

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 24 年 5 月 24 日現在

機関番号：22604

研究種目：若手研究（B）

研究期間：2008～2011

課題番号：20700185

研究課題名（和文）

人とのインタラクションに基づくロボットの状況依存型ジェスチャーの獲得

研究課題名（英文）

Situation-based Robot Gesture acquisition based on human robot interaction

研究代表者

下川原 英理（SHIMOKAWARA ERI）

首都大学東京・システムデザイン研究科・助教

研究者番号：00453035

研究成果の概要（和文）：

自然な対話を可能にするためには、相手の意図を汲み取り、かつ相手にわかりやすく伝えることが重要である。身体表現が可能なロボットにとっては、人が理解しやすいような身振り手振りを踏まえたインタラクションが必要である。本研究では、ジェスチャーに着目し、ロボットのジェスチャー獲得と、環境状況を考慮したインタラクションシステムを開発した。

研究成果の概要（英文）：

This research focused on natural interaction considering gesture interaction and situation. Recently robots can express gesture by these body and arm like a human. I focused on acquiring gestures for a robot, and interaction by using the gestures. However, gestures have many meanings depending on situation. In this research, I developed situation based gesture acquisition system and interaction system.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008 年度	1,400,000	420,000	1,820,000
2009 年度	352,731	105,819	458,550
2010 年度	647,269	194,180	841,449
2011 年度	900,000	270,000	1,170,000
年度			
総計	3,300,000	989,999	4,289,999

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：知能ロボット、ジェスチャー・インタラクション

1. 研究開始当初の背景

ロボット利用のターゲットは、工場などの閉鎖された環境での決まったタスクをこなすものから、一般家庭や公共の場といった環境が常に変動する場へと広がりを見せている。事実さまざまなロボットが福祉施設や一般家庭へ向け開発・販売され始めている。日常生活に溶け込み人と共存するロボットには自然に人と対話し状況に適したサービスを提供することが求められる。ロボットのみならず、情報機器や携帯端末などさまざまな

機器の多機能化・高機能化が進んでいるが、それに比例して操作が複雑になり使用手順・方法を習得するのが難しくなっている。直感的に操作・理解でき、人と対話するような感覚で対話可能なシステムが必要不可欠である

自然な対話を可能にするためには、相手の意図を汲み取り、かつ相手にわかりやすく伝えることが重要である。人と人が対話するような感覚で人とロボットが対話することが、一般家庭に受け入れられる重要な要素であ

ろう。特に身体表現が可能なロボットにとっては、人が理解しやすいような身振り手振りを踏まえたインタラクションが必要である。自然なインタラクションに加え、変化する環境に合わせ、状況を考慮し人とインタラクションしながらサービスを提供することが、これまでエンターテインメントの分野でのみ一般家庭に受け入れられたロボットが、役に立つものとして一般的に受け入れられる基盤となる。状況に合わせたサービスを提供するためには、人の動きや音声だけでなく環境を認識し何に対して行われた動作なのか現在の場所はどこか、といった環境情報を考慮する必要がある。そこで本研究『人とのインタラクションに基づくロボットの状況依存型ジェスチャーの獲得』では人の動きと状況から適したサービスを認識し、さらにロボット自身がジェスチャーによってリアクションを返しながら、インタラクティブにサービスタスクを実行することで、人とロボットの自然な対話を目指す。

ジェスチャーは人と人との対話において、初期に獲得されかつ対話を円滑に進める重要な役割を担っている。言語の異なる人同士であっても、ボディランゲージによってある程度コミュニケーションを取ることができる。しかし人の動作は、同じような動きであっても対象物や状況によって、意味が異なる。環境を考慮して人の動作を理解すること、人にわかり易い動作で対話することは、日常生活において人と共生するシステムにとって重要な技術である。

2. 研究の目的

ユーザのジェスチャーに応じて、環境に付加された知識を利用し、ロボット自身もジェスチャーを交えてインタラクションを行うロボットシステムを構築するためには、どのような動作がどのような状況で行われるか関連付ける必要がある。さらに人を見守ることによって得た動作と状況とのデータベースを基に、ロボット自身が状況に合わせた動作を行わなければならない。そのためには以下の(1)(2)について研究を行った

- (1) 人と対話するためには、人にわかり易いリアクション表現が必要である。1)において得た動作に関する知識を利用し、状況に合わせた動作表現を行うことによって、人の感性に訴えたインタラクションが可能となる。人の注意を引いたり、向けさせたりする場合に人の動作に関する知識から自分の動作として学習することによって、ロボットなりに表現し人-人の対話に近い感覚をユーザに与えることが可能になる。
- (2) 人の動作は同じ動きであっても状況に

応じて意味が異なる。また同じ意味の動作であっても人によって異なった動きをする場合もある。人の動作と状況からジェスチャーの意味を認識し、それに応じてサービスを提供する。

3. 研究の方法

(1)本研究では、様々な状況の中で「ジェスチャーを行う個人」「インタラクションを行う二者間の距離」「インタラクションを行う相手との親密度」に着目し、それらを人とロボットとのジェスチャーインタラクションに適用した。また、人は相手のジェスチャーの意味を理解するとき、そのジェスチャーの特徴的な部分に注目しデフォルメし理解する。以上の事から、人とロボットによるジェスチャーインタラクションから、人が察ししやすいジェスチャーを獲得させるためには、状況に依存したインタラクションからロボットのジェスチャーをデフォルメさせることが望ましい。

本研究ではロボットに察ししやすいジェスチャーを状況に依存したデフォルメを行う。しかし、デフォルメする事によってジェスチャーの意味が失われることや、他の意味を持つジェスチャーに変化することはあってはならない。そこで本節ではデフォルメを行うことでもジェスチャーの意味情報が保存される際に重要な特徴について、ジェスチャー分類の観点から決定した。本研究では図1に示すように、喜多のジェスチャー分類と、広瀬らが3種類に再分類したジェスチャー分類

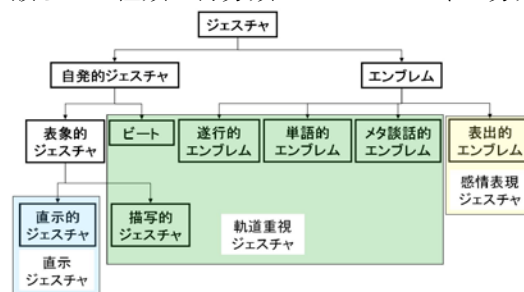


図1 ジェスチャー分類

を用いた。

本研究では、対話型進化計算を用いてジェスチャーの進化を行う。ユーザの負担を減らすためパラメータを変化させる事でその想起範囲の広さを設定できるカオス進化によって対話型進化計算を行う。カオス進化は少数の個体から想起が可能である。また、一般的な対話型進化計算で用いられる遺伝的アルゴリズムの問題点である局所解に陥るという点を、そのカオス性から回避できる。

しかし、実機ロボットによる進化計算は、そのジェスチャー生成に時間がかかる。また、生成したジェスチャーを表示するためには物理的に実機の数を増やす必要がある。そこで本研究では複数固体を評価する際のユーザの負担を減らすため、バーチャルロボットを用いた。

(2) ジェスチャーの意味を理解し、状況に応じたサービスを提供するために、環境の情報を利用したインタラクションシステムを開発した。ジェスチャーによる直視的指示（指さし）と音声による指示を併用した、マルチモーダルインタラクションに、空間の情報を付加することによって、直感的に利用可能なインターフェースとした。

ジェスチャーの認識にはステレオカメラによる OpenCV による SURF 局所特徴点抽出を用いた非拘束な計測とした。音声認識には Julius を用いた。環境の情報を人とのインタラクションから取得し、次のインタラクションへと利用することで、環境情報の取得に必要な大規模センシングが不要となった。さらに RT-Middleware をによるモジュールとすることで、再利用性の高いものを目指した。

4. 研究成果

(1) ジェスチャーの獲得

本研究では 3 つの実験を行った。1 つ目のジェスチャー獲得実験では、被験者 4 人に見真似学習と対話型進化計算によってバーチャルロボットにジェスチャーを獲得させた。また獲得したジェスチャーをデフォルトジェスチャーとし、新規被験者 2 人によるジェスチャー獲得実験を行った。

2 つ目の察しやすさ比較検証実験では、ジェスチャー獲得実験を行った被験者 4 人でバーチャルロボットと実機ロボットによるジェスチャーの察しやすさの比較検証を行った。また、ジェスチャー獲得実験を行っていない新規被験者 4 名を加えた被験者合計 8 名で、状況に依存し獲得したジェスチャーと、そうでないジェスチャーの察しやすさの比較検証をおこなった。

(1)-1 ジェスチャー獲得実験

システムはユーザのジェスチャーを、人の動作認識モジュールにより、頭と手の三次元座標を時系列で取得する。ユーザの正面 250cm の位置に 50 インチプラズマディスプレイが設置してある。50 インチプラズマディスプレイにはバーチャル空間とバーチャルロボットが表示してあり、ユーザはそのバーチャルロボットに対してジェスチャーを行う。バーチャルロボットとバーチャル空間は、本研究で着目した状況のうち「インタラクションを行う二者間の距離」が表現してある。本研究ではパーソナルスペースを参考に近距離 100cm、中距離 300cm、遠距離 500cm の三

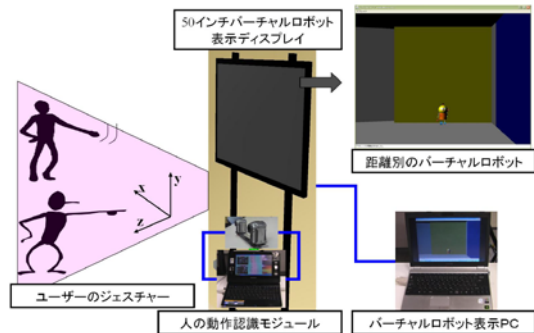


図 2 ジェスチャー獲得実験システム構成

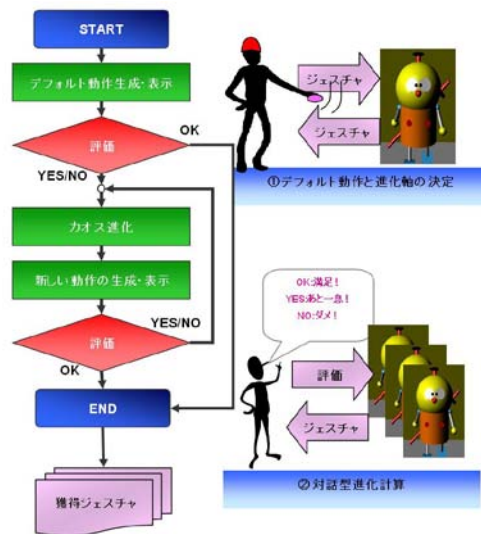


図 3 対話型進化計算のフローチャート

段階に分けた。

進化計算はカオス進化によってバーチャルロボットのジェスチャーのデフォルメを行う。ユーザはまず、バーチャルロボットに対し、「呼ぶジェスチャー」「別の場所を指し示すジェスチャー」を行い、バーチャルロボットが見真似る。ユーザはバーチャルロボットが見真似たジェスチャー（以下デフォルトジェスチャーと呼ぶ）に対し、OK（満足）、YES（もう一息）、NO（ダメ）の 3 段階評価を行う。OK はそのジェスチャーをバーチャルロボットに獲得させ、対話型進化計算を終了する。YES の評価ではカオス進化する想起範囲を狭めて進化させ、新しいジェスチャーを生成・表示する。NO の評価ののでは想起範囲を広げて新しいジェスチャーを生成・表示する。OK の評価を与えるまで進化と評価を繰り返し、ジェスチャーを獲得させる。「インタラクションを行う相手との関係」は親密度として、大（顔なじみ）、中（2、3 回の面識がある）、小（初めて会う）の 3 段階に分けた。

実験は被験者 A, B, C, D の 4 名で行った。

被験者4名がバーチャルロボットに獲得させた「人を呼ぶジェスチャー」の親密度小における「インタラクションを行う二者間の距離」別のバーチャルロボットの手の軌跡について顕著な2名の例を示す。ただし図の原点はバーチャルロボットの肩である。

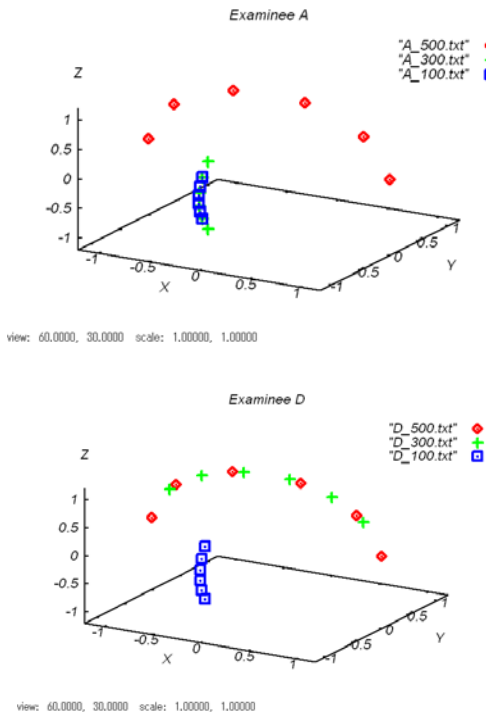


図4 被験A(上)と被験者D(下)がバーチャルロボットに獲得させたジェスチャーの軌跡

実験結果から被験者ごとに違うジェスチャーを獲得させていることがわかった。着目した状況の中で「ジェスチャーを行う個人」に着目してジェスチャーを獲得させたことが有効といえる。

また、近距離と中距離が近いジェスチャーになった被験者と、中距離と遠距離が近いジェスチャーになった被験者に分かれた。よって「インタラクションを行う二者間の距離」に着目した事が有効であったといえる。

また、親密度による変化は親密度が大きくなるにつれてジェスチャーの軌道が小さくデフォルメし、速度は速くデフォルメする被験者と遅くデフォルメする被験者に別れた。「別の場所を指し示すジェスチャー」は状況の違いによる大きな変化は見られなかった。

(1)-2 察しやすさ比較検証実験

ジェスチャー獲得実験を行った被験者4人でバーチャルロボットと実機ロボットによるジェスチャーの察しやすさの比較検証を

行った。実験はバーチャルロボットを表示するディスプレイと実機ロボットを被験者の正面等距離に設置し近距離、中距離、遠距離の各距離で同時2台のロボットがジェスチャーを行った場合の察しやすさをアンケートにより評価した。被験者はジェスチャーを獲得させた被験者A~Dの4名である。その結果、すべての距離において、「実機ロボットの方が気づきやすかった」という回答が優位に多かった。

次にジェスチャー獲得実験を行っていない新規被験者4名を加えた被験者合計8名で、状況に依存し獲得したジェスチャーと、そうでないジェスチャーの察しやすさの比較検証をおこなった。実験は「実機ロボットの方が気づきやすかった」という結果に基づき、2台の実機ロボットを用いて行った2台のロボットのうち1台が状況にあったジェスチャーを行い、もう1台は別の状況での動作を行う。どちらが状況にあったジェスチャーを行うかはランダムである。被験者にはより察しやすいジェスチャーを行ったロボットを選択してもらう。評価は、状況にあったジェスチャーを行ったロボットを選んだ場合を正解とした。試行は近距離・中距離・遠距離で各10回行った。結果を表1に示す。

表1 状況依存のジェスチャーの察しやすさ検証実験

	教示者(4名) 正解数/回数 (正答率)	非教示者(4名) 正解数/回数 (正答率)
近距離	37/40 (92.5%)	32/40 (80.0%)
中距離	36/40 (90.0%)	29/40 (72.5%)
遠距離	38/40 (95.0%)	36/40 (90.0%)
合計	111/120 (92.5%)	97/120 (80.8%)

状況依存のジェスチャーのほうが察しやすという結果が得られた。動作教示者ほど高くはないが、動作非教示者でも8割を越える結果が得られ、自分自身が教示していない状態で、他ユーザの状況依存ジェスチャーを提示することが有効と言える。全体として中距離の察しやすさが他距離より低く、これは前節考察しているように、個人によってジェスチャーの変移点異なる事からと考察できる。

(2)環境状況に依存したインタラクション

ジェスチャー認識と音声認識を併用したマルチモーダルインタラクションを実現するために、まず、ジェスチャーと音声のタイミングの関連性について実験を行った。次にマルチモーダルインタラクションを利用し、人との対話から環境情報を収集し、インタラクションに利用するシステムを開発した。このシステムは、ジェスチャー認識モジュール、音声認識モジュール、環境情報統合インタラクション生成モジュール、から構成されており、RT-MiddlewareによるRTコンポーネントとして構築することによって再利用性の高いものとした。またデータベースにはSQLを用いた。

(2)-1 ジェスチャー認識と発話のタイミング実験

本研究で開発したジェスチャー認識システムによる直示的指示(指さし)の認識の際、ジェスチャーが継続して行われている状態で、単位時間毎に指さし位置を出力することが可能である。しかし、人の指さしは必ずしもある物体1点を指し示しているわけではなく、その周囲を指し示しているものである。今回開発したシステムでは、1.3m離れた対象物を指さした場合、実際のオブジェクトとの平均で291[mm]、最大の誤差が417[mm]であった。

さらにジェスチャー認識と発話のタイミングについて実験を行った。ジェスチャーと発話は必ずしも同時ではなく、むしろ遅延が発生することが指摘されている。そこで、は被験者5名に対して、2つの対象物を順番に指差しながら発話してもらった。実験の結果、発話と関係している指差しのタイミングは発話を基準として前0.8[秒]、後0.3[秒]であることがわかった。この知見から、音声認識モジュールの発話認識の前0.8[秒]から認識後0.3[秒]の動作認識モジュールを指差しの座標として利用することとした。単位時間内に2箇所以上入力指差し位置が検出あった場合は、それらの座標の平均座標を指差し座標として扱うこととした。

(2)-2 環境状況に依存したマルチモーダルインタラクション

マルチモーダルインタラクションモジュールではまず、先に述べたように、動作認識による指差しと音声認識による発話を受け取って、それぞれの認識の結果を合わせて出力する。環境情報としてオブジェクトの位置、色、形状といった情報をデータベースとして保持している。このデータベースを参照し、指差しと発話単語の内容とを合わせて、環境情報と統合して人の指示を判断する。環境情報との統合のために、基本機能として登録・

削除・検索を実装した。また、事例として、指示によって家電の操作を行う実験を行った。

まず指差し範囲からデータベースを検索し、対象物の位置座標から範囲内に指差しが含まれるものを指示物として保持する。そして、指示物の数が0であれば指差している位置に何も存在しない、1であれば指示物が対象物であると確定し、2つ以上であれば対象物に関してより詳細な情報をユーザに問い合わせる。

実験の1例として、家電操作について述べる。家電はネットワークに接続することによって、オブジェクト名や特徴はデータベースに登録されOn/Offは可能になるが位置は登録されていない。初期条件として、データベース上は扇風機とライト(白)並んで設置されている。しかし、実際には扇風機を片付けてしまい、またライト(黒)をもう一つ追加で設置したという状況を想定した。

ユーザがライトを指差し「つけて」と発話をしたのに対し、ロボットは指差し範囲に2つの対象物があると認識し「扇風機ですか?ライトですか?」と問い合わせた。ユーザは「ライト」と答え、続くインタラクションによってデータベースから扇風機を削除した。

次にライト(黒)を指さし「つけて」と発話すると、データベース上では指差し位置にはオブジェクトが存在しないため、ロボットは「何をですか?」と問い合わせた。「ライト」というユーザの答えに対し、「ライトが複数あります。特徴を教えてください」と問い合わせ「黒」という答えを利用し、ライト(黒)を点灯させるとともに、位置情報を更新することが確認できた。

以上のように、環境状況を考慮したマルチモーダルインタラクションシステムは、人とのインタラクションを積極的に活用することによって、人と人との対話に近いインタラクションが可能となった。

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 Yihsin Ho, Tomomi Shibano, Eri sato-Shimokawara, Toru Yamaguchi, Data Mining using Human Motions for Intelligent System, Journal of Advanced Computational Intelligence and Intelligent Informatics 査読有
- ② Yihsin Ho, Kazumasa Murakami, Tomomi Shibano, Eri sato-Shimokawara, and Toru Yamaguchi, An Intelligent Human Support System using Motion Pattern and Environment Information, Journal

of Advanced Computational
Intelligence and Intelligent
Informatics 査読有

[学会発表] (計 12 件)

- ① Eri sato-Shimokawara, Shoichiro Sakurai, Toru Yamaguchi, Human-Robot interaction using indicating behavior for service robot, The 5th International Conference on Ubiquitous Robots and Ambient intelligence (URAI2008)
- ② 大川 智, 下川原 英理, 山口 亨, ロボットサービスシステムの為の状況依存型ジェスチャーインタラクション, 第9回(社)計測自動制御学会システムインテグレーション部門講演会 (SICE SI2008)
- ③ Takahiro Iijima, Eri Sato-Shimokawara, and Toru Yamaguchi, Domestic Robot System Considering Generalize, SICE Annual Conference 2010, pp. 390-391
- ④ Yihsin Ho, Tomomi Shibano, Eri sato-Shimokawara, Toru Yamaguchi, "An indoor location-base smart information offering system by augmented reality on Smartphone", The International Symposium on Intelligent Systems. Paper No. 524
- ⑤ Yihsin Ho, Tomomi Shibano, Eri sato-Shimokawara, and Toru Tamaguchi, System of Recognizing Human Action by Mining in Time-Series Motion Logs and Applications, The 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3347-3452. 査読有
- ⑥ Yihsin Ho, Tomomi Shibano, Kazumasa Murakami, Eri Sato-Shimokawara and Toru Yamaguchi, A Smart Information Offering System by Mining Time Series Motion Logs, Joint 5th International Conference on Soft Computing and Intelligent Systems and 11th International Symposium on Advanced Intelligent Systems, pp. 1289-1294
- ⑦ Yihsin Ho, Tomomi Shibano, Eri sato-Shimokawara and Toru Yamaguchi, The Information Providing System by using Human Motion Recognition, Proceedings of 2011 4th International Conference on Human System Interactions (HSI2011), pp. 110-116
- ⑧ Yihsin Ho, Nobuyuki Sekine, Eri sato-Shimokawara and Toru Yamaguchi, Motion Pattern Recognition using Case-Based Reasoning for Information

Providing, Proceedings of SICE Annual Conference 2011, pp. 1284-1289

- ⑨ Takuya Mizutome, Eri Shimokawara, Yasunari Fujimoto, and Toru Yamaguchi, Electric Power Management System using Home Smart Robot based on User's Life-log Proceedings of 12th International Symposium on Advanced Intelligent Systems (ISIS2011), pp. 435-438
- ⑩ Jiguo Zhen, Hiroataka Aoki, Eri sato-Shimokawara, Toru Yamaguchi, Interactive System for Sharing Objects Information by Gesture and Voice Recognition between Human and Robot with Facial Expression, Proceedings of The Fourth Symposium in System Integration (SII2011), pp. 293-298
- ⑪ 柴野 友美, 青木 広貴, 何 宜欣, 下川原 (佐藤) 英理, 山口 亨, ロボットを用いた高齢者のための見守り支援システム, 第29回日本ロボット学会学術講演会
- ⑫ Eri Shimokawara, Kazumasa Murakami, Yihsin Ho, Shin Ishiguro and Toru Yamaguchi, Clustering action data based on amount of exercise for use--model based health care support, the 2012 International Joint Conference on Neural Networks (accepted)

[産業財産権]

○出願状況 (計 1 件)

名称: 行動認識プログラム、行動認識用の端末および行動認識用の処理装置

発明者: 山口 亨, 下川原 英理 (他 3 名)

権利者: 首都大学東京

種類: 特許

番号: 特願 2011-244951

出願年月日: 2011 年 11 月 8 日

国内外の別: 国内

6. 研究組織

(1) 研究代表者

下川原 英理 (SHIMOKAWARA ERI)

首都大学東京・システムデザイン研究科・助教

研究者番号: 00453035