

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年6月3日現在

機関番号：14303

研究種目：若手研究(B)

研究期間：2010～2012

課題番号：22700272

研究課題名（和文） ワーキングメモリの映像的表現とロボットの語意獲得への応用

研究課題名（英文） Visual Expression of Robot's Working Memory and Application to Lexical Acquisition

研究代表者

尾関 基行 (OZEKI MOTUYUKI)

京都工芸繊維大学・工芸科学研究科・助教

研究者番号：10402744

研究成果の概要（和文）：本研究では、ロボットに埋め込んだディスプレイを通して、ロボットのワーキングメモリの中身を映像で他者に伝える“シースルーワーキングメモリ”というコンセプトの提案を中心に、ワーキングメモリを実現するための視覚的注意モデル（2種類）や長期記憶からの連想手法の構築、ロボットの思惑を適切に伝えるための映像エフェクトの検討・評価を行った。

研究成果の概要（英文）： This study proposed the concept “See-through working memory”, which conveys the internal state (contents of the working memory) of a robot to a user in the form of visual imagery. For realizing this concept, I developed two visual attention models and an association method from long-term memory, and examined visual effects for appropriate communication of robot's thinking.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	1,300,000	390,000	1,690,000
2011年度	1,200,000	360,000	1,560,000
2012年度	500,000	150,000	650,000
年度			
年度			
総計	3,000,000	900,000	3,900,000

研究分野：画像・映像処理・パターン認識・学習

科研費の分科・細目：情報学・認知科学

キーワード：(1) ワーキングメモリ(2)ヒューマン-エージェントインタラクション

1. 研究開始当初の背景

本研究の提案は、ヒトのワーキングメモリの計算モデルをロボットに組み込み、更にそのワーキングメモリの中身を人に伝わりやすい形で映像的に表現することである。遙かな昔から現在に至るまで、「心」や「意識」は我々人間にとって常に最大の関心事であり、その研究の学術的価値に疑う余地はない。そのアプローチは (a) 調べて確かめる立場と (b) つくって確かめる立場の二つに大きく分

かれ、これらは相互に補い合って当該分野の発展に貢献している。(b)のアプローチは更に二つの立場に分かれる。一つは、(b-1) 脳の精緻なモデルをつくり、シミュレーションによって実際の脳の働きを予測・検証しようとするもの。もう一つは、(b-2) 応用しやすいように単純化された脳や心のモデルをロボットなどに取り入れることで、人とのコミュニケーションを改善しようとするものである。本研究は最後の(b-2)の立場に位置づけられるが、映像的表現によるシミュレーション

によってヒトの心にアプローチするという点では(b-1)の立場でもある。

2. 研究の目的

本研究の全体構想は、ロボットに組み込んだワーキングメモリの中身を映像的表現によって他者に伝える“シースルーワーキングメモリ”というコンセプトを提案し、人とロボットの新しいコミュニケーションの可能性を探るとともに、ワーキングメモリを視聴して確かめるという新しいアプローチによって心や意識の研究に取り組もうとするものである。その第一歩として、本研究課題では、(1) ワーキングメモリの計算モデル（特に視覚に関するワーキングメモリ）をロボットに実装し、人に伝わりやすい形で映像的表現すること、及び、(2) ロボットの語意獲得を例題として、シースルーワーキングメモリの有効性を示すことを目標とする。本研究の最大の特徴はワーキングメモリの中身を見せることに主眼を置いていることであり、心のモデルをロボットの中に隠してしまう従来の研究とは異なり、中身を見せることでロボットに心を帰属させようとする点で独創性がある。

3. 研究の方法

本研究では、(1) ワーキングメモリの計算モデルを構築すること、及び、(2) その中身をロボットの意識（思考）として表出するのに適した映像表現についてそれぞれ調べた。語彙獲得についても独立して研究を進めていたが、シースルーワーキングメモリとの統合は今後の課題となっている。ここでは、(1) と(2)について具体的な方法を述べる。

(1) ワーキングメモリの計算モデルの構築

申請当初の計画では、Baddeley&Hitch のワーキングメモリモデルを参考に、視覚（視空間スケッチパッド）・聴覚（音韻ループ）・エピソードバッファの各ワーキングメモリを構築し、最終的に統合する予定であったが、視空間スケッチパッドを中心に研究を進め、音韻ループについては基礎的な検討を実施するに留まった。エピソードバッファについては視覚に関する長期記憶から単純な連想に基づいて情報を引き出す手法を提案した。これらのうち、ここでは視空間スケッチパッドについての取り組みについて詳しく述べる。

視空間スケッチパッドを構築するためにまず必要なことは、ワーキングメモリに一時的に保持する視覚情報（以降、“イメージ”と呼ぶ）を外界もしくは長期記憶から選択する

機能である。外界からのイメージの選択は“視覚的注意”によって行われる。この視覚的注意の計算モデルについて、本研究では①と②の二つの手法を提案した。また、長期記憶からのイメージ選択手法として、「ワーキングメモリに同時に格納された」という共起情報に基づく方法（③）を提案した。以下、それぞれについての方法を簡潔に述べる。

① 粒子フィルタを用いたスポットライト群メタファに基づく視覚的注意モデル

我々は、視点を右に動かしたいと思えば、眼球を右に動かし、その先にある注意の惹かれる領域上で視点を止める。しかし、このような注視点の移動制御は、従来の視覚的注意モデルでは重視されてこなかった。その理由の一つとして注視点の移動は（注意の計算モデルとは切り離して）眼球運動で行うべきという考えがあるが、ヒトの注意は常に視野の中心に位置しているわけではなく、また、眼球運動の機能を備えたロボットは非常に高価である。よって、視覚的注意モデルの機能として、注意の移動制御を容易に行うことのできる仕組みが必要である。

この提案手法の核となるアイデアは、“スポットライト群メタファ”と名づけた新しい注意の捉え方であり、視野上の注意の分布を「多数の小さなスポットライトの集まり」とみなす。このメタファに基づいたモデル化により、移動方向を与えるだけで注意を連続的に移動することができ、かつ、その注意を向ける領域がボトムアップ注意に沿って変形・分割されるという特徴を兼ね備える。本研究では、スポットライト群メタファに基づく計算モデルの一つとして、粒子フィルタを用いた実装方法を取りあげている。

② 注視点からの奥行情報を用いた視覚的注意モデル

（①の研究も含め）従来の視覚的注意モデルは二次元情報しか用いないものが大半であり、奥行情報を使用したものも「近い距離にある領域ほど注意を引く」や「人の手前にある領域は注意を引く」といったモデルしか提案されていない。人間の場合、ある領域を注視すると、その領域と奥行の異なる領域はボケて見える（被写界深度が浅い）。そのため、現在の注視点から奥行の離れた領域が注意を引き付ける度合いは小さくなる（奥行距離の大きくことなる領域には注意が引かれにくい）と考えられる。これを計算モデルに取り入れるには「現在の注視点からの奥行距離」を考慮する必要があり、そのようなモデルはこれまでに提案されていない。

本研究では、①の視覚的注意モデルから求め

られた注視点の奥行距離と、入力画像（距離画像）の奥行距離の差を算出し、各領域が注意を引く強さ（顕著度）をその差に応じて決める手法を提案した。この奥行距離の差に基づく顕著度は①の視覚的注意モデルの構成要素の一つとして取り込まれ、現在の注視点から奥行の離れた領域に注視点が移動しにくくなる。また、その「移動しにくさ」の度合いは **depth-of-focus** というパラメータで調整することができる（トップダウン注意制御）。

③ 長期記憶からのイメージ選択

長期記憶からのイメージ選択のうち、（人間でいうところの）意識的に行われているものは、すでに提案されている種々の検索手法を用いて行えばよい。本研究では、（人間でいうところの）無意識的に行っている連想的なイメージ選択の手法を提案した。人間は何かを見ると、特に目的がなくても、それに関連する物事を思い出してワーキングメモリに置く。それは単純に色・形が似ているものであることもあれば、それと一緒に使うもの、それを使っているシーンなどのこともある。

提案手法では、これらの連想関係を「以前に同時にワーキングメモリに保持された」という関連性で紐付ける。共起性に基づく手法としては、入力カメラに同時に映りこんだ物体を関連付ける方法などが提案されているが、ワーキングメモリ内での共起に絞ることで、比較的重要なものの関連性だけが記録される。また、ワーキングメモリには長期記憶からもイメージが呼び出されるので、外界にあるものと長期記憶にあるものの関連性も紐付けることができる。

(2) シースルーワーキングメモリのための映像表現

(1)で述べた各手法により、ワーキングメモリにイメージが呼び出され、保持される。保持されるのは最大4つのイメージまでとし、時間の経過や新しいイメージが入ってくることにより、最も古いイメージがワーキングメモリから外される。ロボットに搭載された人工知能は、このワーキングメモリに保持されている数個のイメージを「思考の焦点」に置いて種々の処理を行うことができる。

本研究の目的は、このワーキングメモリに保持されたイメージを、ロボットの胴部などに埋め込まれたディスプレイに映像として表示して、相手に心の内を見せることである。実装上、イメージは1枚の画像として管理されており、本研究ではこれをスライドショー形式の映像として出力する手法を提案した。通常は、ワーキングメモリ内の最新のイメー

ジが映像として出力される。最新のイメージが更新されると、スライドショーを進めるように画像が切り替わる。ロボットが相手に何かを伝えようとしたり、思考中や検索中のときには、ワーキングメモリ内の4つのイメージを映像として順番に繰り返し表示する（リピートモード）。

映像を出力するに当たって検討すべきことは、スライドショー形式で表示される画像が

- ・何を伝えようとしているのかわかること
- ・ロボットの心象であることがわかること

である。本研究では、イメージを切り替える際のエフェクト（フェードやワイプなど）によって、これらの印象が相手にうまく伝わるよう操作することを検討した。

4. 研究成果

研究方法で述べた各項目について、その研究成果を以下にまとめる。

(1) ワーキングメモリの計算モデルの構築

① 粒子フィルタを用いたスポットライト群メタファに基づく視覚的注意モデル

実装した提案手法を従来の二つのメタファ（スポットライトとフィルタ）に基づく手法との比較した。その結果、提案手法が、移動方向を与えるだけで注意を連続的に移動することができ、かつ、その注意を向ける領域がボトムアップ注意に沿って変形・分割されるという特徴を持つことを実験的に確認した。また、注目物体推定において、従来の手法に比べて、比較的幅広いシーンに対応できることを示した。

② 注視点からの奥行情報を用いた視覚的注意モデル

まず、人の注視点が仮説どおりの振る舞いをしていることを実験的に確認するため、テーブルに9つの物体を並べ、タスクなし・組合せタスク・探索タスクの三つの状況での視線移動を調べた（12名）。その結果、全体の視線移動の内、直前の注視点から奥行距離の近い物体への視線移動の割合がチャンスレベル（26.7%）に比べて有意に高く（ $p < 0.05$ ）、また、その割合が与えられたタスクによって変化することが確かめられた。次に、Kinect for Windows を用いて実装した提案手法に同じシーンの入力を与え、視線移動を調べたところ、人の視線移動のパターンと似通った結果が得られた。また、depth-of-focus を変化させることにより、3つの状況での視線移動パターンをほぼ再現することができた。

③ 長期記憶からのイメージ選択

長期記憶として、研究協力者1人の1週間の行動を洗い出し、日常的に使われている物体を約190個用いてデータベースを作成した(同時にワーキングメモリに保持されたことによる関連度が付与されている)。タスクとしては、例えば、アイロンをかける、歯磨きをする、りんごを切る、ホットケーキを食べる、パフェを食べるなどである。この長期記憶に対し、いくつかのタスク(コーヒーを淹れるなど)を目の前で起こっている事象として提案手法に入力し、長期記憶からどのような連想イメージが引き出されるかを観察した。その結果、典型的な行動パターンで用いられる物体が順に呼び出されたり、スプーンの代わりに使えるフォークが呼び出されたりなど、単純な連想ではあるが、役立ちそうな情報が取り出せる可能性を見出すことができた。

(2) シースルーワーキングメモリのための映像表現

まず、ロボットが伝え分ける必要のある内部状態(心の内)を、観察・思考-意思・思考-予想・思考-考察・検索-類似・検索-共起・検索-順序・想起の8つに分け、それぞれをうまく相手に伝える映像エフェクトを実験的に調べた。プレゼンテーションソフトウェア(Mac OS X, Keynote)から13種類(21パターン)の画像切替エフェクトを選び実験を行ったところ、特定のエフェクトが特定の内部状態だけを強めるという1対1対応は見られなかった。一方で、13種類のエフェクトを「移動性のあるもの」と「曖昧性のあるもの」という観点から分類すると、一部の内部状態との対応が見られそうであった。

そこで、伝えるべきロボットの内部状態を、観察・意図・予期・思考の4つに絞り、移動性の有無と曖昧性の有無を組み合わせた4つの映像エフェクトとの関係を調べた。なお、4つの映像エフェクトは自作した。その結果、前回の実験と同じく、特定の内部状態だけを選択的に強めるエフェクトはなかったが、各内部状態を伝える際に用いるべき映像エフェクトを実験的に確認することができた。

最後に、胸部にディスプレイを内蔵したロボットを自作し、ワーキングメモリの中身を映像で表示する実験を行った。視覚的注意モデルや連想手法の誤検出等を避けるため、ワーキングメモリの中身は人手で予め用意した。カード作成・文具の分類・プラモデルの組み立ての三つのタスクを人が実施しているときに、ロボットがそれを補助するという状況を設定した。ロボットが出力できる行動とし

て、動作のみ(指を差すことと、モノを渡すこと)、動作+音声、動作+映像の3パターン用意し、ロボットの意図が通じたか、使い続けたいと感じたかなどをアンケートで評価した。その結果、動作+映像(提案手法)は動作+音声と同程度に意図を伝えることが可能であり、また、最も使い続けたいという評価を得た。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計2件)

- ① 尾関基行, 岡 夏樹, 心象映像によるエージェントの内部状態の表出, 人工知能学会誌, 査読無, 特集「人を動かす HAI」, 28 巻, 2 号, pp.302-309, 2013.
- ② 尾関基行, 柏木康寛, 井上茉莉子, 岡 夏樹, 粒子フィルタを用いたスポットライト群メタファに基づく視覚的注意モデル, 電子情報通信学会論文誌, 査読有, Vol.J95-D, No.11, pp.1909-1918, 2012.

[学会発表](計11件)

- ① 藤原菜々美, 小川貴弘, 尾関基行, 岡 夏樹, ロボットの内部処理を映像で伝える際のトランジション効果の印象評価, 信学技法, HCS2012-55, pp. 7-12, 高知, 24 Jan. 2013.
- ② 小川貴弘, 藤原菜々美, 尾関基行, 岡 夏樹, 連想に基づいた心象映像表出によるエージェントの内部状態表示, 第26回人工知能学会全国大会, 301-0S-3a-5, 山口, 14 Jun. 2012.
- ③ 藤原菜々美, 尾関基行, 岡 夏樹, 心象映像によるエージェントの内部状態表現, 第25回人工知能学会全国大会, 2D1-0S5a-4, 岩手, 2 Jun. 2011.
- ④ M. Ozeki, Y. Kashiwagi, M. Inoue, N. Oka, Top-Down Visual Attention Control Based on a Particle Filter for Human-Interactive Robots, Proceedings of 4th International Conference on Human System Interaction, pp.188-194, Yokohama, Japan, 20 May 2011.
- ⑤ 尾関基行, 藤原菜々美, 岡 夏樹, シースルーワーキングメモリ: エージェントの心象表示による新しいコミュニケーションに向けて, HAI シンポジウム, 1B-5, 神奈川, 12 Dec. 2010.

6. 研究組織

(1) 研究代表者

尾関 基行 (OZEKI MOTOYUKI)
京都工芸繊維大学・工芸科学研究科・助教
研究者番号：10402744

(2) 研究分担者
()

研究者番号：

(3) 連携研究者
()

研究者番号：