

当財団では、2018年度より会員相互の情報交換と幅広いネットワーク作りを主目的に、参加者や賛助会員等から要望の強いテーマであるAIやロボットなどIT関連を講演テーマとして「中部社研ITフォーラム」を開催しております。

本レポートは、第11回（2020年7月29日）での講演「視覚的インタラクションにフォーカスした環境・状況・状態認識技術」を元に、講演者の名古屋大学未来社会創造機構特任准教授 平山高嗣氏にその概要を執筆いただいたものです。

視覚的インタラクションにフォーカスした環境・状況・状態認識技術



名古屋大学未来社会創造機構特任准教授 平山 高嗣

1. 視線の魅力

人間と共生するシステムを構築することはコンピュータサイエンスにおける主要な課題の1つです。システムが人間と協調的に活動するためには、人間の振る舞いを外部から観測し、興味や意図などの内部状態を推定することが重要です。その手がかりとして、発話、表情、姿勢、動作などが挙げられ、視線もその1つです。視線分析に関する近年の技術進展と長年の心理物理学知見の蓄積により、視線計測を応用したインタフェース設計への関心が高まっています。

古くから、「目は心の窓、口ほどにものを言う」と言われ、19世紀後半から心理物理学、工学分野を中心に研究されています。目は多くの情報を入力し、かつ内部状態を出力するインタフェースであるため、視覚情報をメディアとするヒューマンコンピュータインタラクション（HCI）を設計する上で、視線は非常に重要な要素となります。システムは時間的に連続するデータとして視線を計測でき、人間はその中に無意識的に内部状態の情報を漏らしてしまうため、コマンドやポインティングだけではない人間同士のやり取りを模した自然なインタラクションを実現できる可能性があります。しかしながら、人間を外から受動的に観察するだけで内部状態を推定することは困難です。我々人間でさえも観察だけに頼って他者の内部状態を推定す

ることは簡単ではありません。

では、人間はどのようにして他者の内部状態を推定しているのでしょうか？人間同士の対話に目を向けると、相手の内部状態を正確に理解したい場合に、積極的に働きかけを行って、その反応を観察していることが分かります。そこで著者らはこれに類推して、システムが主導権を持って積極的に対話を行うプロアクティブな対話戦略を適用し、システムがユーザーに視覚的な働きかけを行い、それに対するユーザーの反応から内部状態を推定するという視覚的インタラクション「Mind Probing」を提案してきました。本稿では、一世紀を超えて行われてきた視線研究の進展を概観し、著者らが取り組んできた「Mind Probing」に関わる研究動向を、インタラクションの参与対象が人間とシステム、人間と環境である場合に分類して紹介



図1 環境・状況・状態認識のための視覚的インタラクションのモデル化

します。ここで、これらのインタラクションが必ずしも厳密に相互作用する構造になっていないことに注意されたい。

2. 視線研究の歴史

視線は19世紀後半から工学、心理物理学分野を中心に、主に計測と分析の観点で研究されています。視線を正確に計測するためには、個人差がある眼球の形状と構造を推定する必要があります。また、視行動は不確定性を多く含む人間の内部状態の影響を受けやすいため、その進展はゆっくりしたものとなっています。

現在の視線計測技術の主流である赤外光照射法の原型といえる光源角膜反射法が1901年にDodge&Clineによって提案されました。また、対話分析で現在よく用いられている頭部装着型のものについては、その半世紀後にHartridge&Thomsonによって開発されました。視線分析への視線計測装置の導入は1950年代に入ってから一般的になり、その多くは眼球運動の分類やコックピットなどのユーザビリティの評価を目的としました。1970年以前は、学習、記憶、注意、認知的負荷といった認知要因に入り込まない研究が主でした。

1960年代後半から工学分野での高まりが一旦落ち着きます。その一方で、心理物理学分野において認知要因を考慮した研究が盛んに行われ、神経科学分野ではサルを用いた視覚分析の研究が活発化しました。Yarbusは、視行動が観察者に課された心的課題に敏感に反応することと、顔などの視対象においては、自由視で注視箇所がほぼ共通する場合があることを明らかにしました。また、Just&Carpenterは、物体や文字の比較という認知要因に注目し、視行動の理論的モデルを提案しました。共同注意という対話現象が研究対象として注目され始めたのも1970年代のことです。

視線計測のリアルタイム処理が可能になった1970年代後半から、視線を入力モダリティとするインタフェースが開発され、HCIへの視線の介入が

注目され始めます。心理物理学分野においても、視覚的注意をめぐって2つの大きな進展がありました。まず、Posnerによって提案された「先行手がかり (pre-cueing)」を用いた実験パラダイムが視覚的注意に関わる研究を活発化しました。そして、工学分野にも影響を及ぼす2つのモデルが提案されました。Treisman&Geladeによる特徴統合理論 (Feature integration theory) とKoch&Ullmanによる顕著性マップ (Saliency map) です。

1990年代に入ると、ユーザビリティ評価への取り組みが再燃します。特徴統合理論に対抗するモデルや顕著性マップに基づく視覚的注意の計算モデルも多く提案されます。対話研究においては、心の理論獲得へと導かれるモジュールとして、視線によるインタラクションを内包するモデルがBaron-Cohenによって提案されました。

近年は、視線計測装置の高精度化、小規模化、低廉化と各分野での知見の蓄積が相まって、学際分野での研究が活発化しています。実験室から外に出た環境での視線計測や多人数対話における視覚的インタラクション分析、視覚的注意の一人称視点モデル設計、霊長類の視線分析、機械学習を導入した視覚的注意や視行動の計算モデル構築といった研究も盛んに行われています。

3. 視線と内部状態の関係

人間の内部状態は多様な認知要因を基底とし、多次元変数で記述されるであろう一方で、視線は低次元情報であるため、それらの関係を一意に定めることは困難ですが、制限された状況や特定のタスクに対しては、関係が明らかになりつつあります。

その中で、知見が蓄積されているものとして、まず、インタフェースのユーザビリティ評価を目的とした視行動の分析が挙げられます。インタフェースから情報を抽出する難しさが固視の持続時間で説明できるという結果が得られています。また、インタフェース全領域での固視回数がモジュールの検索効率と関連し、各モジュール領域に対する固

視回数が各モジュールの重要性を表すとされています。これらの関係に関わる重要な認知要因はインタフェースへの慣れです。

次に、自動車などのシステム操作時の安全性評価を目的として、認知的負荷 (cognitive distraction) と視行動との関係が盛んに分析されています。例えば、自動車運転などの主タスクと並行して暗算や音声遠隔対話といった副タスクを被験者に課し、主タスクのみの場合との視行動の差異が分析されます。主タスクより副タスクに多くの注意資源が割り当てられるほど視行動が単調になり、自動車運転状況であれば視方向が視野前方に集中するという関係が一般的知見とされています。

視覚探索課題においては、注視点を固定して被験者に視覚刺激を探索させ、探索達成時間を計測することで潜在的注意 (covert attention) の分析を行うことが主流でした。視線計測とパターン認識の技術進展とともに、近年は注視点を固定せずに、視線滞留頻度から生成される注視点マップや注視点の時系列パターン (scan path) に基づいた顕在的注意 (overt attention) の分析が多くなっています。いずれの場合にも、探索の先行手がかり (事前知識) の提示の有無と視行動との関係が比較されています。また、注視点マップや scan path をシミュレートする計算モデルも多々提案されています。

上記ほどの知見の蓄積がないものの、視対象への興味とその対象への累積注視時間や注視頻度が相関するという分析結果が得られています。

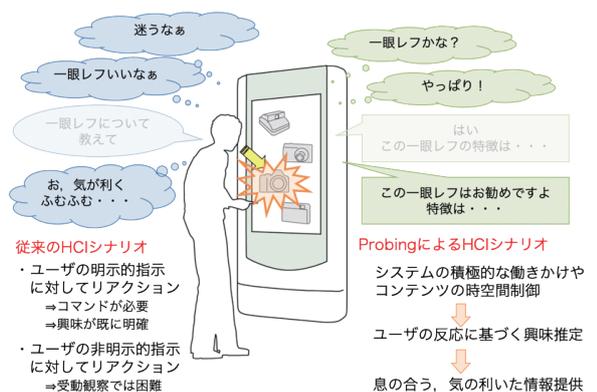


図2 Mind Probing のコンセプト

4. Mind Probing

外環境からの視覚的な働きかけに対する人間の反応に基づいて内部状態を推定する「Mind Probing」では、外環境と視線それぞれのダイナミクス関係、特に各々のイベント間の時間関係 (タイミング) に注目します。先行手がかりを用いた視覚探索実験パラダイムと自動車運転時の認知的負荷を評価するための Peripheral detection task において、外部イベントに対する反応時間が計測されることや、対話における発話交替潜時から相手の気持ちを読み取れることから、内部状態を反映するタイミングという人間の時間感覚が重要であることが分かります。内部状態とタイミングの関係をモデル化できれば、あうんの呼吸で人間とやり取りするシステムを設計できる可能性があります。

5. 人間とシステムの視覚的インタラクション

視覚情報をディスプレイに表示する状況において、ユーザーがどのオブジェクトを注視しているかという情報は、ユーザーの興味や関心を知る上で重要な手がかりになります。そして、興味や関心を推定することによって、気の利いた情報の推薦などが可能になります。また、完全自動運転車の実用化が現実的になってきましたが、周囲の交通参加者による無人車両の走行意図の理解が困難になることが危惧されています。そこで、視覚的インタラクションの設計が重要視されています。

(1) Gaze Probing : イベント提示に基づく注視対象推定

ユーザーの注視対象を推定するために視線の計測が重要になりますが、その計測精度は前述したように眼球の形状と構造の個人差やシステムに対するユーザーの立ち位置や顔向きなどの影響を受けます。そこで、オブジェクトに移動、停止といった特徴的な動き (イベント) を持たせた動的コンテンツを用いて、オブジェクトと同調した眼球

運動を引き出し、双方の動きが同期したタイミングを抽出することによって注視対象を推定するGaze Probingを提案しました。この原理は、視線計測の空間的な精度が十分に望めない状況でも、オブジェクトの動きと相関する眼球運動のパターンを抽出することで注視対象を高精度に推定できます。現在、他の研究者によって改良が加えられて波及しています。

(2) 能動的な情報提示に対する眼球運動の反応遅延に基づいた興味推定

人間は空間に並べられたさまざまな視覚情報から好みのものを選択する状況において、まず、(1) 全体をざっと俯瞰し、その後、(2) いくつかの情報を吟味する傾向があります。吟味状態では、興味対象に能動的に注意を向けるため、情報提供システムがユーザーの興味を推定する場合、まずユーザーが吟味状態であるかどうかを認識すると良いと考えられます。そこで、動的なコンテンツ設計によって、ユーザーを吟味状態に強制的に誘導し、興味対象に能動的に注意を向ける内因性眼球運動を誘発することを試みました。

複数のページで構成されるコンテンツを複数提示する状況を想定します。システムはまず、(1) コンテンツを1つ表示して短い時間間隔でページを最後まで切り替え、続いて、ディスプレイの別の位置にほかのコンテンツを同様に表示していきます。これによって、ユーザーは一通りの情報を俯瞰することができ、脳内の認知地図にはコンテ

ツの空間情報が記録されます。次に、(2) これまでに提示したコンテンツを同じ場所に再び表示します。ただし、ユーザーが興味のあるコンテンツを吟味できるように、ページ切り替えの時間間隔を長く設定します。ここで、コンテンツ間のページ切り替えの順番を、ユーザーに予想させないためにランダムとしますが、ユーザーが認知地図を参照して興味対象に注意を向けることを期待します。

興味の推定には、吟味フェーズにおけるコンテンツのページ切り替えイベントと眼球運動の発生イベント間のタイミング構造に注目します。コンテンツ選択を行う被験者実験により、興味対象を注視している場合に、ほかのページ切り替えへの反応が遅れる傾向や興味がない対象を注視している場合に、興味があるページ切り替えへの反応が早くなる傾向を確かめました。この反応遅延に基づいて、最も興味のあるコンテンツを推定したところ、従来研究で用いられている注視時間と注視頻度より高い精度が得られました。

(3) Gaze Mirroring : ユーザーの興味を顕在化させる注視模倣

5. (1) と (2) で述べたプロアクティブな対話戦略は、おせっかいさや落ち着きの無さを受け手に感じさせてしまう面があります。そこで、リアクティブな対話戦略でユーザーの内部状態を反応に引き出すことを試み、擬人化エージェントがユーザーと共同注意を実現する注視模倣Gaze Mirroringを設計しました。

1. 外因性サックードの誘発 (俯瞰フェーズ)

システム: コンテンツを短い時間間隔で提示 → 決まった順番に提示

ユーザ: 全体をざっと見渡し、脳内にCognitive mapを構築

2. 内因性サックードの誘発 (吟味フェーズ)

システム: コンテンツを長い時間間隔で提示 → ランダムに提示

ユーザ: Cognitive mapを参照し、興味のあるコンテンツの領域に注意を向け、その内容を十分に読み取り、見比べも



図3 興味を反映した眼球運動を引き出す情報提示シナリオ

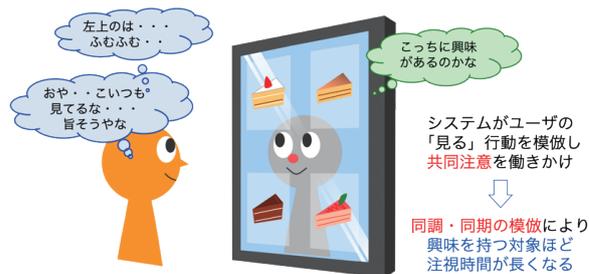


図4 ユーザの興味を顕在化させる注視模倣

共同注意は、他者の心を理解するプロセスの一つで、関心を共有する事物へ他者と共に注意を向けるように行動を調整することです。これに基づけば、提示情報へのユーザーの注視行動を、システムが擬人化エージェントを用いて模倣することで、ユーザーの注視行動に影響を与え、興味を顕在化させられる可能性があります。ユーザーが注視している対象をシステムが同調かつ同期して注視すると、ユーザーがその対象に興味を持っている場合は共同注意を行うことに抵抗がなくそのまま注視を続け、興味を持っていない場合には注視を外す回避行動を取ることが期待できます。

システムが複数のコンテンツをディスプレイに提示し、ユーザーの注視行動に応じてディスプレイ上の目玉エージェントの視線を変化させることでGaze Mirroringを行ったところ、興味があるコンテンツ領域への注視持続時間がGaze Mirroringを行わない場合に比べて有意に増加しました。模倣を扱うHCI研究は盛んに行われており、その先駆的な技術となりました。

(4) 自動運転車の走行意図に対する歩行者の理解を反映する視行動の分析

自動運転車の機能と性能は、安全性を重視して継続的に向上しています。しかし、社会的受容や信頼感の不足などの問題に直面しています。この原因の1つとして、歩行者と自動運転車がインタラクションする際に、歩行者が自動運転車の走行意図を理解できないことが考えられます。

従来は、車載カメラ映像を用いた手動運転車とのインタラクションの分析が多く行われてきましたが、歩行者側の視点からの分析も重要です。そこで、自動運転車とのインタラクションにおける歩行者の視行動を分析しました。歩行者の視線を計測し、手動および自動運転車の走行に対する主観評価を得た結果、手動運転車より自動運転車の安全感、安心感および走行意図に対する歩行者の理解度が低くなり、手動運転車より自動運転車に対する歩行者の注視時間が長くなりました。こ

れらには負の相関が確認されました。これは、歩行者が自動運転車を注視し続けている際に、走行意図を伝達することができれば、安全感と安心感を与え、ひいては社会的受容や信頼感を醸成できる可能性を示唆します。

6. 人間と環境の視覚的インタラクション

対話状況以外の日常活動シーンにおいても動的な視覚刺激が溢れているため、人間は外環境から「Mind Probing」を仕掛けられているようなものです。人間の内部状態を振る舞いに引き出すことができている視覚刺激がどういったものであるか特定できれば、環境設計などに応用することができます。

(1) 映像の顕著性変動と視線運動の時空間相関分析に基づいた集中状態検出

あるタスクを遂行している人間の集中状態を検出することは、システムがサポートするタイミングを計る上で重要です。著者らは、人間が映像を視聴している状況を取り上げ、視聴映像から視線運動に関連するプローブとなる映像ダイナミクスを抽出し、それと視線運動との間の時空間構造から視聴者の集中状態を検出する手法を提案しました。

映像を視聴する際に表出される視線運動は、映像中に存在する顕著な領域（生理的に視線が向きやすい領域）の構成や変化に影響されます。そこで、①顕著性マップを算出して映像中から顕著領域を抽出、②画面中の顕著領域数およびそれらの動きに基づいて、顕著領域の時空間的変動のパターンを分類、③被験者に映像視聴のみをタスクとして課した場合（集中状態）と副タスクとして暗算を課した場合（注意散漫状態）の視線データを蓄積、④各パターンに対して表出された視線運動から特徴量を抽出し、機械学習によって両状態を識別する手法を提案しました。そして、従来手法より20%ほど高精度な集中状態検出が可能であることを確認しました。

(2) 周辺車状況変化と視線遷移の時空間 相関分析に基づいた注意散漫状態検出

集中状態を推定することが有用である状況としては、自動車を運転する状況も挙げられます。そこで、周辺車状況の変化に対する運転者の視行動の時間的關係について、運転集中状態と注意散漫状態における差異を大規模運転行動信号コーパスを用いて分析しました。追い抜き車両を外周刺激として注目したところ、通常走行時（運転集中状態）は音声対話による楽曲検索走行時（注意散漫状態）に比べ500msec程度、有意に早く注視反応が起こることを確認しました。そして、追い抜き車両への注視の時間分布を尤度分布とし、機械学習によって両状態を識別する手法を提案し、従来手法に比べて10%程度の精度向上を達成しました。

しかしながら、得られた知見には、あらゆる追い抜き状況に適用することができる汎用性がなく、そもそも追い抜きが頻繁に生じる保証がありません。そこで、運転状況を機械的かつ網羅的に分析するために発想を転換し、まず、運転者が視線方向を遷移した時区間を検出して、その際の周辺車状況を視線遷移パターンごとに統計解析するデータ駆動型アプローチを提案しました。そして、視線遷移を誘発する周辺車状況を抽出し、その時空間的な位置関係を機械学習することで、上述の手法よりさらに20%程度の高精度かつ高効率な注意散漫状態と集中状態の識別を実現しました。定性

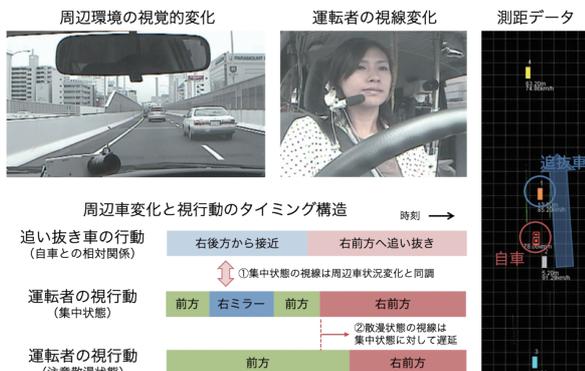


図5 周辺車状況変化と運転者の視行動のタイミング構造分析に基づく注意散漫状態検出

的な分析ですが、注意散漫状態の運転者の視線遷移が周辺車状況と同調しなかったことを示唆する妥当な結果が得られました。

(3) サッカー指導の熟練者と未経験者の 視行動分析

スキル差に関わる視行動の違いは、スポーツを対象に盛んに分析が行われています。著者らは、サッカーの試合映像に対する熟練指導者（S級ライセンス保有者）と指導未経験者の視行動の差異を、映像内で重要度が高い動的対象であるボールと視線運動の関連性に基づいて分析しました。そして、指導未経験者はボールを中心窩視（視野の中心）で捉えて追跡している時区間が多く、一方で熟練指導者は周辺視で捉える時区間が多いことが分かりました。フィールド上でのボールの存在領域とボールへの中心窩視の割合との関係については、指導未経験者がボールの存在領域によって割合をほとんど変化させないことに対して、熟練指導者は、センターライン付近では高く、ゴールラインに近くなるほど低くなる傾向がありました。つまり、熟練指導者は、時々刻々と変化する状況に応じて、注視対象を効率良く変化させると考えられます。

また、視覚環境ダイナミクスと視線運動間のイベント共起構造を記述し、熟練者と未熟者間での差異構造を視行動のコツとして捉え、データ駆動的に抽出するアルゴリズムを設計しました。具体的に

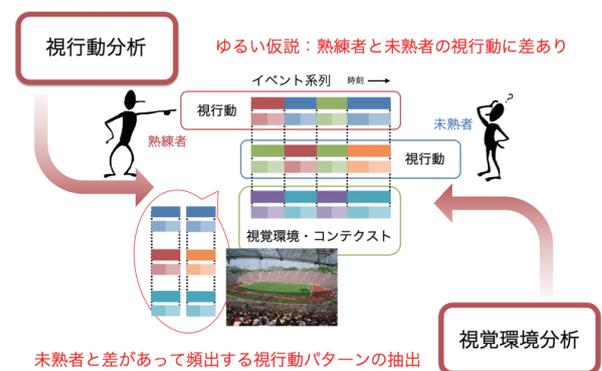


図6 データマイニングによる視行動のコツの抽出

は、サッカーの試合映像を視聴する状況において、ボールだけではなく選手のダイナミクスに対する視線運動パターンの依存性を指導の熟練者と未経験者との間で比較することで、熟練者に特有な視行動の抽出を実現しました。

(4) 運転の熟達に伴う視行動変化のモデル化

運転者は、視覚を通して運転に必要な情報の多くを取得しています。膨大な情報が時々刻々と変化する運転環境の中で、適切な視行動を取ることが交通事故を防ぐことにつながります。そのため、熟練ドライバーの視行動をモデル化し、それに基づいて交通環境を認識する機能を自動運転知能に実装することで、交通事故の削減につながると考えられます。事故が発生する要因が複数存在する箇所特に効率的な視行動が必要となるため、そのような箇所における視行動のモデル化が重要です。

これまでにその試みは多く行われており、近年では車載カメラ画像上での注視点を深層学習を用いて予測する例も報告されています。しかし、周囲の幾何的な情報を明示的に考慮していないため、物標との3次元的な位置関係に基づいた視行動のモデルは実現されていません。そこで、3次元地図上で運転者の頭部位置を推定し、死角あり狭路において運転者がどこから何を見るかを分析しました。そして、熟達に伴って、または車速が速いほど、より前方に存在するリスク要因に中

心視野を向けることを確認しました。

また、計測したデータに逆強化学習を適用して視行動を数理的にモデル化したところ、未熟時においては過剰に視線が変化し、熟達時には遠方に視野を安定させていることを裏付けるモデルが構築されました。

7. ますます熱い視線

本稿では、工学と心理物理学分野における視線研究の動向を概観し、著者が過去に行った視覚的インタラクション研究を紹介しました。およそ一世紀前から視線計測技術の開発が始まり、要素還元を主体としてゆっくりしたペースで視線が分析されてきましたが、近年はタブレットPCやスマートフォンに搭載可能な視線計測モジュールの開発が進展するなど、視線を計測、利用する場が実験室から日常環境へと移りつつあり、今後はデータの蓄積がこれまでとは全く異なる上昇曲線を描き、計算機資源を活用した分析が急速に進められるものと考えられます。

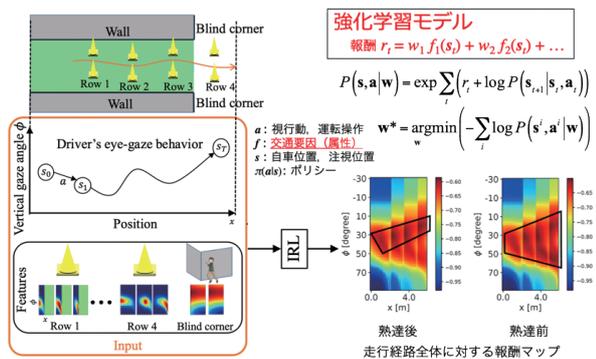


図7 逆強化学習を用いた視行動のモデル化