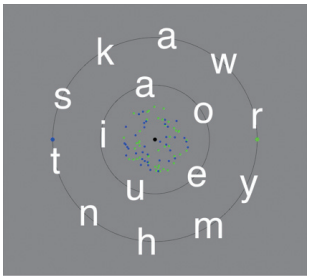


図3　OKNを用いた文字入力画面

合わせた健常者も利用できる汎用システムの開発や，瞳孔反応・心拍変動を用いた心理状態を推定する機能の追加などが期待できる．

7．むすび

瞳孔反応は人が意識的に制御できない無意識的な反応であるため，注意の向け方や感情状態など，内部の心理状態を推定する信頼性の高い指標と考えられる．また，従来の生理学的計測手法（EEG，fMRI，MEG）と比較して，瞳孔反応は測定に特別なトレーニングが不要で，非侵襲的であり，装置が安価かつ簡便に利用できるという多くの利点がある．こうした特徴から，瞳孔反応は今後，ヒューマンマシンインタフェースのツールや新たな障害診断の方法として活用されることが大いに期待される．　（2024年11月5日受付）

〔文　献〕

1) O. Lowenstein and I.E. Loewenfeld: "The pupil", Academic Press, New York (1969)
2) H.J. Wyatt: "The form of the human pupil", Vision Research, **35**, 14, pp.2021-2036 (1995)
3) P. Binda, M. Pereverzeva and S.O. Murray: "Attention to bright surfaces enhances the pupillary light reflex", Journal of Neuroscience, **33**, 5, pp.2199-2204 (2013)
4) S. Mathôt, et al: "The pupillary light response reveals the focus of covert visual attention", PLoS One, **8**, 10, e78168 (2013)
5) M. Naber, et al: "Tracking the allocation of attention using human pupillary oscillations", Frontiers in psychology, **4**, 919 (2013)
6) P. Binda, et al: "Pupil size reflects the focus of feature-based attention", J. Neurophysiol, **112**, 12, pp.3046-3052 (2014)
7) P. Binda and S.O. Murray: "Spatial attention increases the pupillary response to light changes", Journal of Vision, **15**, 2, pp.1-13 (2015)
8) X. Hu, R. Hisakata and H. Kaneko: "Effects of spatial frequency and attention on pupillary response", JOSA A, **36**, 10, pp.1699-1708 (2019)
9) S. Mathôt, et al.: "The pupillary light response reflects exogenous attention and inhibition of return", Journal of Vision, **14**, 7, pp.1-9 (2014)
10) M.M. Bradley, et al.: "The pupil as a measure of emotional arousal and autonomic activation", Psychophysiology, **45**, 4, pp.602-607 (2008)
11) Y. Sasaki, M. Sato and K. Kanari: "Effect of visual saliency on emotions while viewing paintings", the 8th IIEEJ International Conference on Image Electronics and Visual Computing (2024)
12) E.H. Hess: "Attitude and pupil size", Scientific American, **212**, 4, pp.46-55 (1965)
13) T.M. Simms: "Pupillary response of male and female subjects to pupillary difference in male and female picture stimuli", Perception & Psychophysics, **2**, 11, pp.553-555 (1967)
14) K. Kuraguchi and K. Kanari: "Enlargement of female pupils when perceiving something cute", Scientific Reports, **11**, 1, 23367 (2021)
15) H.I. Liao, M. Kashino and S. Shimojo: "Attractiveness in the eyes: A possibility of positive loop between transient pupil constriction and facial attraction", Journal of Cognitive Neuroscience, **33**, 2, pp.315-340 (2021)
16) 堤，清水，富田，二宮，金成：“感情音と輝度が視運動性眼振と瞳孔反応に与える影響”，信学総大（2024）
17) M. Turi, D.C. Burr and P. Binda: "Pupillometry reveals perceptual differences that are tightly linked to autistic traits in typical adults", Elife, **7**, e32399 (2018)
18) X. Fan, et al.: "Abnormal transient pupillary light reflex in individuals with autism spectrum disorders", Journal of Autism and Developmental Disorders, **39**, 11, pp1499-1508 (2009)
19) A. Bellato, et al.: "Is autonomic nervous system function atypical in attention deficit hyperactivity disorder (ADHD)? a systematic review of the evidence", Neuroscience & Biobehavioral Reviews, **108**, pp182-206 (2020)
20) 金成，菊地：“健常者における精神障害の傾向と瞳孔対光反射との関係”，VISION，**34**，4，p.132（2022）
21) 堤，佐藤，菊地，金成：”成人健常者における発達障害傾向及び感覚特性と瞳孔反応の関係”，信学技報，**123**，205，pp.39-42（2023）
22) S. Mathôt, et al.: "The mind-writing pupil: A human-computer interface based on decoding of covert attention through pupillometry", PloS one, **11**, 2, e0148805 (2016)
23) Y. Muto, H. Miyoshi and H. Kaneko: "Eye-gaze information input based on pupillary response to visual stimulus with luminance modulation", Plos one, **15**, 1, e0226991 (2020)
24) 金成，猪股：“視運動性眼振と瞳孔反応に基づいた情報入力手法の検討”，視覚の科学，**44**，2，pp.35-43（2023）
25) D. Kahneman: "Attention and effort", **1063**, pp.218-226, Englewood Cliffs, NJ: Prentice-Hall (1973)
26) 清水，堤，富田，二宮，金成：“視運動性眼振を用いたユーザの注意状態推定によるリアルタイムな文字入力手法の開発”，信学総大（2024）

金成（かなり）　慧（けい）　2014年，東京工業大学大学院総合理工学研究科博士課程修了．2015年，同大学研究員．2017年，玉川大学脳科学研究所嘱託研究員．2020年，宇都宮大学工学部助教．2024年より，東北大学総合知インフォマティクス研究センター特任助教．専門は視覚心理物理学，主に眼球運動，瞳孔反応に関する研究に従事．博士（工学）．

講　座　映像情報×認知科学〔第2回〕

質感の視知覚

正会員 清川宏暁†

1．まえがき

われわれが生きる世界にはさまざまな物体が存在し，それぞれの物体が材質に応じた質感を知覚させる．例えば，磨かれた金属を見ると光沢感を知覚し，蝋細工や瑞々しい果物の果肉を見ると半透明感を知覚する．これらの知覚は，誰もが特別な努力を必要とせずに行うことができる．しかし，質感知覚を支えるための脳内の視覚メカニズムについては未解明な部分が多い．近年，質感知覚のメカニズム解明を目指し，心理物理学や神経科学の面から解明を目指す研究が数多く行われてきた．それらの研究により，質感を知覚するために脳が手がかりとして用いる画像特徴や，脳がどのような情報処理を経て質感を知覚するのかについての知見が集積されてきた．これらの知見は，映像中のリアルな質感再現や，特定の質感を操作するような質感編集技術などの基盤知見として，映像情報メディアの発展に寄与すると期待できる．

本稿では，主に視覚的に知覚される質感である光沢感や半透明感に注目し，質感を知覚するために脳の視覚系が解くべき問題を解説する．その後，どのような画像特徴が質感知覚の手がかりとなるのか，そして，脳内でどのようにしてそれらの手がかりが処理されているのかを解説する．

2．質感を知覚するために脳が解くべき問題

われわれの視覚入力のフロントエンドは眼球内の網膜であり，そこに外界からの光が結像することで網膜像が生じる．この網膜像から，質感の知覚のための手がかりとなる特徴を読み取る必要がある．ある物体を観察した際に，網膜像を構成する要因は主に物体の形状，反射・散乱特性，照明環境の3点が挙げられる．光沢感や半透明感の知覚は，網膜像からこれら三つの要因を分離する逆問題を解き，さらに，反射・散乱特性を推定する問題と言い換えることができる．ところが，この逆問題は，解を一意に求めることができない不良設計問題である．例えば，計算論的に三つの要因のうち一つを求めたければ，他の2要因を何らかの仮定を元に固定する必要がある．実際に，単一の2次元画像の陰影から3次元形状を推定する問題（shape from shading）を解く場合は，照明方向や表面の反射特性について仮定を置いて問題を解くことになる[1]．これは脳にとって非常に困難，かつ，計算効率の悪いアプローチであり，われわれの高速で安定した知覚を支えるメカニズムであるとは考えにくい．過去の研究でも，われわれの視覚系が物体の反射特性の真値を正しく推定できないことを報告している[2)3)]．例えば，Nishida & Shinya[2]は，人間の観察者を対象に，コンピュータグラフィックスを用いて，異なる3次元形状を持つ物体画像間で反射特性のパラメータをマッチさせた．その結果は，マッチング結果に体系的なバイアスを持った誤差が生じることを報告した．また，興味深いことに，文献[2]は，上記の誤差バイアスが，物体画像の輝度ヒストグラムと関連することも報告した．これらの報告は，視覚系による逆推定問題の解決能力に限界があること，そして，単純な画像特徴を用いることで物体の反射特性を推定している可能性を示していた．

3．質感知覚と画像特徴

視覚系が質感を知覚するために，どのような画像特徴を利用しているのかを検討する研究が積極的に行われてきた．先駆的な報告として，Motoyoshiら[4]は，物体表面の鏡面反射率と画像の輝度ヒストグラムの歪度と呼ばれる画像統計量が関連することを見出した（図1）．高い鏡面反射率を持つ物体表面には，入射光が高い指向性を持って反射する鏡面反射光が発生する．一般的に，鏡面反射光は物体表面に，小さな高輝度領域である鏡面ハイライト（光沢ツヤとも呼ばれる）と呼ばれる領域となって現れる．この領域が生まれることで，物体画像の輝度ヒストグラムは正の方向へ歪む．Motoyoshiら[4]の報告は，この物理現象と輝度変化の対応関係を人間の視覚系が光沢感知覚のために利用していることを示唆していた．また，Motoyoshiら[4]は，

† 埼玉大学 大学院理工学研究科
"Visual Information × Cognitive Science (2): The Mechanisms of Material Perception" by Hiroaki Kiyokawa (Graduate school of Science and Engineering, Saitama University, Saitama)

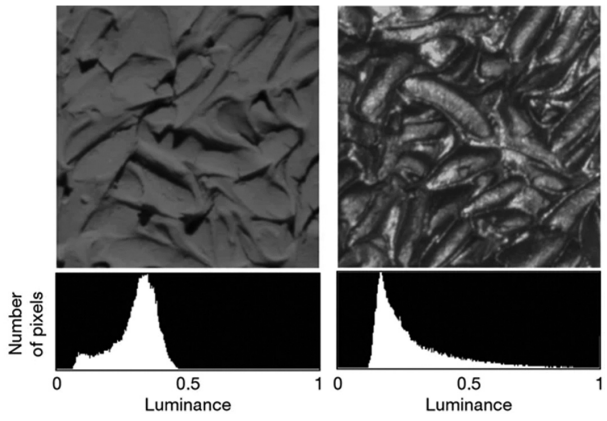


図1　物体表面画像と輝度ヒストグラム[4]
左上の画像からは光沢感を知覚しないが，右上の画像からは光沢感を知覚する．下段はそれぞれの画像の輝度ヒストグラムを表す．輝度ヒストグラムは，右の光沢感を知覚する画像では正の方向へ歪むが，左の光沢感を知覚しない画像では負の方向へ歪む．

輝度ヒストグラムの歪度を計算する神経機構が，外側膝状体や第1次視覚野のような，両眼間の情報が保持されている低次の領野に存在する可能性を実験的に示した．

その他にも複数の研究で，質感知覚と画像統計量の間に関連が有ることが報告されてきた．例えば，輝度ヒストグラムから計算可能な統計量である，輝度コントラストを用いることで人間の光沢感知覚の変化を説明できることを示した報告も存在する[5]．また，半透明感についても，空間周波数帯域別の輝度コントラスト[6]や局所的な輝度勾配の方位分布の異方性[7]により人間の知覚をよく説明できることが報告されている．これらの統計量は，脳内の受容野構造や受容野応答の空間プーリングの理論から，脳内で計算可能と考えられている．

その一方で，輝度ヒストグラム統計量だけでは，人間の質感知覚の説明には不充分であることも報告されている．図2は画像の輝度ヒストグラムの歪度やコントラストは同一である．しかし，陰影成分に対して鏡面ハイライトを回転させることで幾何学的な整合性を損なわせた例である．この処理によって，輝度ヒストグラム統計量は同一であるにも関わらず，鏡面ハイライトが白い塗料による着色に見える．これは，高輝度領域が鏡面ハイライトとして認識されるために満たすべき，画像の幾何学的な制約が存在することを示唆している．さらに，その後の研究によって，まったく同じ輝度勾配の刺激であっても輪郭や両眼視差[9]（図3に輪郭の効果を示す）によって，知覚される3次元形状を明示的に変化させると光沢感の有無が変化することが報告されている．これらの報告は，脳が質感を知覚するために，輝度ヒストグラム統計量のような低次の画像特徴のみではなく，物体の3次元形状の推定に関連する，より高次の画像特徴を用いていることを示唆している．

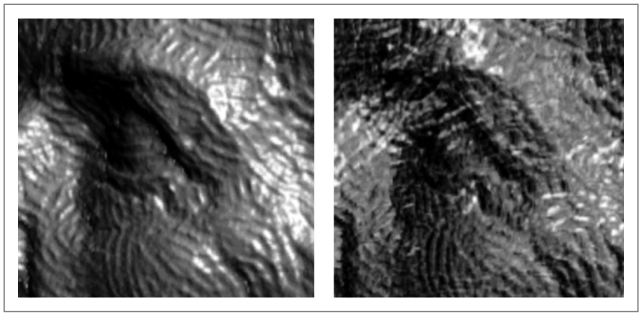


図2　光沢感知覚における陰影と鏡面反射成分の幾何学的整合性の影響[8]
左図が元画像であり，右図は鏡面反射成分のみを回転させた画像である．輝度ヒストグラムは左右の間で同一であるように変換をかけている．左図からは光沢感を知覚するが，右図からは光沢感を知覚できない．

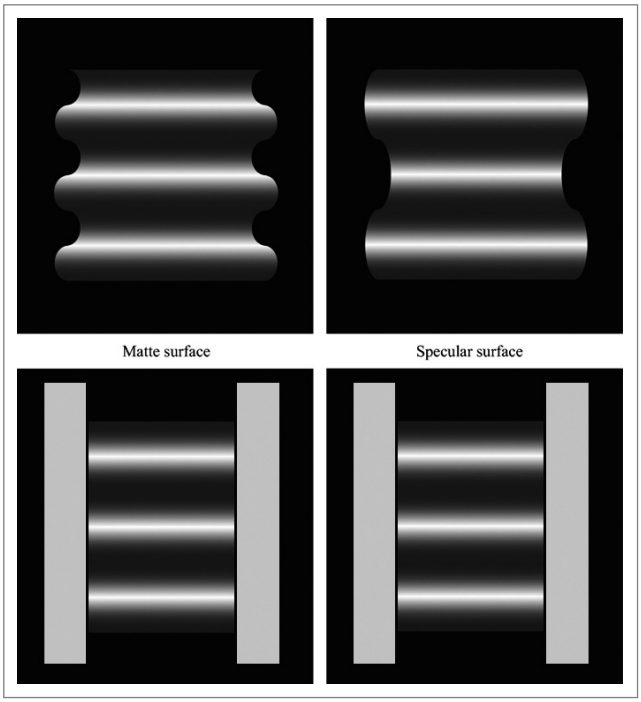


図3　見かけの3次元形状と光沢感知覚[9]
上段：左図は，凸面が三つの光沢のない物体表面に見える．一方で，右図は凸面が二つの光沢の有る物体表面に見える．下段：上段画像の左右の輪郭を遮蔽した画像を示す．どちらもまったく同じ輝度分布のパターンを持つことがわかる．このデモから，同一の輝度パターンであっても，輪郭によって異なる3次元形状の知覚が誘発されることで，知覚される光沢感も変化することがわかる．

このように，光沢感や半透明感のような質感について，知覚のための手がかりとなるさまざまな画像特徴が報告されている．また，今回報告した研究結果は，利用できる手がかりに応じて質感知覚の応答戦略も柔軟に変化することを示唆している．

4．質感知覚のための脳内の情報処理

脳内でどのような情報処理を経て質感が知覚されるのかについても研究が行われている．これまでに，人間やサルを対象とした脳活動計測により，質感知覚のための神経基

盤について検討が行われてきた．例えば，さまざまな素材画像を観察している際に脳活動計測を行うと，腹側経路の低次領野では画像特徴と相関する活動パターンを示すが，高次領野では素材の違いと相関する活動パターンとなる．つまり，低次から高次の領野に処理が進むにつれて，画像特徴の情報が素材の違いのような質感に関連する表現へと徐々に変換されていくことが報告されている[10) 11)]．

また，特定の画像特徴の変化に選択的に応答するニューロンの存在も報告されている．例えば，サルの脳を対象とした電気生理学的実験から，腹側経路の高次領野であるInferior Temporal（IT）野には，鏡面反射成分と陰影成分のコントラストや，鏡面反射成分のぼやけ具合の変化に選択的に反応するニューロン群が存在することが報告されている[12)]．また，Babaら[12)]の研究では，物体画像の観察時に，IT野の光沢感へ応答するニューロン群付近に電気刺激を与えたり，反対に，抑制性の化学物質投与によりそれらの活動を抑えた場合に光沢感が変化することを報告した．この結果は，ニューロンの活動と光沢感知覚の間の因果関係の存在を示唆している．

5．むすびと今後の展望

本稿では，質感知覚のメカニズム解明を目指す心理物理学や神経科学の研究を振り返り，特に，光沢感や半透明感について，人間の視覚系が質感の知覚に用いる画像特徴や神経基盤をまとめた．これまでの質感研究では，光沢感や半透明感のような，物体の反射・散乱特性のような物性的特性の推定に関連する属性が対象となってきた．しかし，現実世界の質感は多種多様であり，なおかつ，視覚に基づく感覚だけではない．また，質感は物体認識や価値判断のような高次の情報処理にも寄与すると考えられる．多種多様な質感の知覚のために脳がどのような情報処理を行っているのか，そして，質感の知覚がさらに高次の処理にどのように寄与しているのかなど，興味深い未解決課題は多く残る．それらを解決することで，われわれのリッチな質感知覚を支える脳内の情報処理メカニズムの全貌が解明されると期待している．また，科学的な発展のみならず，映像情報メディア向けの新規技術への工学的な発展が期待できる．例えば，質感の知覚に寄与する画像特徴への理解がさらに進めば，任意の物体画像への画像特徴操作のみで所望の質感を表現できる質感編集技術等が生まれることが期待できる．また，質感知覚のための脳内の情報表現や神経基盤の理解が進むことで，より人間らしい判断ができる質感認識システムなどのコンピュータビジョンへの発展も期待できる．引き続き，質感知覚に関連する研究動向に注目したい．

（2025年1月31日受付）

〔文　献〕

1）B.K.P. Horn: "Obtaining shape from shading information", In Shape from shading, pp.123-171 (1989)
2）S. Nishida and M. Shinya: "Use of image-based information in judgments of surface-reflectance properties", JOSA A, **15**, 12, pp.2951-2965 (1998)
3）P. Vangorp, J. Laurijssen and P. Dutré: "The influence of shape on the perception of material reflectance", ACM Transactions on Graphics 26, 3 (Proceedings of ACM SIGGRAPH 2007), **77**, 1-77:9 (2007)
4）I. Motoyoshi, S. Nishida, L. Sharan and E.H. Adelson: "Image statistics and the perception of surface qualities", Nature, **447**, 7141, pp.206-209 (2007)
5）C.B. Wiebel, M. Toscani and K.R. Gegenfurtner: "Statistical correlates of perceived gloss in natural images", Vision research, **115**, pp.175-187 (2015)
6）I. Motoyoshi: "Highlight-shading relationship as a cue for the perception of translucent and transparent materials", Journal of vision, **10**, 9, 6-6 (2010)
7）H. Kiyokawa, T. Nagai, Y. Yamauchi and J. Kim: "The perception of translucency from surface gloss", Vision Research, **205**, 108140 (2023)
8）B.L. Anderson and P.J. Marlow: "Perceiving the shape and material properties of 3D surfaces", Trends in Cognitive Sciences, **27**, 1, 98-110 (2023)
9）P.J. Marlow and B.L. Anderson: "Material properties derived from three-dimensional shape representations", Vision research, **115**, pp.199-208 (2015)
10）C. Hiramatsu, N. Goda and H. Komatsu: "Transformation from image-based to perceptual representation of materials along the human ventral visual pathway", NeuroImage, **57**, 2, pp.482-494 (2011)
11）N. Goda, A. Tachibana, G. Okazawa and H. Komatsu: "Representation of the material properties of objects in the visual cortex of nonhuman primates", Journal of Neuroscience, **34**, 7, 2660-2673 (2014)
12）M. Baba, A. Nishio and H. Komatsu: "Relationship between the activities of gloss-selective neurons in the macaque inferior temporal cortex and the gloss discrimination behavior of the monkey", Cerebral Cortex Communications, **2**, 1, tgab011 (2021)

清川（きよかわ）　宏暁（ひろあき）　2016年，山形大学工学部情報科学科卒業．2021年，同大学大学院理工学研究科電子情報工学専攻博士後期課程修了．その後，東京工業大学特別研究員，産業技術総合研究所特別研究員を経て，2023年より，埼玉大学大学院理工学研究科助教．視覚心理物理や深層学習の技術を用いた質感知覚のメカニズム解明や感情認識の研究に従事．博士（工学）．正会員．

講　座　映像情報×認知科学〔第3回〕

錯　視

杉原厚吉†

1．まえがき

錯視とは，見たものが実際とは違うように知覚される現象である．目の前の状況を誤って判断することは事故の原因ともなるから，錯視の解明は重要な研究課題である．人の視覚データを学習させることによって機械でも錯視が起きることは観察されつつある[1]．しかし，その方法ではなぜ錯視が起きるかという仕組みを知ることは難しい．仕組みを知るためには，学習によって真似をさせるという汎用的な方法ではなくて，情報処理の中身に踏み込んだ個別の議論が必要である．

錯視には多様な種類が知られているが，ここでは仕組みがある程度わかっている三つの典型例を，私自身の錯視作品を使って紹介する．いずれも，脳が，不完全な情報から目の前の状況を認識しようとして頑張った結果の勇み足だということをわかっていただけるであろう．

2．静止画が動いて見える錯視

図1に示す図形を見ると，UFOが列ごとに異なる動きをするように見える．1枚の静止図形が場所ごとに異なる動きをするはずがないから，これは錯視である．この錯視が起きる理由は，動きを検出するそれぞれの脳細胞が網膜上の狭い範囲だけを受け持っていることから説明できる．

私たちがものを見たとき，外から入ってくる光の情報は網膜に投影されて画像となる．この画像は脳に送られ，脳がそれを解釈する．したがって，見たものの動きを認識するのも脳である．

私たちの脳には，視覚情報を処理する膨大な数の細胞があり，それぞれが固有の役目を受け持っている．動きを検出する初期段階の脳細胞は，網膜上のある狭い領域を受け持ち，その中の一つの方向のエッジの動きを監視している．この事実は最初にネコの神経細胞で発見され，その発見はノーベル賞も受賞している[2]．図2に示す円領域の内部の直線Aで示す方向のエッジの動きを監視している脳細胞に着目しよう．今，Aの位置にあったエッジが，Bへ移動したとしよう．このとき，AがBへ移ったことはわかるが，A上の1点が，図の矢印で示すどの方向へ動いたかはわからない．わかるのは，エッジAがそれに垂直な方向へどれだけの距離動いたかだけである．このように局所的な領域

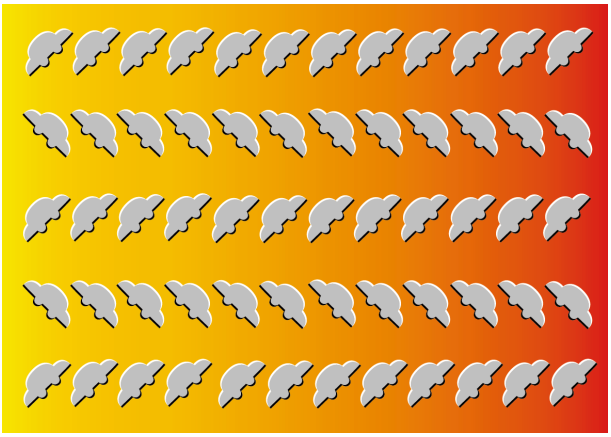


図1　「UFOのラインダンス」(杉原，2013)

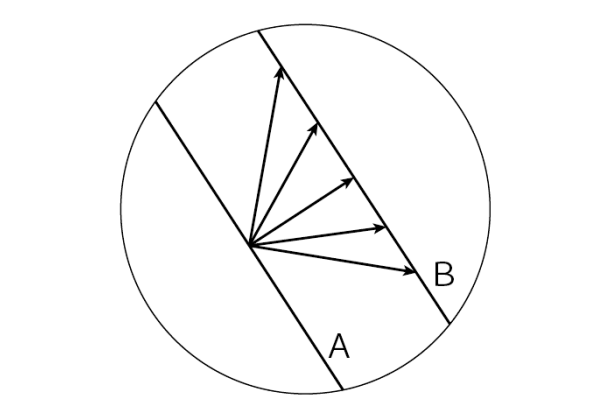


図2　動き検出細胞が検出できる情報の限界

† 明治大学

"Visual Information × Cognitive Science (3): Optical Illusion" by Kokichi Sugihara (Meiji University, Tokyo)

映像情報メディア学会誌 Vol. 79, No. 4, pp. 373～377 (2025)

を監視している動き検出細胞にとっては，受け持った方向のエッジの垂直な方向への動きがわかるだけで，実際にどの方向へ動いたかはわからない．この限界は「窓枠問題」などと呼ばれている[3]．

この動き検出細胞の限界から，図1の動きの錯視が説明できる．物を見るとき，私たちは目で見る方向（視線方向）を動かす．すると，目でとらえた画像も網膜上を動く．そしてその動きが，動き検出細胞でとらえられる．一般には視野の中にいろいろな方向を向いたエッジが混在しているから，それぞれの方向のエッジを検出する細胞が働き，全体として同じ動きが知覚される．

一方，図1に描かれているそれぞれのUFOは，輪郭が白と黒で塗り分けられている．一番上の列のUFOは，白の輪郭も黒の輪郭も右上から左下に伸びていて，向きがそろっている．上から2列目のUFOは，それとは直交する方向に白と黒の輪郭が伸びている．すなわち，主要なエッジ方向が1列目と2列目では直交している．したがって，網膜上でこの画像がランダムに動いたとき，1列目付近を受け持っている細胞と2列目付近を受け持っている細胞では，検出される動きの方向も直交する．その結果，画像が全体として同じ動きをしても，列ごとに異なる動きが検出される．これが，図1の静止画が動いて見える錯視の仕組みである．

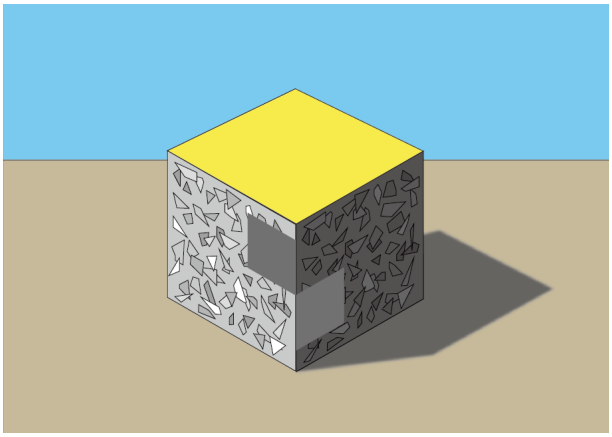
この仕組みがわかると，動いて見える錯視図形を創作する一般指針も得られる[4]．図1は，この指針に基づいて創作したものである．

3．明るさの錯視

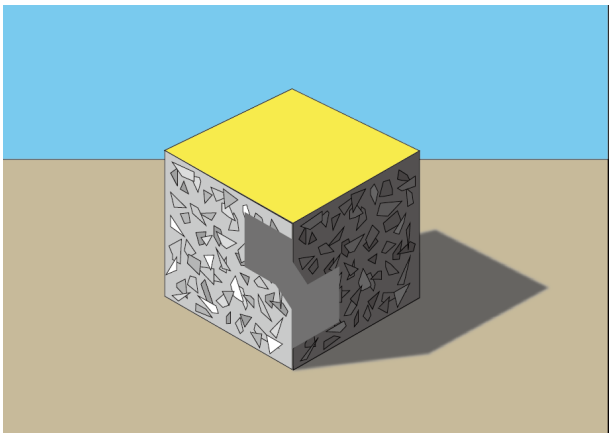
図3（a）に描かれた箱の側面の二つの灰色の四角形は，左の方が右より濃く見える．しかし，二つの灰色は紙面上でまったく同じ濃さである．このことは，図3（b）に示すように，二つの四角形をまたぐ同じ色の帯領域を追加すれば確認できる．

同じ濃さの二つの灰色領域を，明るさの異なる背景で囲むと，明るい背景で囲まれた方が暗い背景で囲まれたものより濃い灰色に見える．この視覚効果は古くから知られており，「明るさの対比」と呼ばれている[5]．ただし，明るさの対比自体は，それほど強い錯視ではない．注目する灰色領域と，それを囲む二つの背景領域の明るさや大きさをうまく調整しないと錯視図形を作ることは難しい．

背景を一様な明るさから，明るさの変化するテクスチャを持ったものに変更すると，明るさの対比が強まることを，ギルクリスト（Gilchrist）等が見つけた[6]．図4に，テクスチャを持った背景に置かれた同じ濃さの灰色図形の例を示す．明るい背景に置かれた左側の灰色の方が，右より濃いと感じるであろう．一様な背景の場合には，このように角の1点で接触する位置に二つの灰色四角形を置くと，ほとんど明るさの対比効果は観察されない．背景にテクスチャ



（a）異なる濃さに見える二つの灰色領域



（b）同じ濃さの灰色であることの確認

図3　錯視絵「ギルクリストの箱」（杉原，2023）

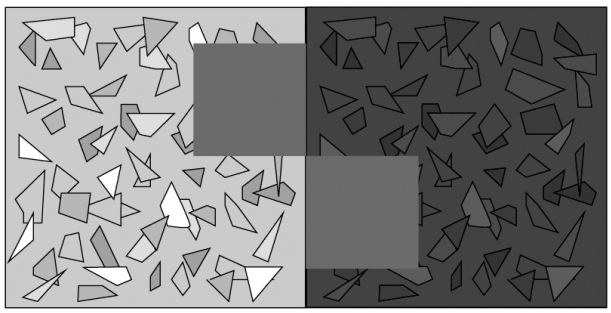


図4　テクスチャを持った背景での明るさの対比

を加えたギルクリスト等の効果がこの図からもわかる．

図4を図3（a）と見比べると，図3（a）の方が明るさの対比効果がさらに強まっていることが観察される．これは，図4が奥行きのない平面図形として描かれているのに対して，図3（a）が奥行きのある立体を表す絵として描かれているためである．すなわち，図3（a）を見た私たちの脳は，左から照明の当たった箱を認識し，箱の左の面には強い照明が当たっているのに対して，右の面にはあまり照明が当

たっていないと解釈する．そして，この照明の強さの違いを考慮して，目に届いた灰色の濃さを補正する．この観察から，図4の明るさの対比も，実は照明の強さの違いを感じた脳が補正した結果だろうと推察できる．

今私たちは，立体に照明が当たった状況を思い浮かべた脳が灰色の濃さを補正することを理解した．ここで改めて，図3(a)の左右の四角形の灰色領域を，「立体の絵という解釈をしないで，純粋に紙面上に塗られた色」とみなして比較したら同じ濃さだと確認できるかどうかを自分自身に問いかけてみよう．同じ濃さには見えないであろう．「立体の絵という解釈を無視する」ことが，私たちの脳にはできないのである．

私たちは奥行きのある3次元の世界に住んでいる．そこでは，手を伸ばしてほしいものをつかんだり，ぶつからないように障害物をよけたりなど，ものの奥行きを知ることが大変重要である．そして，その役目を担っているのが視覚である．外の世界には奥行きがあるのに，それを見て得られる網膜像には奥行きの情報がない．奥行きのないところから奥行きを取り出す（実際には想像する）ことが，視覚の重要課題である．図3(a)のような紙面に描かれた絵も，立体を見たとき直接に得られる網膜像も，脳にとっては同じものであり，見たとたんに元の立体を思い浮かべる任務を遂行しようとするのであろう．立体が描かれているのにその立体を忘れて紙面上の2次元の図形とみなして灰色の濃さを比較せよなどということは，脳にとっては意味のない課題なのである．

同じように，立体が描かれているために，脳が純粋に2次元の図形だと見なして判断することができない視覚現象を「錯視」と呼ぶ例はたくさんある．図5に一例を示す．二つの箱の上面を表す平行四辺形は紙面上でまったく同じ形である．しかも，描かれている葉っぱの向きもそろっている．すなわち，二つの平行四辺形をぴったり一致するように重ねると，葉っぱの向きもそろう．この図からは，左は葉っぱの幅方向に広がった平行四辺形で，右は葉っぱの高さ方向に伸びた平行四辺形に見える．しかし，葉っぱの向きも含めて重ねるとぴったり合うのである．信じられない読者は，是非確かめてみてほしい．この錯視はシェパード錯視と呼ばれる[5]．

色の濃さにしろ，平行四辺形の形にしろ，描かれている奥行き構造から離れて純粋に2次元の図形としての性質を判断することは，私たちの脳にとってはとても難しい．これは，奥行き情報を含まない網膜像から奥行きを推測するという難しい問題を生活の中でいつも課されている脳が，絵を見ると条件反射のように奥行きを思い浮かべてしまうからであろう．

4．平行移動錯視

図6は，立体の後ろに平面鏡を垂直に立てたところである．立体を直接見ると，鏡に向かって飛ぶアゲハ蝶に見える．しかし，普通のアゲハ蝶だったら，鏡に映ったときこちら向きに飛ぶ姿に向きを変えるはずであるが，このアゲハ蝶は振り向かないで同じ姿勢のまま鏡の中へ平行移動しているように見える．これが，平行移動錯視[7]の例である．このアゲハ蝶の振る舞いは，鏡面反射の光学的法則に反しており，あり得ないものに見えるから，この立体は不可能立体に属す．

この錯視が起きるのも，画像には奥行きがないことに起因している．2次元の画像には奥行きの情報が含まれていないから，画像からそこに映っている立体の形を一意的に決めることはできない．同じ姿に見える立体の奥行きには無限の可能性がある．その中には，鏡に映したとき振り向かないアゲハ蝶に見えるものもあるのである．

振り向かないアゲハ蝶の作り方を図7に示す．まず，水平な平面を固定し，互いに逆方向から同じ角度でこの平面を見下ろす二つの視線方向を決める．そして，第1の視線方向から見たとき遠ざかる方向へ飛ぶアゲハ蝶のシルエットと，第2の視線方向から見たとき遠ざかる方向へ飛ぶアゲハ蝶のシルエットを，この平面上に描く．二つのシルエットは，図に赤で示した垂直な面に関して面対称となる．

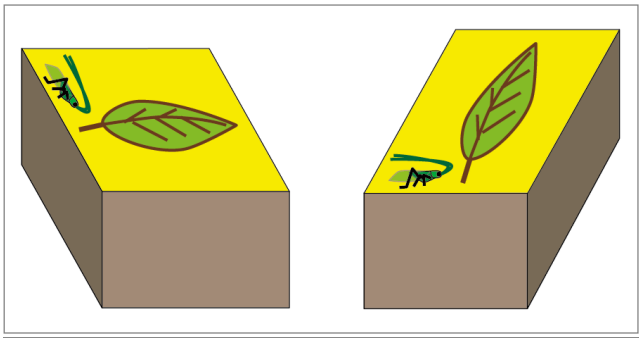


図5　錯視絵「葉っぱの描かれた箱」(杉原，2016)

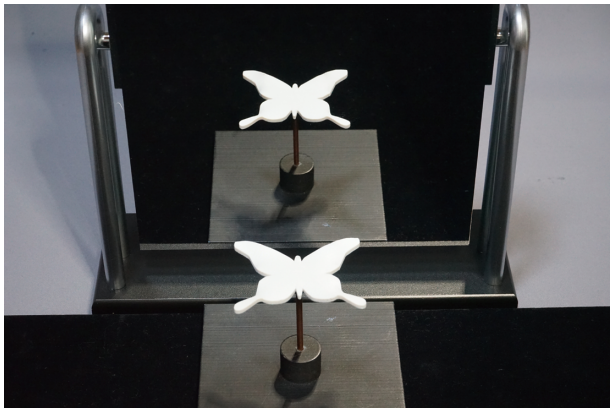


図6　不可能立体「振り向かないアゲハ蝶」(杉原，2022)

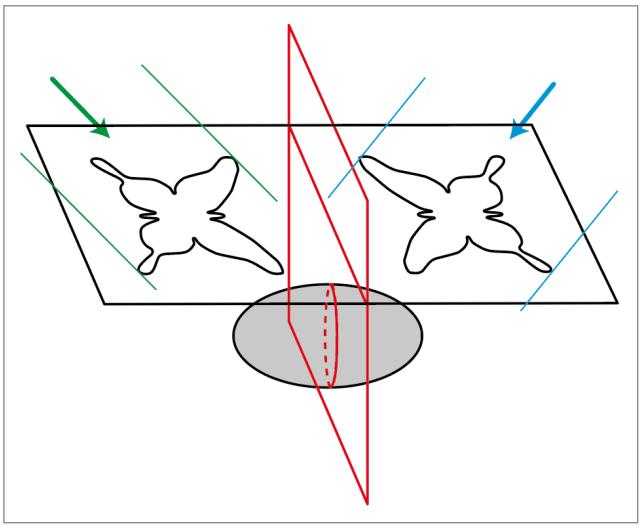


図7　振り向かないアゲハの制作手続き

次に，二つの視線方向からそれぞれ望みのシルエットに見える空間曲線を求める．そのためには，二つのシルエット上の点に1対1対応を定め，対応する点を通ってそれぞれの視線方向に平行な直線の交点を求める計算を繰り返せばよい．詳しくは，この立体の設計法[8]を参照されたい．最後にこの空間曲線の中を石鹸膜のような滑らかな面で張り，一様な厚みをつければよい．この方法で設計計算した結果を，3Dプリンターで作った結果が，図6の立体である．

図7に示した立体設計手続きの全体は，赤で示した垂直な面に関して面対称である．そのため，設計結果である立体自身もこの面に関して面対称となる．このアゲハ蝶を真上から見下ろしたところを図8に示すが，この図からこの立体が面対称であることが確認できる．

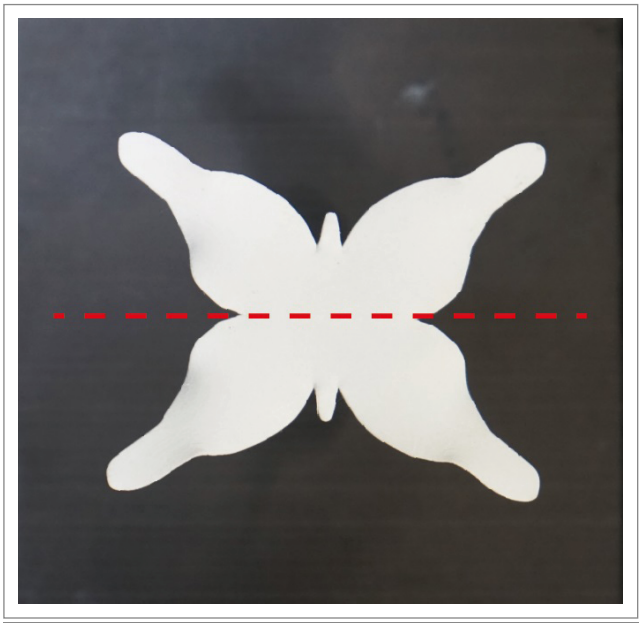


図8　振り向かないアゲハ蝶の面対称性

面対称な立体をその対称面が鏡に平行になるように置くと，立体の鏡像は，その立体を平行移動したものに一致する．したがって，鏡に映った像は，立体の鏡像という解釈と，立体を平行移動したものという解釈の両方が可能である．図6のアゲハ蝶の場合は，斜めの方向から見下ろしたとき面対称であることがわかりにくいため，鏡の中の姿は平行移動したものという解釈だけが知覚される．これが平行移動錯視が起きる仕組みである[7]．

ところで，立体が面対称で，その対称面が鏡に平行であるという性質は，立体と鏡に関かかわるものであって，それを見る視点位置に依存しない．すなわち，どこから見るかにかかわりなく成り立つ性質である．したがって，この立体は，どこから見ても鏡の中へ平行移動しているという知覚が成立する．

図6のシーンを，視点を左へ大きく移動させてみたところを示したのが図9である．このように視点を動かしても，アゲハ蝶が鏡の中へ平行移動しているという知覚は消えない．このように，平行移動錯視は，広い視点範囲で起こることが，その面対称性から保証される[7]．

その姿や振る舞いがあり得ないように見える立体は，「不可能立体」と呼ばれる．不可能立体は，最初は「不可能図形」と呼ばれる絵で表現された．その時点では，絵を見たとき人の脳に思い浮かぶけれど実際には作れない架空の3次元構造という意味であった[9]．オランダの版画家エッシャー(Escher)が作品の中で使ったことでも有名である[10]．

その後，不可能図形と同じに見える立体を，錯視を利用して物理的に作るトリックもいくつか見つかった．特定の視点から見たときつながっているように見えるが実際には不連続な構造で作るトリック[11]，平面のように見えるが実際には曲面で作るトリック[12]，直角のように見えるが実際には直角以外の角度を使って作るトリック[13]などである．しかし，これらのトリックで作った立体は，特定の視点位置から見たときだけ不可能立体に見え，視点を動かしたり

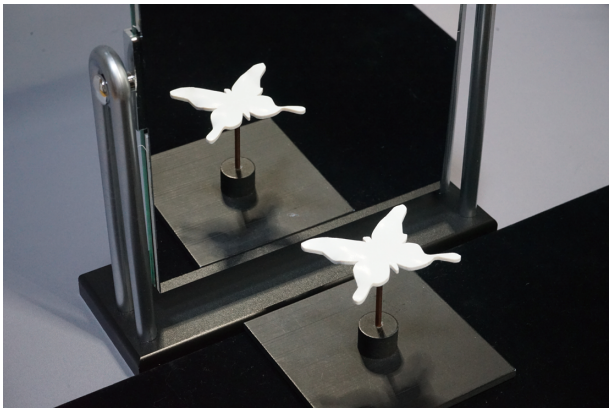


図9　平行移動錯視の視点位置に対する頑健性

両目で見たりすると本当の形がわかって，錯視は消滅する．特定の視点から見るということは，その点を投影中心として2次元の画像へ投影することと等価であるから，立体とはいっても絵の延長にすぎなかった．

それに対して，ここで示した平行移動錯視立体は，図9に示すように視点を大きく動かしても成立し続ける．その意味で，不可能立体が真に絵から飛び出して3次元の立体で実現できるようになったと言ってよいであろう．

視覚芸術の歴史は，新しい視覚効果を開拓する挑戦の歴史であった[14]．遠近法によって見えたとおりに絵を描くことができるようになり，印象派によって見えた姿を忠実に描くのではなくて印象を強調できるようになり，キュビズムによって多数の視点の見え方を1枚の絵に盛り込めるようになったりである．

不可能図形の発明も，あり得ない3次元構造を絵で表現するという視覚芸術の歴史の1ステップといえる．そして，視点位置に強い制限を設けなくてもよい平行移動錯視などの立体錯視は，あり得ない立体を実在する立体で表現できるようになったという意味で，不可能図形を超える視覚芸術の新しい1ステップと言ってよいであろう．すなわち，彫刻分野に新しい芸術表現法を提供できたと思っている．

この認識に基づいて，最近では，錯視を利用した立体アートの創作活動にも力を入れている．図10にその一例を示す．鏡に向かって飛ぶ虫たちが振り向かないで平行移動する作品であるが，理論的に最もきれいな姿が見えるはずの正面ではなくて，このように斜めから見ても安定して錯視が成立する．この方向の活動にも今後力を注ぎたい．

図10 「振り向かない」（杉原，2022，第106回二科展彫刻部入選）

5．むすび

仕組みが比較的明確に説明できる錯視を三つ取り上げ，私自身の作品を例に使って紹介してきた．いずれも，網膜像という2次元情報しか得られない目の構造，局所的な特徴抽出を組み合わせるしかない脳の構造という制約の中で，私たちの脳が生きるために必要な視覚の機能を発揮しようとしてやむを得ず起こってしまう現象であることがわかっていただけたら幸いである．

ここで紹介した研究は，科研費課題番号23K21712および24K22325の援助を受けている．（2024年9月20日受付）

〔文　献〕

1) T. Kobayashi, A. Kitaoka, M. Kosaka, K. Tanaka and E. Watanabe: "Motion illusion-like patterns extracted from photo and art images using predictive deep neural networks", Nature Portfolio Scientific Reports, 12, 3893 (2022)
2) D.H. Hubel and T.N. Wiesel: "Receptive fields of single neurons in the cat's striate cortex", J. Psychology, 148, pp.574-591 (1959)
3) D. Marr: "Vision", W.H. Freeman and Company (1982)
4) K. Sugihara: "A framework for creation of anomalous motion pictures", Art & Perception, 19, pp.59-66 (2021)
5) 北岡明佳：“錯視入門”，朝倉書店（2010）
6) A. Gilchrist et al.: "An anchoring theory of lightness perception", Psychological Review, 106, pp.795-834 (1999)
7) K. Sugihara: "Translation illusion of 3D objects in a mirror", The Journal of the Society for Art and Science, 22, pp.4:1-4:12 (2023)
8) K. Sugihara: "Ambiguous cylinders: A new class of impossible objects", Computer Aided Drafting, Design and manufacturing, 25, 3, pp.19-25 (2015)
9) L.S. Penrose and R. Penrose: "Impossible objects: A special type of visual illusion", British Journal of Psychology, 49, pp.31-33 (1958)
10) M.C. Escher: "M. C. Escher, The Graphic Work", Taschen GmbH, Korn (2005)
11) R.L. Gregory: "The Intelligent Eye", Weidenfeld and Nicholson (1979)
12) G. Elber: "Modeling (seemingly) impossible models", Computers and Graphics, 35, pp.632-638 (2011)
13) K. Sugihara: "Machine Interpretation of Line Drawings", MIT Press (1986)
14) E.H. ゴンブリッチ：“美術の物語”，河出書房新社（2019）

杉原（すぎはら）　厚吉（こうきち）　1971年，東京大学工学部計数工学科卒業．1973年，同大学院修士課程修了．電子技術総合研究所，名古屋大学，東京大学などを経て，現在，明治大学先端数理科学インスティテュート研究特別教授．2010年，作品「なんでも吸引四方向すべり台」が第6回世界錯覚コンテストで優勝したのをはじめ，同コンテスト優勝4回，準優勝2回．主な著書に，『新錯視図鑑』，『見て，知って，つくって！　錯視で遊ぼう』（以上，誠文堂新光社），『立体トリックアート工作キットブック』（金の星社），『鏡で変身！？ ふしぎ立体セット』（東京書籍），『錯覚クイズ』（だいわ文庫）などがある．工学博士．

講　座　映像情報×認知科学〔第4回〕

視覚と画像フィルタ

齋藤　豪†

1．まえがき

眼球の内面の網膜には明所視で働く錐体細胞（cone cell）や暗所視で働く桿体細胞（rod cell）という光受容細胞が並び，それぞれが受光による刺激に応じて反応する．ここからものを見るための視知覚がはじまる．錐体細胞や桿体細胞は，ディジタル画像を撮影をする際の撮像面に並べられた光学センサに対応する．

網膜で受け取った像の刺激は脳の後頭部にある視角野へと伝達されるが，光受容細胞一つ一つの応答の信号が送られているわけではない．その信号は網膜内の層状の神経伝達経路にて加工されて脳へと送られる．

網膜から脳へと信号を伝達するのが網膜神経節細胞（Retinal Ganglion Cell: RGC）である．網膜神経節細胞には平均約100個の光受容細胞からの信号が水平細胞，双極細胞を経て入力される．つまり，平均約100個の光受容細胞の網膜上に分布する範囲が一つの網膜神経節細胞の刺激を受け取る範囲であって，この範囲を網膜神経節細胞の受容野と呼ぶ．網膜神経節細胞は同じ光刺激であっても受容野に対する刺激の位置で応答が変わる．網膜神経節細胞で最も多い種類は受容野の中心部への刺激と周辺部で同じ刺激に対して興奮と抑制という異なる応答を生じさせる中心周辺拮抗型である．中心周辺拮抗型にはON中心型とOFF中心型がある．ON中心型神経節細胞は周辺部より相対的に明るい光が受容野の中心部に照射された時に興奮応答する一方，周辺部に受容野中心部より相対的に明るい光が照射された時には抑制応答する．OFF中心型神経節細胞は逆に周辺部より相対的に暗い光が受容野の中心部に照射された時に興奮応答する一方，周辺部に受容野中心部より暗い光が照射された時には抑制応答する．

この光受容細胞から脳へと伝達される応答の数理モデルにはDifference of Gaussian（DoG）関数が用いられる．

網膜神経節細胞によって眼球から脳へと向かった信号は外側膝状体を経て脳のV1野へ送られるが，そこには網膜像の異方的特徴に反応する細胞があることが知られている．網膜像に対するこの反応はGabor関数でモデル化されることが多い．

さらに多段の神経回路の興奮の伝達により処理される一連の初期視覚系に対するディジタル画像処理の実装形態として畳み込み計算が利用される．

畳み込み計算が画像の局所特徴に応じた値を返すことを数式を見ながら確かめてみよう．

2．内積計算

畳み込み計算の説明の準備として内積計算を復習する．

k次元の任意のベクトル$\boldsymbol{v}=(v_1 \ldots v_k)$と長さ1のもう一つの任意のベクトル$\boldsymbol{e}=(e_1 \ldots e_k)$が与えられた時，内積は式（1）で計算される．

$$v \cdot e = \sum_{i=1}^{k} v_i e_i \tag{1}$$

これは，$\boldsymbol{v}$の$\boldsymbol{e}$への写像であり，$\boldsymbol{v}$と$\boldsymbol{e}$の方位が近ければ大きな値，離れていれば小さな値を取る．そこで内積の値は$\boldsymbol{v}$の$\boldsymbol{e}$らしさを表していると見なすことができる．

3．離散畳み込み計算

個々の光受容細胞が受け取った刺激は複数の神経伝達を経て集約加工され，視野のある一点の光刺激への応答から視野のある局所部分の面刺激への応答へと変化する．この面への刺激の応答は面に対応する視野の部分像の特徴によって強弱する．

ディジタル画像でこれに対応する処理として多用されるのが畳み込み計算である．そして畳み込み計算は内積の応用として考えることができる．

画像$\boldsymbol{I}$を$W \times H$ピクセルの輝度画像とする．検出したい局所特徴に対応する$m \times n$次元の特徴ベクトル$\boldsymbol{F}$を用意し，画像のある$m \times n$の局所部分領域の値を$m \times n$次元ベクトルと考えて内積を取れば，その局所領域の検出したい局所特徴の度合いがスカラー値として得られる．注目局所領域

† 東京科学大学
"Visual Information × Cognitive Science (4): Vision and Digital Image Filter"
by Suguru Saito (Institute of Science Tokyo, Tokyo)

を少しずつずらしてそれぞれの内積値をその局所領域の中央の値とした新たな画像$\boldsymbol{I'}$を作れば，それは元の画像$\boldsymbol{I}$の局所特徴$\boldsymbol{F}$を取り出した画像となる．画像の特徴を抽出することから$\boldsymbol{F}$を画像フィルタと呼ぶ．

処理を式で表すと，式(2)となる．これが離散畳み込みである．要素毎の積の和の計算は内積と同じであること，$\boldsymbol{I}$と$\boldsymbol{F}$との対応付けをkとlで移動させていることがわかる．

$$\boldsymbol{I'}(k,l)=\sum_{i=-m/2}^{m/2}\sum_{j=-n/2}^{n/2}\boldsymbol{I}(k-i,l-j)\boldsymbol{F}(i,j) \tag{2}$$

ただし，$m/2$，$n/2$は，小数点以下を切りすてた商．

4．連続畳み込み計算

網膜像は光受容細胞により離散化された像であるし，ディジタル画像も撮像素子により離散化された画像であることから，前節では離散畳み込みを説明した．本節では畳み込み計算をもう少し踏み込んで考えるために連続関数の畳み込み計算について見てみよう．

式(3)は，離散畳み込み計算の式(2)に対応する連続畳み込み計算の式である．

$$I'(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}I(x-s,y-t)F(s,t)dsdt \tag{3}$$

積分区間が$-\infty$から∞となっているが，$F(x,y)$は値が非ゼロとなる変域が狭く，実際の積分区間は$F(x,y)$の非ゼロとなる変域であることと等しい．

上の二次元の畳み込みでは二重の積分があるが，これからの説明では式の記述が複雑になるため，以後では一次元連続関数の畳み込み計算を扱う．

改めて，連続関数$I(x)$に$F(x)$を畳み込むことで得られる関数$(I*F)(x)$は式(4)で定義される．

$$(I*F)(x)=\int_{-\infty}^{\infty}I(x-t)F(t)dt \tag{4}$$

ここで，連続畳み込み計算に関する四つの性質を確認する．

畳み込み計算の交換則

$$\boxed{(I*F)(x)=(F*I)(x)}$$

$$\begin{aligned}(I*F)(x)&=\int_{-\infty}^{\infty}I(x-t)F(t)dt\\ &\quad u\equiv x-t\text{と置く}\\ &=\int_{\infty}^{-\infty}I(u)F(x-u)-du\\ &=\int_{-\infty}^{\infty}F(x-u)I(u)du\\ &=(F*I)(x)\end{aligned} \tag{5}$$

畳み込み計算の結合則

$$\boxed{((I*F)*G)(x)=(I*(F*G))(x)}$$

$$\begin{aligned}((I*F)*G)(x)&=\int_{-\infty}^{\infty}(I*F)(x-t)G(t)dt\\ &=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}I(x-t-u)F(u)duG(t)dt\\ &=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}I(x-t-u)F(u)G(t)dudt\\ &\quad v\equiv t+u\text{と置く}\\ &=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}I(x-v)F(u)G(v-u)dudv\\ &=\int_{-\infty}^{\infty}I(x-v)\int_{-\infty}^{\infty}F(u)G(v-u)dudv\\ &=\int_{-\infty}^{\infty}I(x-v)(F*G)(v)dv\\ &=(I*(F*G))(x)\end{aligned} \tag{6}$$

畳み込み計算の分配則

$$\boxed{(I*(F+G))(x)=(I*F)(x)+(I*G)(x)}$$

$$\begin{aligned}(I*(F+G))(x)&=\int_{-\infty}^{\infty}I(x-t)F(t)dt\\ &\quad+\int_{-\infty}^{\infty}I(x-t)G(t)dt\\ &=(I*F)(x)+(I*G)(x)\end{aligned} \tag{7}$$

畳み込み計算と微分

$$\boxed{\frac{d}{dx}(I*F)(x)=(\frac{d}{dx}I*F)(x)=(I*\frac{d}{dx}F)(x)}$$

$$\begin{aligned}\frac{d}{dx}(I*F)(x)&=\frac{d}{dx}\int_{-\infty}^{\infty}I(x-t)F(t)dt\\ &=\lim_{\delta\to0}\frac{\int_{-\infty}^{\infty}I(x+\delta-t)F(t)dt-\int_{-\infty}^{\infty}I(x-t)F(t)dt}{\delta}\\ &=\int_{-\infty}^{\infty}\lim_{\delta\to0}\frac{I(x+\delta-t)-I(x-t)}{\delta}F(t)dt\\ &=\int_{-\infty}^{\infty}\frac{d}{dx}I(x-t)F(t)dy\\ &=(\frac{d}{dx}I*F)(x)\end{aligned} \tag{8}$$

$$\begin{aligned}\frac{d}{dx}(I*F)(x)&=\frac{d}{dx}(F*I)(x)\quad\text{交換則}\\ &=(\frac{d}{dx}F*I)(x)\quad\text{式(8)より}\\ &=(I*\frac{d}{dx}F)(x)\quad\text{交換則}\end{aligned} \tag{9}$$

四つの性質を確認した．次節では，これらを踏まえた上で先に述べた視覚の初期知覚を模倣するDifference of Gaussian関数，Gabor関数の実装や性質を見ていこう．

5．DoGフィルタ

網膜神経節細胞の受容野応答の数理モデルには二つのGauss関数の差，Difference of Gaussian（DoG）関数（式(10)）が使われる．

$$DoG(x,y)=a_1e^{-(x^2+y^2)/\sigma_1^2}-a_2e^{-(x^2+y^2)/\sigma_2^2} \tag{10}$$

ON中心型の網膜神経節細胞のモデルは，$\sigma_1 < \sigma_2$とする．また，入力像が均一のとき出力が0となるように，第一項と第二項の積分値が同じになるように係数は決めておく．なお，二次元ガウス関数の積分値は式(11)のとおりである．

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ae^{-\frac{(x^2+y^2)}{\sigma^2}}dxdy &= a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2}{\sigma^2}}e^{-\frac{y^2}{\sigma^2}}dxdy \\ &= a\int_{-\infty}^{\infty} e^{-\frac{x^2}{\sigma^2}}dx\int_{-\infty}^{\infty} e^{-\frac{y^2}{\sigma^2}}dy \\ &= a\sigma\sqrt{\pi}\,\sigma\sqrt{\pi} = a\sigma^2\pi\end{aligned} \tag{11}$$

DoG関数は，最も反応する空間周波数が一つある緩いバンドパス型のフィルタを作り出す．網膜神経節細胞により網膜から出発した信号は外側膝状体を経て脳の視覚野に到達する．そこには空間周波数の帯域別に反応する仕組みがある．

ディジタル画像処理では，複数の周波数帯の画像特徴を分離して取り出すために複数のDoG関数による畳み込み計算が用いられる．図1に三つの分散の異なる二次元DoG関数の例を示す．暗いほど低い値，白いほど高い値を表し，負の値も取るため灰色が0の値に相当する（図2，図4，図5の濃淡表現も同様である）．また，Gauss関数の組み合わせ方の違いでDoG関数は特に強く反応する空間周波数帯が変わる．

DoG関数の畳み込み計算には，計算量を削減する実装が採用されることが多い．それがラプラシアンピラミッドである．ラプラシアンピラミッドでは，まずGauss関数の畳込みを多段で行い，その各段階の畳み込み計算結果を保存する．

式(12)のように，Gauss関数$G_1(x)$とGauss関数$G_2(x)$の畳み込みによる結果はGauss関数になる．

$$\begin{aligned}(G_1 * G_2)(x) &= \int_{-\infty}^{\infty} a_1 e^{-\frac{(x-y)^2}{\sigma_1^2}} a_2 e^{-\frac{y^2}{\sigma_2^2}}dy \\ &= a_1a_2\int_{-\infty}^{\infty} e^{-\frac{1}{\sigma_1^2\sigma_2^2}(\sigma_1^2y^2+\sigma_2^2(x-y)^2)}dy \\ &\quad \sigma_1^2y^2+\sigma_2^2(x-y)^2 \\ &= (\sigma_1^2+\sigma_2^2)y^2-2\sigma_2^2xy+\sigma_2^2y^2 \\ &= (\sigma_1^2+\sigma_2^2)\left(y-\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}x\right)^2+\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}x^2 \\ &= a_1a_2e^{-\frac{x^2}{\sigma_1^2+\sigma_2^2}}\int_{-\infty}^{\infty} e^{-\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}\left(y-\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}x\right)^2}dy \\ &\quad z \equiv y-\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}x \text{と置く} \\ &= a_1a_2e^{-\frac{x^2}{\sigma_1^2+\sigma_2^2}}\int_{-\infty}^{\infty} e^{-\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}z^2}dz \\ &= a_1a_2\sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\pi}e^{-\frac{x^2}{\sigma_1^2+\sigma_2^2}}\end{aligned} \tag{12}$$

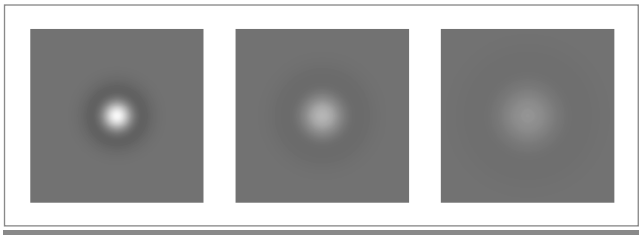


図1　二次元DoGフィルタ

よって，畳み込み計算の結合則より，多重にGauss関数を畳み込むことは，分散が大きくなったGauss関数を一回畳み込むことに等しい．またGauss関数を畳み込むことは局所的な重心偏重加重平均を計算していることになるので，画像をぼかす効果があり，低周波通過型フィルタの特徴を持つ．そのため空間高周波情報が取り除かれた畳み込み計算結果の画像はサンプリング密度を下げることができる．これらにより，広い受容野に相当する離散畳み込み計算中の積和計算を効率的に削減している．こうして作られるGauss関数の多段畳み込み計算結果をガウシアンピラミッドと呼ぶ．また，段階間の差分画像がDoGを畳み込んだ結果と等しくなる．この画像はラプラシアン画像と呼ばれる．

二次元のラプラシアン演算は，$\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}$であり，上記の計算に二階偏微分は出てこないのでラプラシアン画像と呼ぶことに疑問であろうが，実際Gauss関数にラプラシアン演算を施した関数とDoG関数は完全一致ではないが形状が似ており，それらを使った画像への畳込み計算結果も類似する．なお，畳み込みの性質の**畳み込み計算と微分**でのとおり，Gauss関数にラプラシアン演算を施した関数を画像に畳み込むことは，画像にラプラシアン演算を施してからGauss関数を畳み込むことと等しい．

6．Gaborフィルタ

DoG関数には異方性がなく，模様の傾きを検出することはできない．脳の視覚野のV1野には，網膜像の異方的特徴に反応する細胞があることが知られている．網膜像に対するこの反応は三角関数とGauss関数の積で作られるGabor関数でモデル化される．そのフーリエ変換は三角関数の周期に対応する周波数を中心とするGauss関数で広がった形をしており，周波数空間で信号処理を理解する上で便利である．

Gabor関数$Gb(x,y)$はcos型とsin型があり，式(13)で与えられる．θは縞の方向を決める角度である．

$$Gb(x,y) = \begin{cases} ae^{-\frac{(x^2+y^2)}{\sigma^2}}\cos(b(x\cos\theta+y\sin\theta)) & \text{cos型} \\ ae^{-\frac{(x^2+y^2)}{\sigma^2}}\sin(b(x\cos\theta+y\sin\theta)) & \text{sin型} \end{cases} \tag{13}$$

網膜神経節細胞の数理モデルのDoG関数と異方性のGauss関数を畳み込むことで，図2のようにcos型Gabor関数に類似した関数を作り出せる．sin型Gabor関数に類似し

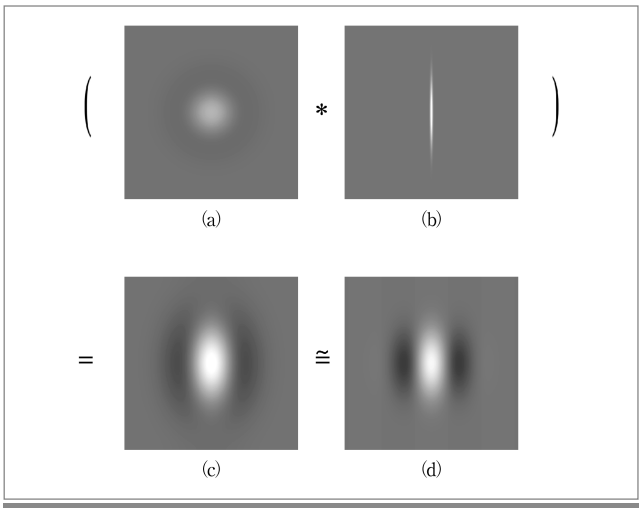


図2　DoG関数（a）と細長い異方性ガウス関数（b）の畳み込み計算（c）とGabor関数（d）

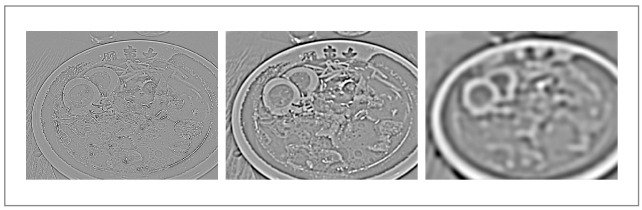


図4　異なるDoG関数との畳み込み計算結果

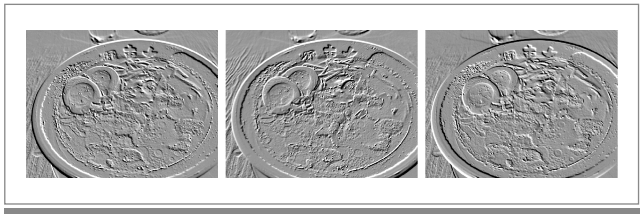


図5　異なるGabor関数との畳み込み計算結果

た関数は一つのDoG関数の隣に符号反転させた同関数を並べたものを図2（a）の代わりとすることで作り出せる．

V1野中の複雑細胞は異なる位相の模様に対して安定した応答をするが，その応答の数理モデルにはsin型Gabor関数とcos型Gabor関数の畳み込み計算のそれぞれの二乗の根が用いられる．なお，sin型とcos型のGabor関数には，位相の違いだけではなく，sin型は積分値が0であるがcos型は0ではないという違いもある．これはcos型だけは局所的な平均値に依存した応答となることを意味するので，場合によっては注意が必要である．

7．画像への適用例

図3に広がりの異なるDoG関数を畳み込んだ結果が図4である．DoG関数はラプラシアンと近い特徴を持つと述べたが，画像の輪郭線の位置は輝度勾配の変曲点となることが多く，応答の符号が反転するところは輪郭であることが多い．さらに広がりの異なるDoG関数を畳み込んだ結果で変曲点の位置が変化しないところは代表的な輪郭線であることが多い．

図5はGabor関数の方向選択応答性を示している．テーブルの木目の部分で模様に沿った方向のGabor関数による応答が他の方向のそれより強く出ていることがわかる．

図3　対象画像

8．むすび

本稿では，視覚の初期知覚の数理モデルについて，線形畳み込みフィルタで説明できるごく単純な範囲を説明した．非線形の応答の合成やさらに多層化して複雑な模様への反応に特化した関数も取り入れるCNNの理解のきっかけにもなってもらえば幸いである．網膜上の刺激の比較的単純な伝達処理でも触れなかったことはまだあり，明所視で活躍する光受容細胞の三種類の錐体からの入力の組み合わせにより網膜神経節細胞の段階で反対色コントラストの信号が生じること，また，周辺抑制がない網膜神経節細胞があること，光受容細胞であることが近年明らかとなった神経節細胞の一種のipRGCの応答について等々あるので，それらの数理モデルへも興味を拡げてもらいたい．

（2025年7月3日受付）

齋藤（さいとう）　豪（すぐる）　1999年，東京工業大学大学院情報理工学研究科博士課程修了．同大学助手，同大学准教授，同大学情報理工学院准教授，お茶の水女子大学大学院准教授を経て，現在，東京科学大学情報理工学院准教授．視覚特性の工学応用，アニメーションの技法支援法，画像加工処理に関する研究に従事．博士（工学）．

講　座　映像情報×認知科学〔第5回〕

マルチメディアデータから捉える言語発達

佐治伸郎†

1．言語習得研究におけるデータの問題：〜日誌を超えて〜

心理学は基本的にデータの不足を前提とする学問である．心理学はそもそも人間の心という直接観察が不可能な対象を研究対象する学問領域であり，それ故にその推定を可能とする人の行動データを量・質ともにどれだけ収集できるかが学問的生命線であった．実験心理学者は実験室に被験者を集め，巧みにデザインされた実験により行動のパターンを明らかにしようとしてきた．質的心理学者は研究者自身がフィールドに入り，人が社会的環境との関わりの中でどのような行動を行うのかそのパターンを観察することでこれを明らかにしようとしてきた．20世紀前半に確立した推測統計学は，限られたデータから背後に存在する心のはたらきの規則性を推測するのに大きな役割を果たした．

発達心理学は，心理学の中でも殊さらにデータの不足問題と対峙してきた学問分野である．子どもは，大人のように自ら実験室に来てはくれないし，子どもが生活する場である家庭は，部外者である研究者が最もフィールドとしづらいプライベートな空間の一つである．そこで初期の発達心理学の研究は，研究者にとって身近な子どもを対象とした観察研究の形をとり，探索的に理論構築へと繋げようとする研究が多かった．例えば，発達心理学の祖とも言われるジャン・ピアジェによって打ち立てられた初期の多くの理論が，自分の3人の子どもを対象とした縦断観察データに基づいていることはよく知られている．また，著者の専門領域である言語習得研究においても，特に20世紀中盤においては，限られた数の子どもを対象にした観察研究をベースに多くの理論構築が行われてきた[1)]．後の研究者に日誌研究の時代（the period of diary studies: Khodareza et al., 2015）とも言われるこの時代の観察データは，非常に多くの重要な理論的構築に貢献し，後に行動実験が発達心理学の主要な手法となった後も，検討すべき理論的予測を生み出す役割を担ったのである．

† 早稲田大学 人間科学学術院
"Visual Information × Cognitive Science (5): Exploring Language Development through Multimedia Data" by Noburo Saji (School of Human Sciences, Waseda University, Tokyo)

2．発達心理学・言語習得研究分野におけるマルチメディアデータの利用

日誌研究は多くの理論的貢献を成し遂げたが，研究者による書記による記録だけではなく，音声・映像など，よりリッチな情報を含むマルチメディアデータを求める声も日に日に大きくなっていった．子どもが話す言語はそもそも音声であるために，音声言語の習得過程を調査するには日誌における書き起こしデータだけではなく，実際の音声データを得る必要がある．また言語を用いて何を参照しているのかという意味の問題や，どのような社会的なやりとりをしているのかという語用の問題を調査しようと思えば，それが発せられた場面を記録した映像データも必要である．本章では20世紀後半から現在に至るまでの発達心理学におけるマルチメディアデータの利用について概観する．

2.1　CHILDES

日誌研究は，研究者の実子など研究者にとって身近な子どもを対象としたものが多く，それぞれの研究室の「引き出し」の中のデータであった．したがって，外部の研究者には直接検討がしづらいという問題点を抱えていた．より一般的な理論を構築するためには，このようなデータを「引き出し」から出して，多くの研究者で共有する必要があった．

さらに研究史における理論的な背景を挙げれば，20世紀中盤のチョムスキーの生成文法理論の隆盛以降，言語習得研究の主たる関心は理論的に想定される生得的言語知識にあった．そのような中，言語的知識の構築がどの程度，経験（すなわち子どもに与えられる言語的インプット）から説明できるのかという問題は，このような理論に対する反証材料として強い関心を集めていたのである．

上記の方法論的・理論的関心から，言語習得研究者にとってマルチメディアデータへの関心は非常に高いものであったと言える．1984年に始まったCHIDES（Child Language Data Exchange System: https://talkbank.org/childes/）プロジェク

トはこのような関心を踏まえての，最も大きな試みの一つであったと言えるだろう[2]．CHIDLESプロジェクトは発達心理学者であるブライアン・マックウィニーとキャサリン・スノウによって開発され，それまで研究者が個別に保持していた子どもの発達過程を記録した映像データ，音声データ，発話書き起こしデータを統一した規格で保存，共有することを可能にしたのである．CHILDESのデータを用いたコーパス研究は枚挙に暇がないが，特に言語知識の獲得において経験的要素を重要視する立場の研究者たちの重要なデータとなった．例えばLievenら[3]はCHILDES収録データにおける1歳から3歳の子どもの発話を分析し，子どもの初期の文における語の組み合わせが抽象的な文法規則に基づき自在に組み合わされる訳ではなく，特定の語同士の定型的な組み合わせによってなされることを明らかにし，生得的な統語的知識の存在について疑問を投げかけた．

現在から見ればオープンサイエンスの走りとも言えるこのCHLDESプロジェクトは，言語的知識を支える経験的要因の検討，さらに言語習得の多言語比較研究を強力に押し進めた．現在でも30以上の個別言語のデータが収録され，発達心理学，言語学分野の多くの研究において利用されている．

2.2　Human Speechomeプロジェクトと Dense corpus

2000年代，記録媒体の大容量化，撮影機器の小型化・高精度化により，映像や音声を含むマルチメディアデータを用いた研究が認知科学を席巻した．言語習得分野もその恩恵を受けた分野の一つである．それまでのCHILDESに収録された映像データは発達心理学・言語習得分野におけるデータ不足の問題に大きく貢献したが，集められるデータが「まばら」(sparse)になりやすいという問題があった．すなわち，子どもの身体的・心理的発達は数日のうちに大きく変化するという非連続的な特徴があるのに対し，研究者が家庭に撮影に出向く頻度はどうしても数日おき，数週おきの「まばら」な撮影になってしまう．すると，発達の微視的変化を緻密に捉えることが難しいという問題が生じるのである．この問題に対し撮影機器や記憶媒体の技術発展は，「高密度」(dense)の映像記録データを提供することで応えようとしたのである．

この問題に対する回答として，言語習得分野において最も象徴的であったのはMITのデブ・ロイらのグループによるHuman Speechomeプロジェクトだろう[4)5)]．Human Speechomeプロジェクトでは，子どもが生活する家屋の各部屋の天井に小型カメラを設置し，生後2年間にわたって映像・音声を撮影し続けた．結果として得られたマルチメディアデータは80,000時間に及ぶ膨大な容量かつ高密度なものになり，その後，さまざまなデータ分析の対象となった．一例を挙げると，このデータを用いて，子どもの語の産出を予測する要因を日常空間から探った研究では，養育者から話される語の特異性（時間的に特異的に使われるか，空間的に特異的に使われるか，一緒に話される語との関係が特異的かの3点を統合した指標）が強力な予測用要因となることを報告したのである[4]．Human Speechomeプロジェクトのデータ収集方法はプライバシーや協力者側の負担の問題から，多くの研究者にとって手軽に用いることができるとは言い難いものであったが，子どもが日常生活に最も近い経験データをこれまでにはない規模で収集可能であることを示した点で画期的であった．

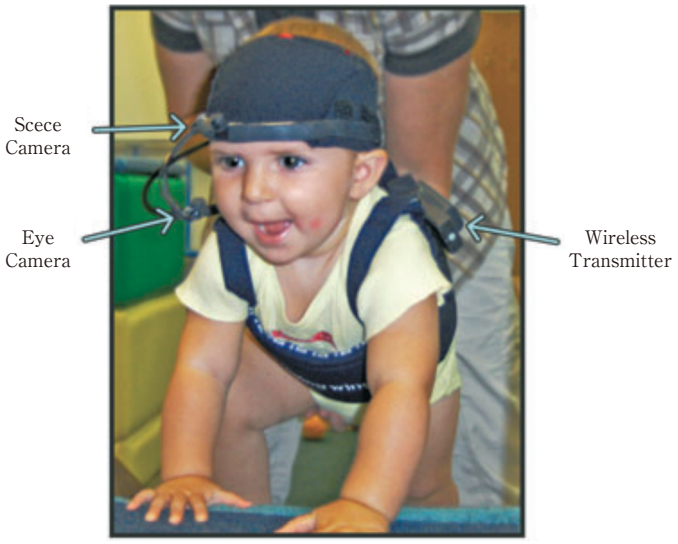
このように，2000年代は「まばら」ではなく高密度に発達過程を記録するマルチメディアデータを利用する研究への機運も生まれ，多くの高密度コーパスの記録および研究がなされた[6]．一方で，これまで研究者が経験したことのないマルチメディアデータ，ビッグ・データの蓄積は，このデータをどのように処理すれば何がわかるのかという分析手法の問題を顕在化させたとも言えるだろう．

2.3　ヘッドマウントカメラを用いた一人称視点研究

2000年代後半から2010年代にかけて，発達研究に大きな影響を与えたもう一つの技術革新は撮影機器の小型化である．カメラの小型化は，発達研究において新たなマルチメディアデータの利用の可能性を開いた．それは，子どもが日常生活の中で何をどのように見ているのかを一人称的(egocentric)視点から探ろうという研究プロジェクトである[7)8)]．

それまでの研究で用いられていたマルチメディアデータは，CHILDESにしても，Human Speechomeプロジェクトにしても，子どもの様子を三人称視点から場面を撮影，その行動を記録したものであった．しかし，子ども自身が世界をどのように見ているかというデータは子どもの心の発達過程を探るのに決定的に重要である．というのも，子どもの場合，身体のサイズや運動能力の短期間での変化が大きく，発達の過程に応じて一人称視点からの「見え」が大きく異なる．例えば生後0～3ヵ月の子どもにとって，視覚的に確認できる世界のは「天井」や時折自分を覗きこんでくる「親の顔」であろう．しかし，3ヵ月以降に首が座れば上腕の運動が可能になり視界には「自分の手」が加わるだろうし，縦向きに抱っこされるようになれば，より高い目線から「天井」以外の奥行きのある世界を眺められるようになるだろう．このように子どもの一人称的視点からの「見え」は，生後1年の間に非常にダイナミックに変化し，それは少なからず心理・言語の発達に影響を与えていることが示唆されてきた．撮影機器の小型化は，この子どもにとっての「見え」の変化を記録可能にし，子どもが日常においてどのような対象に視点を向けているのか，それがどのように変化するのかを推定することを可能にした．図1に文献[7)]で用いられたこのデバイスの一つの典型的な装着事例を示す．この装置では子どもの頭部にシーン撮影用，および視線検出用のカメラを装着し，子どもが見ている世界をリアルタイム記録することを可能にしている．

Franchakら[7)]は，生後12ヵ月の乳児にこのようなヘッ



(a)　(b)

図1　文献[7]にて用いられたヘッドマウントカメラ
（引用：Di Martino et al., BMC Microbiology (2011), licensed under CC by 2.0.）

ドマウントカメラを装着し，乳児がどのような状況において，外界の視覚的探索を行っているのかを多方面から検討した（図1）．この結果，乳児の視覚的探索は外部からの働きかけというよりは乳児が置かれた状況に依存することを報告している．例えば，乳児が養育者の顔へと視線を向けるのは，養育者の発話に対してではなく，養育者が乳児の目の高さで座っている状況において頻繁にみられた．言語習得研究に関しても，ヘッドマウントカメラを用いた一人称視点研究は新たな研究の展開を生み出した．例えばYuら[9]の研究では，日常的な遊びの中で子どもがモノをどのように持ち，またそれを見ている時に養育者が語りかけることが，語の習得を促進するのかが検討された．この研究では，やはりヘッドマウントカメラを装着した養育者と1歳半の子どもが新奇の名称をラベリングされたおもちゃで相互に遊ぶ場面を設定した．一定時間，両者におもちゃで遊んでもらった後，子どもがどれくらいおもちゃの名前を学んでいるかがテストされた．分析では命名の成功と，遊びの中で子どもがそのおもちゃをどのように見ていたかの関係が分析された．すると，おもちゃが子どもの視界の一定の割合を，安定して占有しているタイミングで養育者が名付けしている場合，命名が成功しやすいことがわかったのである．

このように2010年代には，発達心理学のさまざまな領域においてヘッドマウントカメラを用いた一人称視点マルチメディアデータが採用されたが，さらに近年では，一人称視点のマルチメディアデータについても大規模データベース化が進んでいる[10]．一般的に，人間の子どもは機械学習と比べても非常に少ないデータからパターン発見，一般化を行うヒューリスティクスを持つことはよく知られている．ロングらのBaby view datasetプロジェクトでは，この理由の一つが子どもの一人称的視点の中に潜んでいることを想定し，一般家庭で導入可能なヘッドマウントカメラのキットを開発，これを配布することにより一人称視点のマルチメディアデータを収集している．このプロジェクトでは現在までに6ヵ月から5歳までの子どもの一人称視点からの映像を収めた493時間のデータベースが作成され，深層学習や言語習得領域を含むさまざまな研究者との共同研究が進められている．

3．これからの発達・言語研究におけるマルチメディア・データの利用と今後の課題

ここまで，発達心理学・言語習得分野においてマルチメディアデータの高密度化，大規模データベース化，子どもにとっての多様な「見え」の分析が進んでいることを論じてきた．本稿の最後に，現状の問題と今後の展望について論じてみたい．

3.1　撮影手法の簡素化

第一に，利用可能はマルチメディアデータの蓄積はbaby view datasetプロジェクトに見られるように，汎用性のある録画機器キットを協力者に配布することによって大きな進展を見せている．しかし，育児中の保護者にとって，ヘルメット型の機器であるカメラを装着したまま家庭内で育児を進めることは簡単なことではなく，未だ撮影の敷居は低いとは言えないだろう．また，家庭の外，例えば，幼稚園や保育園などで幼児の集団保育場面における子どもの一人称視点データを集め，子どもたちの対人，対物のやり取

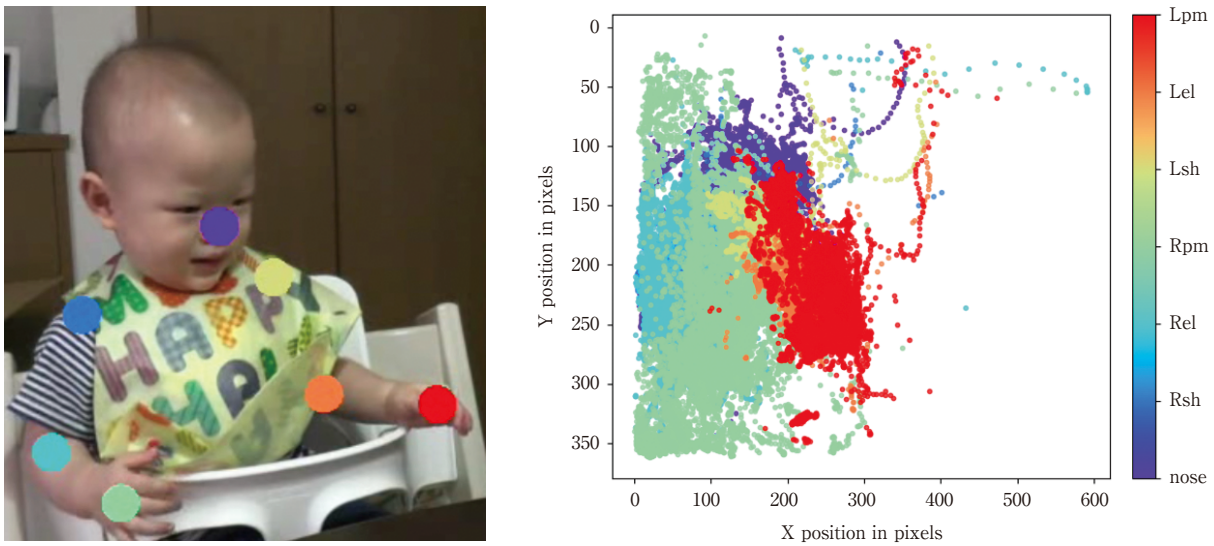


図2　幼児の食事場面におけるマーカレスモーションキャプチャの例

りをデータとする研究を行う場合には，カメラが保育の障害になるのを避けなければならず，更なるカメラの小型化，着脱のしやすさの向上，および事故防止策を講じなければならないだろう．乗り越えるべき技術的な問題はまだ大きいが，もしこのようなデータが取得できるとすれば，発達心理学分野における観察研究は非常に大きな前進を見せることが期待される．

3.2．データ分析手法の確立

マルチメディアデータの大規模化に伴い浮上する第二の問題は，取得したデータをどのように分析するかという問題である．従来，映像を含むマルチメディアデータの分析は，人手によるアノテーションを基にした分析が主流であった．しかし，今後増え続ける大規模データセットに対して手動のアノテーションを行うことは現実的に非常に難しいことが予想される．近年このような問題に対して，大人や生物一般のデータに関してはOpen pose[11]のような自動姿勢推定システムや，DeeplabCut[12]のようなマーカレスモーションキャプチャを用いた分析のが盛んに行われている．このような分析はもちろん子どもに対しても有効であると考えられる．例えば図2は，筆者の記録した子どもの食事場面における映像データに対し，頭部と右手と左手の動きを計測するために腕関節部マーカをDeeplabcutで推定，その座標の分布を表したものである．このデータからは，右手は食器と頭部の間の移動に集中しているのに対し，左手は大きなばらつきを持って動いているのがわかる（実際には左手は指さしなど食事以外の対象へと関心を他者に向けさせるような行動が多く見られた）．マーカレスモーションキャプチャによりこのような分析を，多くのデータに対して実施できるのは，非常に重要な技術的進展である．ただし子どもにこのような自動推定を用いることの問題は未だ多い．例えば，運動量の多い子どもの動きを捉えるために精度の高い奥行きの推定，物体や人との重なりの処理が不可欠であるが，やはり現状ではその精度は著しく落ちる．また多くの姿勢推定システムでは大人のデータをベースにモデルの学習を行っているために乳児や子どもの姿勢推定には問題があることなども指摘されている[13]．

3.3　多様なデバイスとの協働

さらにマルチメディアデータと，さまざまなウェアラブルデバイスとの連携は発達心理学・言語習得研究に大きなブレイクスルーを引き起こすと考えられる．例えば子どもに装着することで子どもの言語環境，すなわち周囲からどのような言語的入力を得て，また子ども自身が産出を行っているのかを測定・解析するLENAシステムは，子どもを取り巻く音声データの利用可能性を大きく広げた[14]．映像を含むマルチメディアデータとこのような言語解析システムの共同は，子どもが日常的な生活の中で見ているもの，聞いているものをこれまでにない精度で記録し，言語習得研究の領域に非常に重要な知見を提供できるだろう．

4．むすび

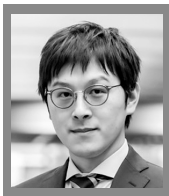
本論では，発達心理学・言語習得研究においてマルチメディアデータ利用の経緯を概観し，現在に至るまでどのような関心・課題が存在しているかを論じた．撮影機器や映像や音声の分析を巡る技術革新は，これまでも発達心理学・言語習得研究に新しい地平を提供してきたし，またこれからも大きなブレイクスルーを起こし続ける非常に大きな可能性を秘めた領域である．発達心理学・言語習得研究者と，撮影機器，マルチメディアデータ，映像情報を扱う研究者とがより強く

連携した研究が一層進むことを強く願う．

（2025年7月29日受付）

〔文　献〕

1) E.V. Clark: "WHAT's in a WORD? on the CHILD's ACQUISITION of SEMANTICS in HIS FIRST LANGUAGE", Cognitive Development and Acquisition of Language, pp.65-110 (1973), https://doi.org/10.1016/b978-0-12-505850-6.50009-8
2) B. MacWhinney: "The CHILDES Project: Tools for analyzing talk: the database", Lawrence Erlbaum Associates Publishers, 2, 3 (2000), https://psycnet.apa.org/record/2000-03631-000
3) E.V. Lieven, J.M. Pine and G. Baldwin: "Lexically-based learning and early grammatical development", Journal of Child Language, 24, 1, pp.187-219 (1997), https://doi.org/10.1017/s0305000996002930
4) B.C. Roy, M.C. Frank, P. DeCamp, M. Miller and D. Roy: "Predicting the birth of a spoken word", Proceedings of the National Academy of Sciences of the United States of America, 112, 41, pp.12663-12668 (2015), https://doi.org/10.1073/pnas.1419773112
5) D. Roy, R. Patel, P. DeCamp, R. Kubat, M. Fleischman, B. Roy, N. Mavridis, S. Tellex, A. Salata, J. Guinness, M. Levit and P. Gorniak: "The Human Speechome Project", Symbol Grounding and Beyond, pp.192-196 (2006), https://doi.org/10.1007/11880172_15
6) R.J.C. Maslen, A.L. Theakston, E.V.M. Lieven and M. Tomasello: "A dense corpus study of past tense and plural overregularization in English", Journal of Speech, Language and Hearing Research: JSLHR, 47, 6, pp.1319-1333 (2004), https://doi.org/10.1044/1092-4388(2004/099)
7) J.M. Franchak, K.S. Kretch, K.C. Soska and K.E. Adolph: "Head-mounted eye tracking: A new method to describe infant looking", Child Development, 82, 6, pp.1738-1750 (2011), https://doi.org/10.1111/j.1467-8624.2011.01670.x
8) L. Smith, C. Yu, H. Yoshida and C.M. Fausey: "Contributions of head-mounted cameras to studying the visual environments of infants and young children", Journal of Cognition and Development: Official Journal of the Cognitive Development Society, 16, 3, pp.407-419 (2015), https://doi.org/10.1080/15248372.2014.933430
9) C. Yu and L.B. Smith: "Embodied attention and word learning by toddlers", Cognition, 125, 2, pp.244-262 (2012), https://doi.org/10.1016/j.cognition.2012.06.016
10) B. Long, V. Xiang, S. Stojanov, R.Z. Sparks, Z. Yin, G.E. Keene, A.W.M. Tan, S.Y. Feng, C. Zhuang, V.A. Marchman, D.L.K. Yamins and M.C. Frank: "The BabyView dataset: High-resolution egocentric videos of infants and young children's everyday experiences", arXiv cs.CV (2024), http://arxiv.org/abs/2406.10447
11) Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei and Y. Sheikh: "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", arXiv cs.CV (2018), http://arxiv.org/abs/1812.08008
12) A. Mathis, P. Mamidanna, K.M. Cury, T. Abe, V.N. Murthy, M.W. Mathis and M. Bethge: "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning", Nature Neuroscience, 21, 9, pp.1281-1289 (2018), https://doi.org/10.1038/s41593-018-0209-y
13) F. Diaz-Rojas and M. Myowa: "Estimation of human body 3D pose for parent-infant interaction settings using azure Kinect and OpenPose", MethodsX, 13, 102861, pp.102861 (2024), https://doi.org/10.1016/j.mex.2024.102861
14) A. Cristia, M. Lavechin, C. Scaff, M. Soderstrom, C. Rowland, O. Räsänen, J. Bunce and E. Bergelson: "A thorough evaluation of the Language Environment Analysis (LENA) system", Behavior Research Methods, 53, 2, pp.467-486 (2021), https://doi.org/10.3758/s13428-020-01393-5



佐治　伸郎（さじ　のぶろう）　2010年，慶應義塾大学大学院政策・メディア研究科後期博士課程単位取得退学．2011年，博士（学術）取得．2011年，日本学術振興会特別研究員（PD）．2014年，鎌倉女子大学児童学部子ども心理学科専任講師．2022年より，早稲田大学人間科学学術院准教授．実験心理学的手法を用いて，子どもの言語習得過程および言語習得と認知発達の関連について研究を進めている．博士（学術）．

講　座　映像情報×認知科学〔最終回〕

マルチモーダルインタラクション

伝　康晴†

1．まえがき

マルチモーダル（Multi-modal）とは，テキスト・画像・音声など，複数の異なる種類の情報を同時に扱うことを指す．とくに，マルチモーダルインタラクション（Multi-modal Interaction）は，音声のパラ言語情報（声の高さや抑揚）や身体動作による非言語情報（視線・姿勢・表情・身振りなど）と，言語情報（発話の意味内容）を同時に使って，人間同士や人間とシステムがコミュニケーションすることを指す．人類学者のBirdwhistellは，人間同士の会話において，言語情報が占める割合は3割程度であり，残りの7割程度はパラ言語・非言語情報が占めると見積もっている[1]．マルチモーダルインタラクションの重要性はAI分野でもますます高まっており，毎年開催される対話システムライブコンペティション（聴衆の前で実際に対話システムを動作させ，評価を行うイベント）においても，近年はマルチモーダル対話システムが対象となっている[2]．

本稿では，マルチモーダルインタラクションにおける非言語情報のうち，まず視線と頷きに注目し，それらの特徴と機能を筆者自身の研究事例をまじえて紹介する．ついで，言語情報と非言語情報を合わせて分析するマルチモーダルインタラクション分析の事例を紹介する．これらの研究事例を通じて，マルチモーダルAIシステムが取り組むべき課題について議論する．

2．視線と順番交替

視線（Gaze）の第一義的な機能は，物や人を見ることであるが，マルチモーダルインタラクションにおいては，他の会話参与者を見ることがとりわけ重要な意味を持つ．会話の聞き手は通常，話し手を見ていることが多いが，このような聞き手による視線は，関心や意欲，発話の促進として働くことが指摘されている[3]．一方，話し手の視線は，会話運用においてとくに重要な役割を果たしている．ここでは，会話のもっとも基本的な運用規則である順番交替（Turn-taking）を取り上げ，それと話し手・聞き手の視線との関係について述べる．

順番交替とは，一人の話者の発話が終わったところで他の話者が次の発話を始めるということが規則的に繰り返される現象である[4]．一般に，順番交替の前後の発話に大きな重なりや隔たりはなく，テンポよく話者は移り変わっていく．このような円滑な順番交替を実現するために会話参与者たちが用いている技法の一つとして，現在の話し手による次話者選択というものがある．これは，質問や依頼などの他者に働き掛ける発話を，名前による呼びかけや特定の参与者への視線といったアドレス手段とともに用いるというものである（例えば，Aさんに視線を向けながら「明日来る？」と尋ねる）．このように，視線は順番交替において重要な役割を果たしている．

Kendonは，会話の録画資料の詳細な分析に基づき，話し手が次話者を注視することで順番交替を合図し，次話者が発話末まで話し手を注視し，その後目をそらすことで順番交替を受け入れるとしている[5]．つまり，発話末での相互注視（Mutual Gaze）が順番交替の契機となる．したがって，次話者選択のためには，話し手が次話者に視線を向けるだけでなく，次話者のほうもその話し手の視線を見て自身が選択されていることを知る必要がある．そのため，もし話し手が聞き手からの視線を得ていなければ，発話を休止したり再開始したりして，聞き手の注目を得ようとする[6]．

Kendonの分析は2人会話のデータに基づくものである．次話者選択の技法は3人以上の参与者による会話でも利用可能であるが，実際のデータではどうだろうか．筆者らは『千葉大学3人会話コーパス』の人手による視線アノテーションデータ（図1）を用いて，話し手の視線の向け先を分析した[7]．3人会話では，話し手の発話の聞き手は2人いる．直後に順番交替が生じるとすると，聞き手は2種類に分類できる．すなわち，次話者になる聞き手と，次話者にならない聞き手（非次話者）である．話し手が発話中にいずれの聞き手に視線を向けるかに応じて，以下の4通りのパター

† 千葉大学 大学院人文科学研究院
"Visual Information × Cognitive Science (final study); Multi-Modal Interaction" by Yasuharu Den (Graduate School of Humanities, Chiba University, Chiba)

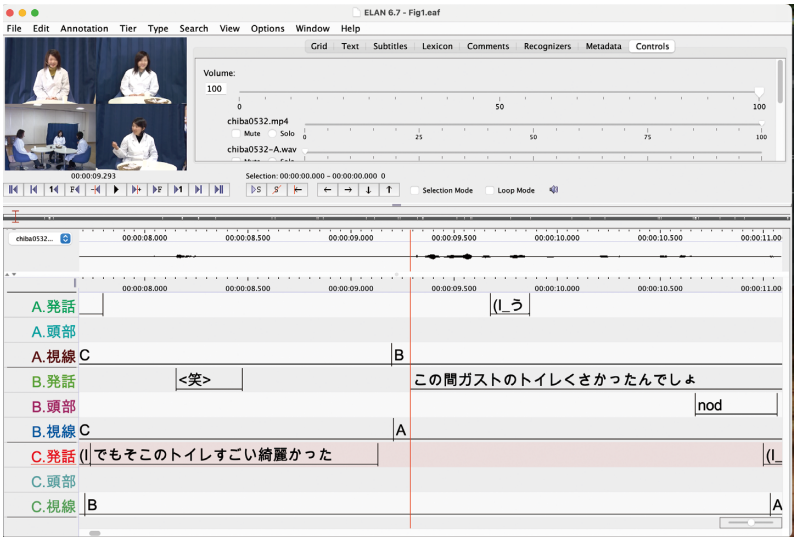


図1　映像アノテーションソフトELANによる発話と非言語情報のアノテーションの例

ンが想定できる．なし：どちらにも視線を向けない，次話者：次話者にだけ視線を向ける，非次話者：非次話者にだけ視線を向ける，双方：両方の聞き手に交互に視線を向ける．図2（左）は働き掛け発話のときの話し手の視線の向け先の分布を示す．次話者にだけ視線を向ける場合がほとんどであり，働き掛け発話に伴って話し手が特定の会話参与者に視線を向け，その参与者が次話者になる，すなわち，次話者選択技法の利用が頻繁に観察されることを示している．一方，図2（右）の働き掛け発話以外（働き掛け発話への応答や，とくに反応を要求しない発話）のときの分布を見ると，働き掛け発話ほど顕著ではないものの，やはり話し手が次話者にだけ視線を向けている場合が多いことがわかる．このことは，話し手によって特定の会話参与者の次発話が義務付けられない（＝誰もが自主的に発話を開始してよい）ときでも，話し手に視線を向けられていた参与者が次話者になりやすいことを示している．例えば図1では，Cの発話「でもそこのトイレすごい綺麗かった」は働き掛け発話ではないが，視線を向けられていた聞き手Bが次話者になっている．したがって，話し手に選択的に視線を向けられた聞き手が次の話者になるというのは，次話者選択技法を超えた一般的な傾向と言えよう．

ここまでは，話し手の視線について見てきた．次に，聞き手の視線について見てみよう．ここでは，3人会話における次話者と非次話者の視線行動の違いを分析した研究事例を紹介する[8]．まず，聞き手は，話し手の発話中，その話し手を見ていることが多い．この点に関しては，次話者と非次話者で大きな違いはない．一方，話し手の発話中に，話し手以外の会話参与者（＝自分以外の他の聞き手）に視線

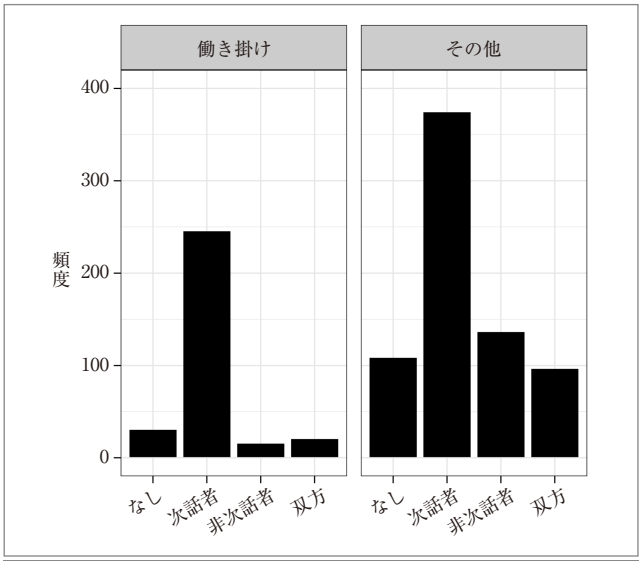


図2　話し手の視線の向け先の分布

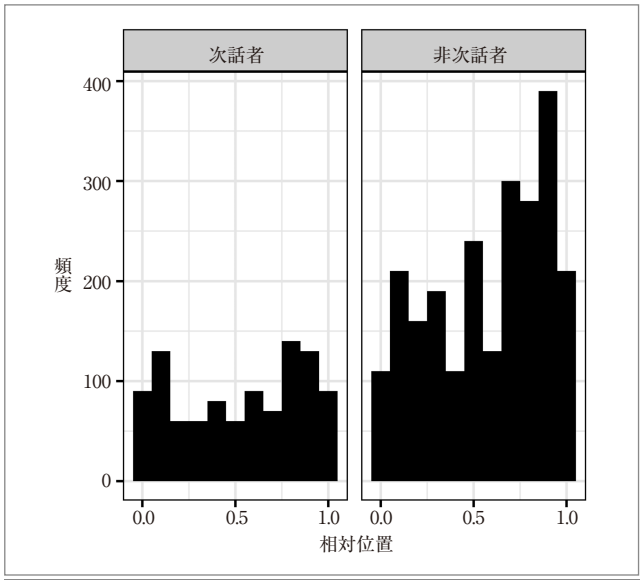


図3　聞き手による他の聞き手への視線の開始位置の分布

を向ける行動もしばしば観察される．図3は，現在の話し手の発話開始時点を0，次話者の次発話開始時点を1としたときに，他の聞き手に視線を向け始める時点を相対位置で算出し，その分布を示したものである（順番交替の前後に休止があれば，その休止は0～1の区間に含まれ，逆に，重複があれば，重複開始箇所が1である）．まず，全体的な頻度から，この行動は非次話者に顕著であることがわかる．つぎに，非次話者における分布の形状を見ると，この視線の開始位置は相対位置1直前に集中している．つまり，現在の話し手の発話末付近，ないしは，次話者の発話開始直前で，非次話者は次話者に視線を向け始めるのである．例えば，図1では，Cの発話の終了直後，次話者Bの発話開始直前に，非次話者であるAは現話者Cから次話者Bに視線を向け変えている．このことは，非次話者は，必ずしも，次話者の発話開始を聞いてから視線を向けるのではなく，次話者の発話開始にあわせて（予測して）そこに視線を向ける場合が多く見られることを示している．したがって，視線行動に関して言えば，発話順番を受け渡す当事者（現話者と次話者）以外の会話参与者も，順番交替に志向しているのである．

以上のように，会話における話し手・聞き手の視線は順番交替と深く関わっている．このことは，3人以上の参与者による多人数会話において，とくに重要になるだろう．AI分野におけるマルチモーダル対話システムはほとんどが2人会話を想定したものであり，多人数の対話システムはほとんど例がない（数少ない先駆的な事例として，松坂らによる研究がある[9]）．しかし，近年のAI技術の飛躍的な発展によって2人対話システムがある程度成熟した次のステップとして，多人数対話システムを本格的にターゲットとする時代が来るであろう．本節の知見は，そのようなマルチモーダル対話システムの設計に示唆を与える．

3．頷きの諸特徴

頷き（Nodding）もまた，会話において重要な役割を果たしている．聞き手は話し手の発話を黙って聞いているだけではなく，会話への参与を示す反応を積極的に産出する．聞き手が発する「うん」，「そう」などの反応は，一般にあいづちと呼ばれ，頷きをあいづちの一種と考える研究者もいる．本稿では，頭部運動（上下運動）による反応を頷きとし，言語によるあいづちとは区別する．聞き手による頷きは話し手の承認欲求を満足させ，発話量を増大させることが指摘されている[10]．また，頷きの頻度は言語・文化差が大きく，日本語話者はアメリカ英語話者と比べて，約3倍も多く頷くという報告もある[11]．

頷きは単独で産出されることもあるが，あいづちとともに用いられることも多い．以下では，どのような条件で頷きがあいづちと共起しやすいのか，また，あいづちと共起する頷きの形態（振幅や反復回数など）があいづちのどのような特徴から影響されるのかについて分析した研究事例を紹介する[12]．あいづちとして，以下の3種類の形態を対象とした．応答系感動詞：受容や承認を表す感動詞（「はい」，「うん」など），感情表出系感動詞：気づき・驚き・感心を表す感動詞（「ああ」，「へえ」など），語彙的応答：理解や同意を表す慣習的な表現（「そう」，「ね」など）．『千葉大学3人会話コーパス』に出現するこれらのあいづちについて，頷きとの共起，および，共起する場合の頷きの物理的特徴を分析した．結果を**表1**にまとめる．まず，頷きがあいづちと共起するかどうかを予測する統計モデルを構築したところ，応答系感動詞が他の形態のあいづちより頷きと共起しやすく，また，継続長が長いあいづちや，基本周波数が低いあいづちほど頷きと共起しやすいことがわかった．さらに，あいづちと共起する頷きについて，各参与者の正面映像から各頷きの振幅（＝最高点と最低点の差）を画像処理で抽出し，あいづちの緒特徴からの影響をモデル化したところ，応答系感動詞と共起する頷きは他の形態のあいづちと共起する頷きよりも振幅が小さく，話し手の発話途中に打たれたあいづちと共起する頷きは発話末尾や第3位置（働き掛け－応答の次の承認を与える位置）で打たれたあいづちと共起する頷きより振幅が小さかった．また，継続長が長く，基本周波数が高く，音量が大きいあいづちと共起する頷きほど振幅が大きいことがわかった．同様に，頷きの反復回数（1回の頷きの中での上下運動の回数）についても，話し手の発話直後に打たれたあいづちと共起する頷きがそれ以外の位置のあいづちと共起する頷きより反復回数が少なく，また，継続長が長く，音量の大きいあいづちと共起する頷きほど反復回数が多かった．このように，同時に産出される言語情報と非言語情報の間には深い結びつきがある（発話と身振りの結びつきについては，McNeilの成長点理論[13]を参照）．

表1　各統計モデルにおける有意な特徴（+/-は正負の相関）

あいづちとの共起	応答系＞感情表出系・語彙的 継続長＋，基本周波数－
振　幅	応答系＜感情表出系・語彙的 発話途中＜発話直後・第3位置 継続長＋，基本周波数＋，音量＋
反復回数	発話直後＜発話途中・第3位置 継続長＋，音量＋

頷きは1回の上下運動で終わることもあるが，何回か繰り返されることもよくある．上記のように，この反復回数は共起するあいづちの諸特徴に影響される．しかし，この分析は頷きとあいづちが共起する場合に限ったものであり，頷き全般の性質を検討したものではない．例えば，反復頷きでは，最初の上下運動の振幅が大きく，その後，振幅が徐々に小さくなっていくことが知られているが[14]，その減衰の仕方に規則性があるか検討した研究は見られない．筆者らは『千葉大学3人会話コーパス』のすべての頷きを対象

に，その時間的構造を分析した[15]．画像処理によって各頷きの上下運動の軌跡を抽出し，各サイクル（1回の上下運動）の振幅（山と谷の差）を算出した（図4（上：Example1）は単独頷き，図4（下：Example2）は反復頷きの例）．図5はサイクルの位置（1回の頷きの中での何番目のサイクルか）による振幅の変化を反復回数の異なる頷きごとにプロットしたものである（点は平均値，エラーバーは標準誤差）．この図から以下の三つのことがわかる．①振幅は反復が進むごとに（後の位置のサイクルほど）一定の割合で減衰していき，②反復回数が多い頷きほど最初のサイクルの振幅が大き

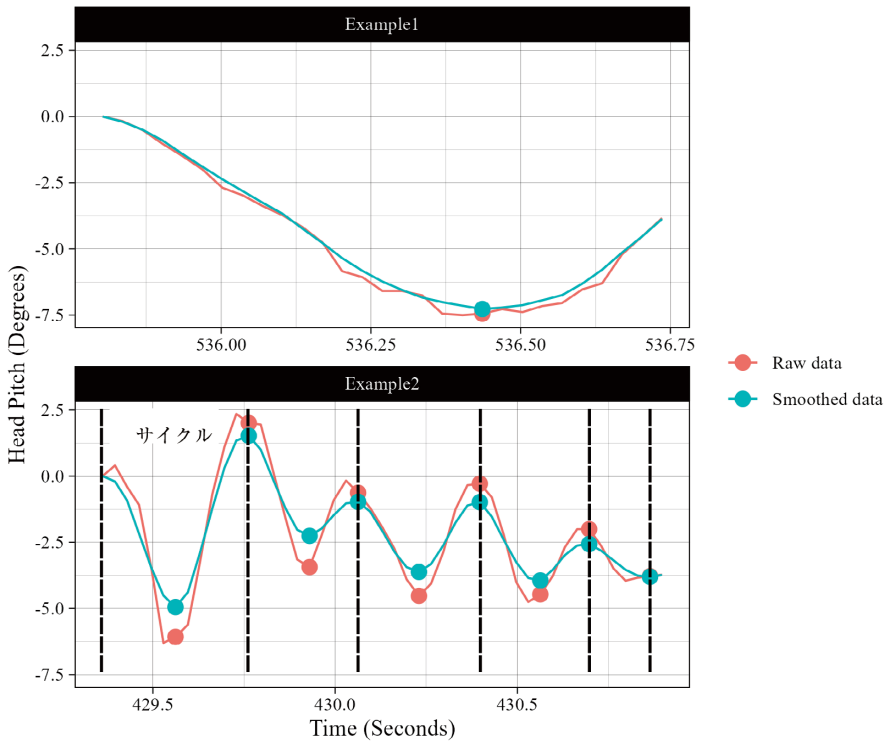


図4　頷きの上下運動の軌跡とピーク点（山と谷）の抽出の例（Raw dataは生データ，Smoothed dataはスムージング後のデータ）（文献[15]のFig 3に加筆して転載）

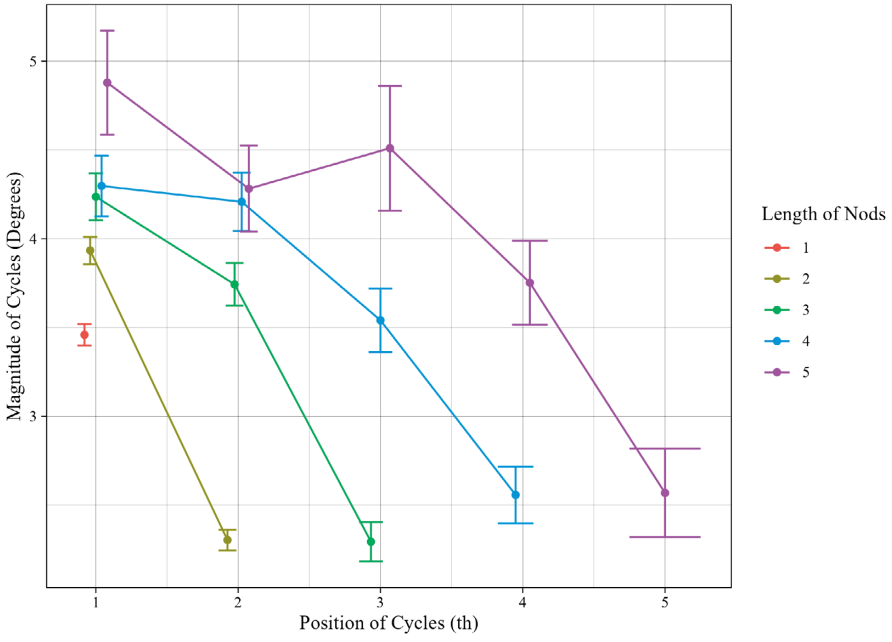


図5　頷きの反復回数ごとのサイクル位置による振幅の変化（文献[15]のFig 5を転載）

く，③最後のサイクルの振幅はそれまでの減衰傾向から予測されるよりも顕著に小さい（これらの傾向は統計モデルにより有意であることが示されている）．これらの特徴は音声における韻律構造が示す特徴と類似している．すなわち，①音声のピッチ（基本周波数）は発話末に向かって低くなっていき（Declination），②長い発話ほど冒頭のピッチは高く始まり（Anticipatory rising），③発話末尾のピッチは顕著に低くなる（Final Lowering）[16)～18)]．このことは，コミュニケーションにおける人間の反復的な行為一般に見られる普遍的な傾向（生理的・心理的要因による減衰傾向，長期的な計画に基づく冒頭からの予示，末尾境界のマーキング）が存在する可能性を示すものである．さらに，マルチモーダルインタラクションにおいては，これらの特徴が他の会話参与者にとっても利用可能かもしれない．例えば，特徴2により，頷きの冒頭の振幅によって反復回数が予測できるし，特徴3により，反復頷きの完結点を確認できる．また，特徴1により，振幅がそれまでの減衰傾向から増幅に転ずることは，当初とは別の新しい対象や情報への頷き反応が開始されたと理解できる[19)]．

以上見てきたように，頷きは共起するあいづちとの間に深い結びつきを持ち，また，それ自体，規則的な時間的構造を持つ．マルチモーダル対話システムが自然な非言語行動を産出するためには，このような知見もふまえなければならないだろう．

4．マルチモーダルインタラクション分析

ここまで，視線と頷きについて，その特徴と機能を筆者らの研究事例をまじえて紹介してきた．最後に，言語情報と非言語情報を合わせて分析した研究事例を紹介する．マルチモーダルインタラクション分析（Multi-modal Interaction Analysis）は，とくに会話分析などの質的分析の分野で2010年代以降盛んになってきている．物質的世界における身体動作と言語の組織化に焦点を当てた学際的研究[20)]，複数の活動を同時に行う場面に焦点を当て社会的・言語的・身体的な現象として捉えた研究[21)]，言語使用と社会的相互行為を「複雑性」という観点から考察したマルチモーダル会話分析研究[22)]などがまとまった論文集として刊行されている．これらの研究の中で，さまざまな非言語情報と発話，さらには外界の物質とが密接に結びついて，会話を含む社会的活動が組織化される様が詳らかに記述されている．

ここでは，マルチモーダルインタラクション分析に関する筆者の研究事例を紹介する[23)]．図6は『日本語日常会話コーパス』に収録されている日常的なインタラクション場面からの抜粋である．この場面では，父親と息子2人が公園で野球の練習をしており，母親がそれを見守っている．母親はこの場面をコーパスに収録するために，頭に取り付けたウェアラブルカメラにより撮影している．図6（上）の囲み内は，母親の発話と身体動作，および，捕手を務める

T003_019：31.9-35.4
（（捕手を務める息子がボールを後ろにそらし，あわてて拾いに行く））

01 母親　ほ*らーそれじゃいっ(0.4)一点
母_身　　*息子を指差し------>
02 母親　入っ+ちゃ*うよー**取ん*ないとちゃんとー**
母_身　　*捕球の身振り-------->
*prep->*str>*ret---->
息_身　　+母親と視線を合わせる---->

図6　マルチモーダルインタラクション分析の例

息子（兄）の身体動作を書き起こしたものである．ここでは，マルチモーダル会話分析で標準的に用いられている転記方法[24)]を簡略化して援用している．発話は01，02のように番号が付与された行に記述し（発話中の(0.4)は休止の秒数），同時に生じている身体動作はその下の行に記述している（「母_身」，「息_身」はそれぞれ母親と息子の身体動作）．身体動作と発話のタイミングは“*”と“+”の記号によって示し（前者は母親，後者は息子の身体動作に対して使用），“-->”の記号は動作がその間継続していることを示す．身振りは，その時間的構造を細分化して次の行に示している．身振りは一般に，①定常位置（この場合は膝の上）から身振りをする位置までの準備的移動（Preparation），②その位置での身振りの実行（Stroke），③そこから定常位置への復帰的移動（Retraction）の三つの段階からなる[25)]．図6（上）では，これらの区間がそれぞれprep，str，retで示されている．図6（下）は，この身振りのstrの瞬間を捉えた，母親のウェアラブルカメラのスナップショット映像である．

さて，02行目の母親の発話「取んないとちゃんとー」（太字部分）に注目しよう．この発話は，述部「取んないと」と修飾部「ちゃんとー」の語順が逆転した倒置構文になっている．従来，倒置構文が生じる要因は言語学的観点から論じられてきた（誤用論的な有標化や強調など）．しかし，マルチモーダルインタラクションの観点からは，少し違った要因が見えてくる．この場面では，父親が投げたボールを弟が空振りしたのに対して，捕手を務める兄がボールを後ろ

にそらしてしまう．あわててそれを拾いにいく息子を指差しながら，母親は「ほらーそれじゃいっ（0.4）一点入っちゃうよー」と声を掛ける（01－02行目）．その直後に，母親は捕手の捕球の身振り（図6（下））をしながら，この倒置構文を発するのである．この身振りのストロークは「取ん」（「取る」の未然形「取ら」の撥音化）の発話と正確に同期している．身振りのストロークとそれが表現する語の発話とが同期する現象は広く観察される[13) 25)]．したがって，「取んないと」がこのタイミングで発されたのは，身振りと同期させるためと考えられる．では，この身振りはなぜまさにこのタイミングで産出されたのであろうか．言い換えると，「ちゃんと取んないと」という標準的な語順で発話して，その「取ん」に身振りを同期させることもできたはずである．ここで，息子の視線に注目しよう（02行目の最下段）．ボールを拾いに行って，母親から声を掛けられた息子は，母親に振り返り，視線を合わせる．母親の身振りが開始されるのはまさにその直後である．つまり，息子と相互注視が成立したすぐ直後に（息子の視線がそれないうちに）身振りを開始し，そのストロークに同期して「取んないと」が発話されている．結果として，「ちゃんと」の発話は先送りされ，倒置構文になっているのである．

このような発話・視線・身振りの精密なタイミング調整は，それぞれのモダリティを独立に産出していたのでは成立しない．つまり，人間のマルチモーダルインタラクションにおいては，さまざまなモダリティが極めて密接に結びつきながら産出/理解されている．もし日常生活における人間の自然なマルチモーダルインタラクションをAIシステムが真に実現しようとするならば（そのことの是非は別として），このことは極めて挑戦的な課題となるだろう．

5．むすび

本稿では，マルチモーダルインタラクションにおける非言語情報，とくに視線と頷きの特徴と機能，および，言語情報と非言語情報を合わせて分析するマルチモーダルインタラクション分析に関する研究事例を紹介した．また，これらの研究事例を通じて，マルチモーダルAIシステムが取り組むべき課題について議論した．マルチモーダルインタラクションについては，ビデオデータの質的分析を駆使する会話分析などの分野において，その組織化に関わる会話参与者たちの実践の記述が蓄積されつつあるが，その背後にある認知的基盤に関する研究は進んでいない．それゆえ，AIシステムへの実装についても極めて挑戦的な課題として残されている．今後，認知科学やAI分野におけるブレイクスルーが期待される．

（2025年10月28日受付）

〔文　献〕

1）R.L. Birdwhistell: "Kinesics and Context: Essays on Body Motion Communication", University of Pennsylvania Press (1970)
2）東中竜一郎，高橋哲朗，稲葉通将，斉志揚，佐々木裕多，船越孝太郎，守屋彰二，佐藤志貴，港隆史，境くりま，船山智，小室允人，西川寛之，牧野遼作，菊池浩史，宇佐美まゆみ：“対話システムライブコンペティション6”，人工知能学研究会資料，SIG-SLUD-099，pp.84-89（2023）
3）M. Argyle and M. Cook: "Gaze and Mutual Gaze", Cambridge University Press (1976)
4）H. Sacks, E.A. Schegloff and G. Jefferson: "A Simplest Systematics for the Organization of Turn-Taking for Conversation", Language, 50, pp.696-735 (1974)
5）A. Kendon: "Some Functions of Gaze Direction in Social Interaction", Acta Psychologica, 26, pp.22-63 (1967)
6）C. Goodwin: "Conversational Organization: Interaction between Speakers and Hearers", Academic Press (1981)
7）榎本美香，伝康晴：“話し手の視線の向け先は次話者になるか”，社会言語科学，14，1，pp.97-109（2011）
8）伝康晴：“多人数会話におけるしぐさの語用論”，月刊言語，36，12，pp.48-55（2007）
9）松坂要佐，東條剛史，小林哲則：“グループ会話に参与する対話ロボットの構築”，信学誌，J84-D-II，pp.898-908（2001）
10）J.D. Matarazzo, G. Saslow, A.N. Wiens, M. Weitman and B.V. Allen: "Interviewer Head Nodding and Interviewee Speech Durations", Psychotherapy: Theory, Research & Practice, 1, pp.54-63 (1964)
11）メイナード泉子：“会話分析”，くろしお出版（1993）
12）森大河，伝康晴：“相槌の特徴に一致した頷き生成モデル”，人工知能学論，37，pp.IDS-H_1-12（2022）
13）D. McNeill: "Hand and Mind: What Gestures Reveal about Thought", University of Chicago Press (1992)
14）U. Hadar, T.J. Steiner and F. Clifford Rose: "Head Movement during Listening Turns in Conversation", Journal of Nonverbal Behavior, 9, pp.214-228 (1985)
15）T. Mori, Y. Den and K. Jokinen: "Structure of Nods in Conversation", PLOS ONE, 20, 5, e0323448 (2025)
16）M. Liberman and J. Pierrehumbert: "Intonational Invariance under Changes in Pitch Range and Length", M. Aronoff and R. Oehrle (eds.), Language Sound Structure, MIT Press, pp.157-233 (1984)
17）K. Maekawa: "Five Pieces of Evidence Suggesting Large Lookahead in Spontaneous Monologue", Proceedings of the 9th Workshop on Disfluency in Spontaneous Speech, pp.7-10 (2019)
18）K. Maekawa: "A New Model of Final Lowering in Spontaneous Monologue", Proceedings of INTERSPEECH 2017, pp.1233-1237 (2017)
19）森大河，伝康晴：“反復的うなずきのセグメンテーション単位に関する分析”，人工知能学会研究会資料，SIG-SLUD-102，pp.146-151（2024）
20）J. Streeck, C. Goodwin and C. LeBaron (eds.) :"Embodied Interaction: Language and Body in the Material World", Cambridge University Press (2011)
21）P. Haddington, T. Keisanen, L. Mondada and M. Nevile (eds.): "Multiactivity in Social Interaction: Beyond Multitasking", John Benjamins Publishing (2014)
22）P. Haddington, T. Eilittä, A. Kamunen, L. Kohonen-Aho, I. Rautiainen and A. Vatanen (eds.): "Complexity of Interaction: Studies in Multimodal Conversation Analysis", Palgrave Macmillan (2023)
23）Y. Den: "When Gestures Affect Syntactic Structures: A Case of Postposed Construction in Japanese Conversation", Oral presentation at the 8th Conference of the International Society for Gesture Studies (2018)
24）L. Mondada: "Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multi-Modality", Research on Language and Social Interaction, 51, pp.85-106 (2018)
25）A. Kendon: "Gesture", Cambridge University Press (2004)

伝　康晴（でん　やすはる）　1993年，京都大学大学院工学研究科博士後期課程研究指導認定退学．ATR音声翻訳通信研究所研究員，奈良先端科学技術大学院大学情報科学研究科助教授などを経て，現在，千葉大学大学院人文科学研究院教授．人間の日常コミュニケーションを統計モデリングから相互行為分析・フィールドワークまで多用な方法論で分析している．博士（工学）．