

令和 5 年 6 月 22 日現在

機関番号：34303

研究種目：若手研究

研究期間：2020～2022

課題番号：20K19899

研究課題名（和文）ロボットの視覚の自律性獲得のための視る行為と気配についての研究

研究課題名（英文）Research on the act of viewing/looking/seeing and signs for acquiring visual autonomy in robots

研究代表者

佐藤 啓宏（Sato, Yoshihiro）

京都先端科学大学・工学部・講師

研究者番号：00607591

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：この研究では、人間の「視る行為」のデータ収集、視覚に関連する動作のモデル化、ロボットによる行動の再現、の3つの領域を探索した。しかし、COVID-19の影響で人を対象とする実験の実施が困難であった。データ収集では、簡易防音暗室を設置し、監視カメラと環境センサーを用いてデータを収集できる装置を開発した。動作のモデル化では、自己観察を行い、基本データを取得したが、温度や湿度などの変動幅が小さく、人手でラベル付けすることが困難であるため自動ラベル付けアルゴリズムを考察した。一方で、AIの進歩に伴い研究分野が大きく進展したため、新たに脳波デバイスを用いたAIとのコミュニケーションの基礎研究を行った。

研究成果の学術的意義や社会的意義

本研究では、COVID-19パンデミックの制約下で「人間の視る行為」についてデータ収集とモデル化を試みた。また人間の自然行動を模倣するロボットの開発により、人間の行動理解とAIとの新たなコミュニケーション手法の開発を目指した。開発したロボットの設計デザインはオンラインにて公開している。さらに、脳波デバイスを用いたAIとのコミュニケーションの基礎研究を開始した。これらの成果は、今後の実験によるデータ取得によって補完されて、当初目的のように自然なロボットの振る舞いを実現できるようになる。また、大規模言語モデルの大きな進展により、その外側を担うロボットの重要性が上がっており、開発の土台となるだろう。

研究成果の概要（英文）：The study explored three areas: collecting data on the human "act of seeing," modeling vision-related behavior, and reproducing the behavior with a robot. However, COVID-19 made it difficult to conduct experiments on human subjects. For data collection, we set up a simple soundproof darkroom and developed a device that could collect data using a surveillance camera and environmental sensors. In modeling the behavior, I conducted self-observation and obtained basic data. However, we considered an automatic labeling algorithm because the fluctuation range of temperature and humidity was small and it was difficult to manually label the data. On the other hand, since the research field has greatly advanced with the progress of AI, we newly conducted basic research on communication with AI using EEG devices.

研究分野：ロボティクス

キーワード：ロボティクス

1. 研究開始当初の背景

近年、ロボットの自律性を高めるための研究が増え、特に「意識」の探求が注目を浴びている。深層学習の一部分野では単機能としての実用化が進んでいたが、一般化を目指した「教師なし深層学習」の実用化は未だ達成されていなかった。これはロボットが自律的にデータを取得し選択する能力に直結しており、人工知能研究の大きな課題であった。また、意識と身体の関係について、生物の進化の観点から研究するアプローチも存在するが、その核心に到達するのは難しいとされている。直接観測できない「意識」を工学に落とし込むのは大きな挑戦である。

そこで本研究では「見る行為」の重要性に焦点を当てる。視覚情報学は既に広く研究されている領域だが、視点移動機構を伴う能動視覚についてはまだ十分に研究されていない。生物は、光学情報だけでなく、聴覚や嗅覚、そして「気配」のような視覚以外の情報も取り扱う。「気配」を「意識下の視覚に頼らずに外界を知覚する能力」と定義すれば、感覚や記憶に基づく空間知覚が含まれる。

しかしながら、「気配」は物理的な定義が曖昧であり、見る行為と意識の関係も不明確なままである。それらを扱った研究は少なく、基礎的なデータもほとんどない状況である。そこで本研究では、微弱な外部刺激に対する人間の内的反応から生じる運動を観察し、データの蓄積を進めることを主要な課題とする。

2. 研究の目的

本研究は、「気配を感じて見る」という行為を理解し、その過程をモデル化することを目指す。そしてその結果を用いてロボットの能動視覚の自律性を向上させることを最終的な目標とする。具体的には、研究は次の3つのフェーズに分けられる。

1. 人間の「見る行為」に関するデータの収集

実験により、微弱な刺激が人間の視覚行動にどのように影響するかを調査する。

2. 視覚に関連する動作のモデル化

収集したデータを基に、視覚と運動の関連性をモデル化し、それを言語表現可能な形で記述する。

3. ロボットによる再現

2で作成した動作モデルを用いて、人間と同様の反応をするロボットを開発し、その性能を評価する。

研究の成果は、ロボットの視覚制御の自律性向上に寄与し、商業的にも暗黙の視線誘導による宣伝効果を上げる可能性がある。また、人間の視覚が視覚以外の刺激に対してどのように反応し、それがどのように考え制御されているのかを理解することで、人間の視覚の本質的理解にも貢献できる。

3. 研究の方法

三つのステップで人間の「気配を感じて見る」行為を分析し、ロボットの能動視覚の自律性を高めるモデルの開発を試みる。

初めに、簡易防音暗室で光、音、におい、風といった刺激を被験者に提示し、その反応を記録する。被験者の行動の各パラメータを観測し、視覚の運動に関する反応を抽出、分析対象とする。また、観測不能な内的要素については聞き取り調査により取得する。

次に、得られたデータに基づき、人間の内部処理機構を深層学習モデルで表現し、ブラックボックスの関数を作成する。ただし、観測不能なパラメータの存在から、動作モデルのみでの理解は不十分である。古典的な要素の切り出しと状態遷移モデルによる記述を通じて、不明なパラメータの影響を加味したモデルを作り上げる。仮説として、刺激の強度に応じて反応パターンが変わる階層的な状態遷移モデルを持つと考え、時間パターンやストーリー性のある刺激に対する「先読み」のような反応も観察する。

最後に、2で得たモデルを用いて、人間の反応を模倣するロボットを開発する。特に、6自由度のロボットアームに各種センサーを取り付けた疑似頭部を使用する。ロボットへの実装により、能動視覚の自律性向上を確認する。

これらのステップを経て、本研究は人間の視覚行為の理解とそのモデリングを進めることで、ロボットの視覚能力を高めることを目指す。

4. 研究成果

研究期間が令和 2-4 年で COVID-19 による混乱期と重なったことから、1 と 2 の人を対象とする実験はほぼ実施できなかった。さらに、最終年度の令和 4 年には、Stable Diffusion を始めとする Generative AI と呼ばれる分野が急速に発展し、特に、2022 年 11 月末に発表された ChatGPT を契機として、ロボット・AI 分野の研究方法が大きく変革を遂げた。これらの背景を踏まえ、科研としての研究期間の延長申請は見送った。また、研究後期には AI の進歩に合わせて研究分野が大きく進展することを見越して、脳デバイスを用いた研究を試験的に行っている。以上の事情を注記として、以下に研究成果をまとめる。

(1) 人間の「見る行為」に関するデータの収集

観察のための実験用簡易防音暗室(図 1)を製作し、それを研究室内に設置した。アルミフレームで骨組みを組み立て、格子を形成し、カメラと環境センサーがほぼ均等になるよう配置した。当初は複数の Raspberry Pi と各種センサーを組み合わせた計測器を自作する計画だったが、Covid-19 の影響で半導体部品の入手が困難になったため、既製品の監視カメラと環境センサーを使用することに変更した。総数 28 セットのセンサーが配置され、常時録音とデータ記録が行われている。

当初の設計図 1(a)からサイズを縮小し、図 2(b)に示すように 2.4m 四方の空間を密接に観察できるように変更した。空間はカメラの画角約 90 度に対して狭いが、複数のカメラにより観察が行われているため、観察対象はどのカメラからでもほぼ確実に観察されている。観察対象が立っているときにはこの構成が非常に狭く感じるため、座れるように変更し、簡易畳を導入した。これにより観察対象がリラックスしやすくなるという副次的な効果も期待した。

観察装置は稼働しているものの、感染症対策の観点から広範囲での実験を行う目標はまだ達成されていない。

使用しているカメラは ATOM CAM2、環境センサーは OMRON 2JCIE-BU01 である。

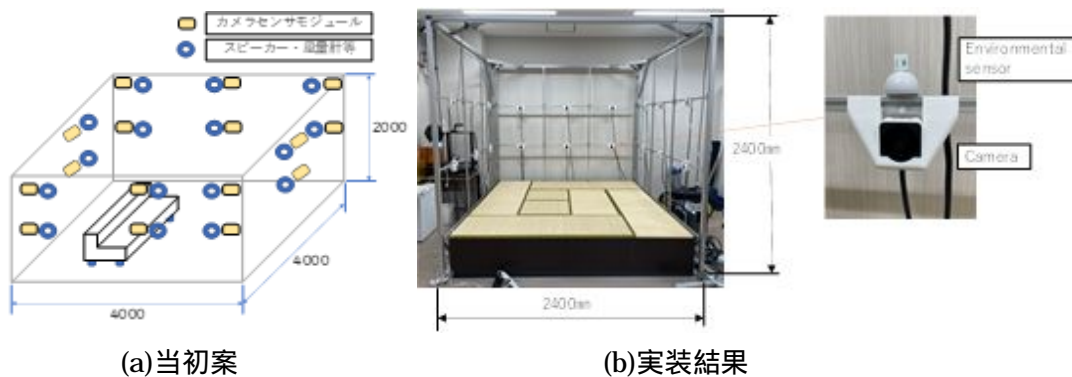


図 1. 実験用簡易防音暗室

(2) 視覚に関連する動作のモデル化

自分自身を観測対象として、基礎データをいくつか取得した。図 2 にそのときのイメージと環境データの様子を示す。図 2a は実験中の様子を表しており、被験者が長時間その場所に滞在して、わずかに動く様子とその動作が与える環境への影響を環境センサーのデータとして黒くした。そのデータ例を図 2b に示す。そこから学習用の教師データを作ろうとしたときに、温度、湿度などの環境データの変動幅が非常に小さくて人の手でラベル付けすることが困難であることが分かった。そこで、環境データのみからイベントを検出しラベル付けする方法を考え、one-class SVM を階層的に組み合わせてデータ特徴を分解していくといった random forest に似た考えで、なおかつクラス分けの根拠が分かりやすいと期待できる方法を考案して、テストした。その方法を AROB2022 という会議で発表している。しかし、視覚データから検出されるイベントと環境センサーのデータから検出されるイベントの関連性の有無を評価したところ結果は良いものではなかった。

ここで計測したデータの種類は、音量、みなし CO2、大気圧、温度、湿度、照度である。いずれの値も計測対象との距離の関数の形で反応するものの、その感度が桁違いなので正規化せずに統合することに無理があった。

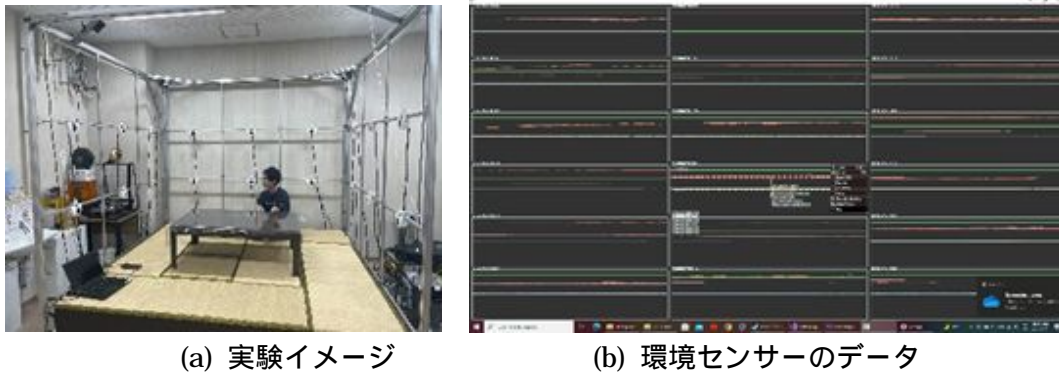


図 2. 環境センサーのデータ例

(3) ロボットによる再現

ロボットを通じて人間の行動を模倣することで、人間の無意識な自然行動を表現するモデルの評価を目指した。その観点から、人間を模倣する動作を表現できるロボットを製作した。胴体部分は、図 3(a)の初期の案通り、アームロボットで置き換え、頭部部分を新たに製作した。人間の顔の大きさや目と耳の位置を可能な限り人間に近づけるよう配慮し、図 3(b)の実装結果のように設計した。各眼部分は独立したパンチルト機構を有し、独立したカメラを装備している。耳に相当する部分には、それぞれ内蔵マイクが設置され、耳介に相当する部品の影響を受ける音が計測される。また、首部分にはパンチルトと回転の機構が設けられ、3自由度の方向制御が可能である。

さらに、このロボット自体はオープンソースとして公開することを前提に設計した。そのため、ほぼ全ての部品は 3D プリンターで作成可能であり、また安価に入手可能な電子部品で構成されている。このロボットの設計データは、<https://github.com/cvl-kuas/20k19899> にて、Creative Commons License 4.0 の下で公開している。

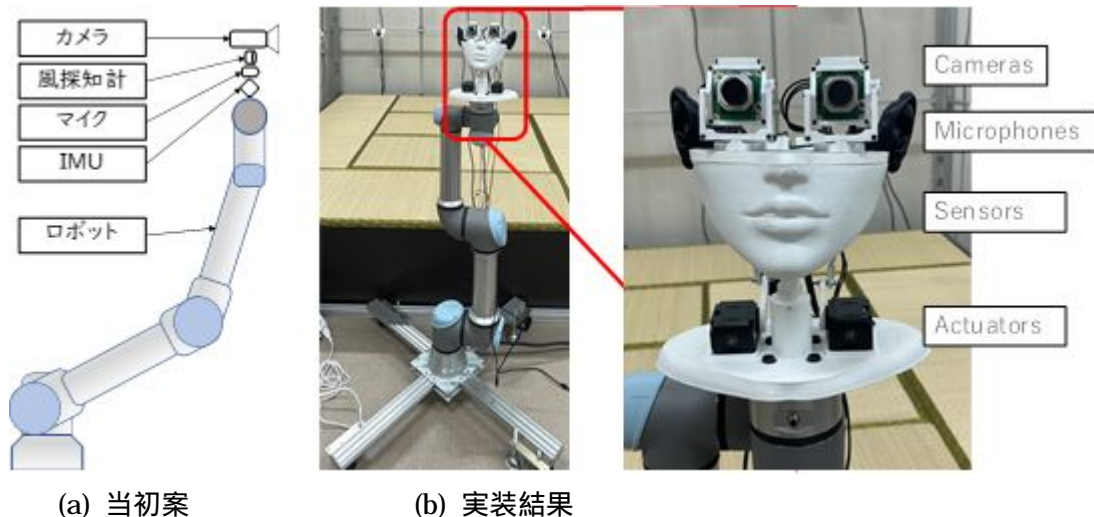


図 3. モデル検証用ロボット

(4) 脳波デバイスを用いた AI とのコミュニケーション

研究の終盤では、AI が脳活動に迫る研究が大きく進展したため、今後の研究の方向性を見据え、当初の予定には含まれていなかった脳波デバイスを用いた AI とのコミュニケーションについての基礎研究を行った[2]。AI と脳が類似性を持つことが明らかになりつつある。しかし、お互いにコミュニケーションを取ろうとすると、文字や音声だけでは不十分であり、脳波デバイスもまた、AI の進歩に対して不十分であることが明確になった。インターフェースの能力が不足していると、高度な対話を実現するためにはシステムを複雑化しなければならないというジレ

ンマが浮かび上がった。

[1] Y. Sato, “ Basic Experiments for Erasing from Memory of Audio and Visual Data based on Somatosensory Imitation by Integrated Sensors, ” AROB2022.

[2] E. Chatani, Y. Sato, “ A Study of Communication Formats for Human and Large Language Models Using EEG Devices, ” FIT2023.

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 Eisuke Chatani, Yoshihiro Sato
2. 発表標題 脳波デバイスを用いた人と大規模言語モデルのコミュニケーション形式の検討
3. 学会等名 第22回情報科学技術フォーラム
4. 発表年 2023年

1. 発表者名 Sato Yoshihiro
2. 発表標題 Basic Experiments for Erasing from Memory of Audio and Visual Data based on Somatosensory Imitation by Integrated Sensors
3. 学会等名 27th International Symposium on Artificial Life and Robotics (国際学会)
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------