

発話内容に応じたロボットの集団適応動作

Motion generation system of robots for a group according to the utterances and the attribute of speakers

小川 研^{1*} 片上 大輔¹ 新田 克己¹
Ken Ogawa¹ Daisuke Katagami¹ Katsumi Nitta¹

¹ 東京工業大学 大学院総合理工学研究科

¹ Interdisciplinary Graduate School of Science and Engineering,
Tokyo Institute of Technology

Abstract: In this paper, we investigated a kind of the teaching motion and difference in number of the motions by the difference of the group in the equivalence utterance, and to suggest the system which worked that a robot adapted the group depending on utterance contents. I collect the motion data which the member of the group direct taught to a robot with utterance contents and make an adaptation rule. As for the rule set provided in this way, the rule is built by motions to work in the group and utterance in the group so that can be the common knowledge in the group. The motion adaptation degree is higher in a group when the motion is generated from a new text as input by this.

1 はじめに

近年, ASIMO[1] や商業施設などで人間の代わりに案内などのサービスを行うサービスロボット [2] 等の, 人間とのインタラクションを目的としたヒューマノイドロボットが開発されており, 人間と同じ環境で活躍する人間共生型ロボットの普及が高まっている. ところで, 人間同士のインタラクションにおいては 65% がノンバーバルなもの [3] であり, ジェスチャーや仕草が非常に大きな役割を果たしている.

そこで, 本研究ではヒューマノイドロボットをコミュニケーションツールとして用い, 遠隔地にいる人とコンピュータネットワークを介して対話する際に, 自分の発話内容に応じて遠隔地の人のそばにいるロボットに適した動作をさせて, 相手の人に適した情報を伝達することを想定する. 従来研究では, 予め実験者が動作を用意し, 発話テキストに合わせてその動作をロボットが出力する [4] というものが多い. しかし, そのテキストと動作の組み合わせは実験者が決めたものであり, 一般的な組み合わせとは言い難い. さらに, 人間は同じ発話内容からいつも決まった動作をするとは限らない. 発話相手に応じて動作を無意識に変えていると考えられ, さらには相手と同じ属性を持つ場合には同じ

ような動作を行うと考えられる. また, エージェントであるロボットがいつも同じ発話内容から同じ身振りをさせると不自然だと感じられ, 動作にはバリエーションが必要になる. 従来, 動作生成には確率モデルを採用 [5] したり, モーションキャプチャを用いたりして動作に関する研究がされている. 本研究ではこれらの手法を統合し, 用意した発話内容に適切な動作を複数人でロボットに教示 [6] し, 動作のバリエーションを増やす. そして, 同一発話内容における集団の違いによる教示動作の種類と動作数の相違を調査することを目的とする.

第2章では人間, 及びエージェントの集団適応について述べ, 第3章では本研究で用いる集団に適応した動作選択システム, 及び関連研究について説明する. 第4章では実験内容とその結果を示し, 第5章でまとめる.

2 人間及びエージェントの集団適応

人間は成長するとともに, 自分の意思決定で行動する場面が多くなっていく. 人間が環境に適応してゆく過程を, 対人場面におけるパーソナリティ調査 [7] などを行うことで調査がされている. 環境に適応するというと, 人間が周辺環境や相手集団に適応していくのは先生など周りの人間から通じて学ぶものがほとんどである.

*連絡先: 東京工業大学 大学院総合理工学研究科
知能システム科学専攻
〒226-8502 神奈川県横浜市緑区長津田町 4259-J2-53
E-mail: ogawa@ntt.dis.titech.ac.jp

人間の成長過程の適応と同様に，エージェントの人間社会への進出に伴い，人間社会に適応していくためには集団適応という社会的な考慮が必要 [8] であると考えられる．エージェントが社会的学習を実現するためには，従来のエージェントと人間という一対一のインタラクションではなく，エージェントと複数の人間という一対複数のインタラクションについて考えることが重要である．つまり，社会的知識や情報は一人の人間とのインタラクションにより創発されるのではなく，複数の人間とのインタラクションにより生まれるものである．

本研究では，エージェントであるロボットの動作も相手の集団に適応したものを作成する必要があると考える．人間が行う動作については，年齢だけを考慮しても年上の人に対する動作，同い年の人に対する動作，年下の人に対する動作など，相違することが考えられる．他にも初対面同士や親密度による動作の変化など様々あり，これらは同じ発話内容であっても相手の集団である立場を認識し，相手に適応した動作を行っている．相手に適応した動作を行うことは，人間にとって重要なことである．間違った動作を行うと相手に不快感を与えたり，相手の人との関係性に問題が生じる可能性がある [9] からである．

本研究にて調査を行うにあたり，図 1 に示すような場面を想定する．

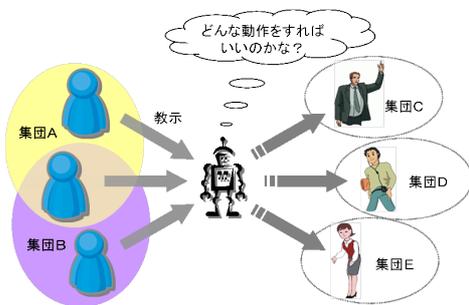


図 1: 想定する集団適応

図 1 は，ロボットが人間と対話を行うときに，相手に応じて動作を選択している様子を示したものである．事前に様々な集団（A や B）に属する人間より動作を収集し，作成したルールに則り，対話相手集団（C や D，E）に動作を行う．

動作収集において，被験者の所属している集団を調査するため，被験者のデータを集めることが重要になる．また，被験者に対して仮想する対話相手を何通りか用意する．対話相手を仮想し，発話内容に応じた動作を教示することで，教示動作に変化が起きると考えている．その後，教示動作の分類を行い教示者の集団が明らかにする．

3 集団に適応したロボットの動作選択システム

発話内容に応じたロボットの動作選択システムを作成するにあたり，本研究で提案するシステムは動作と発話内容のルール作成モジュールと，動作出力モジュールからなる．

3.1 集団適応動作選択システム

発話内容に応じたロボットの動作選択システムの概要を図 2 に示す．このシステムは教示，動作モデル作成，集団ごとのルール作成，動作出力の 4 つのステップからなる．各ステップの説明を以下に行う．

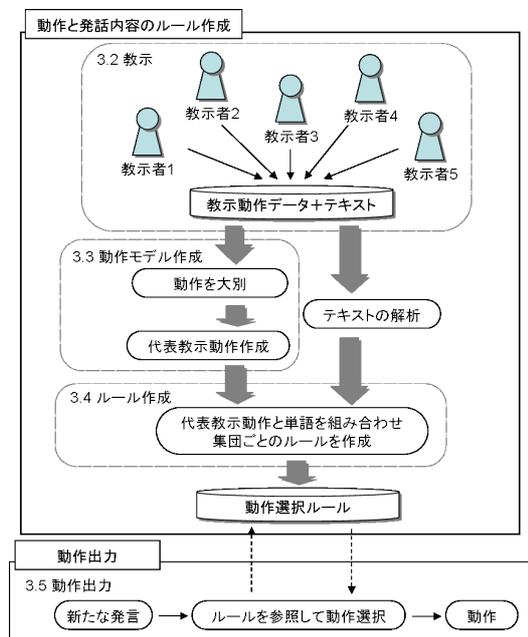


図 2: 集団適応動作選択システムの概要

3.2 教示

ロボットの動作作成には，ロボットへの教示が必要となる．教示とは，ロボットに動作を教え示すことである．その手法として，人間らしい動作を実現するために最も有力なモーションキャプチャを用いる [10] 方法や，近年ヒューマノイドロボットの運動制御への応用が期待されている見まね学習 [11] を用いたものがある．

本研究では，教示手法として直接教示 [6][12] を採用した．直接教示とは，実際にロボットの関節ごとに付いているモータを直接操作し，モータデータを取得する方法である．教示者は関節部分を手で操作して動作

を教示するが、ヒューマノイドロボットの場合などは全身にモータが付いているので、直接教示の際はある程度動作を限定する必要がある。実験ではロボットの上半身だけを用い、スクリプトを用意してその内容に応じた動作を直接教示することで動作を収集する。

3.3 集団適応動作モデル作成

集団適応動作モデルを作成するにあたり、本研究では被験者から教示された動作を大別し、動作ラベルを付ける。明らかに異なる動作、例えば手を振る動作と手を前に出す動作は誰が見ても異なるので、別のラベルとする。ラベル付けされた動作内に複数の教示が存在する複数人教示では、従来研究[5]のような全ての教示が単一の方針であると仮定しているような手法では対応できない。そこで、教示データをクラスタリングすることにより同一の方針の教示をクラスタ化することで、この問題を解決する。

本研究では小竹[12]の提案する動作生成手法を用いて代表教示動作モデルを作成する。代表教示動作とは、複数の似た教示データを統合し、一つの代表される動作のことである。このモデルは、教示動作データを各教示方針別に階層的な手法によるクラスタリングを行い、各クラスタ内でデータの平均化を行い動作を生成する手法である。クラスタ内の複数のデータを平均化することで全体的な教示数を圧縮する効果がある。クラスタリングと平均化の際は時間軸を考慮して動的計画法(DP マッチング)を用い、時間に依存するタイミングや静止時間などの差を考慮している。ただし、クラスタ内の教示の数が一つしか無い場合はこれを代表教示動作とする。

3.4 集団ごとのルール作成

上記で作成した代表教示動作と単語のルールを、仮想した集団ごとに作成する。集団ごとに適応したルールとは、話者間の集団の違いを認識した上での出力動作の違いを考慮するルールのことである。

ルール作成において、被験者に提示したスクリプトを解析し、文中に出現する単語と代表教示動作を組み合わせる。なお、単語と代表教示動作を組み合わせる際は単語に対して代表教示動作を付与して組み合わせる。そのため、被験者は提示されたスクリプト中の好きなところ(単語)で動作を入力し、その単語に印を付ける。これにより、被験者は好きな単語の場所で自由に動作を教示することが出来る。

以上の手続きから得られた代表教示動作と単語の組み合わせを蓄積していく。蓄積された代表教示動作と単語の組み合わせから、各集団の各単語において出力

される代表教示動作に確率を付与する。その際、代表教示動作は一つとは限らず、一つの単語に複数の代表教示動作が組み合わさることもあり、それぞれの代表教示動作に確率を付与する。集団ごとにこの確率が変わり、例えば年上の集団に対しては「こんにちは」という単語について、お辞儀をする確率が非常に高くなり、右手を上げる確率が非常に低くなることが考えられる。尚且つ、3.3で述べた動作生成により、お辞儀の深さにも考慮する。このようにして作成した代表教示動作と単語のルールをルール集合として蓄積していく。

3.5 動作出力

テキストを入力とし、作成したルールを参照して出力動作を選択する。入力されたテキストの形態素解析を行い、単語を抽出する。自動的に単語を抽出することで、ロボットはリアルタイムで動作を選択することが出来る。本研究では、ロボットは入力されたテキストの発話を行わず、動作を行うのみである。ロボットが動作を行うタイミングに関しては、テキスト中に出現する単語に関連付けして動作をさせる場合が多い。例えば、実際のプレゼンテーションのデータを分析し、テキストにジェスチャーを自動的に付与してナレーション音声と同期したエージェントアニメーションを作成[4]したりする際の動作も、この手法を用いている。また、話し手における強調語句の発話音声に対する身体動作の分析[13]を行う際にも、また同様である。しかし、今回は入力テキストは短文に限り行うため、タイミング制御は行わない。つまり、テキストが入力され、形態素解析後ルール付けされた単語が出現すれば、すぐに動作を行うようなシステムである。短文であれば、被験者は動作タイミングに違和感を感じることはないと考えられるためである。

4 実験

4.1 目的

同一発話内容における、集団の違いによる教示動作の種類と動作数を調査することを目的とする。今回行った実験では、数ある集団の中でも、社会的関係の上下関係と親密度に考慮して調査を行った。予めスクリプトを用意し、複数の被験者に提示してその内容に応じた動作をロボットに教示する。

4.2 実験環境

実験ではスピーシーズ社製のロボット SPC101-C[14]を用いた。ヒューマノイドロボットを用いることで、動

きに意味を持たせてジェスチャーとして扱うことが出来る．そのロボットと、実際の教示風景を図3に示す．

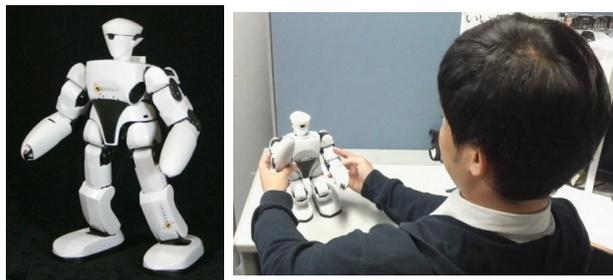


図 3: ロボット SPC101-C[14]

このロボットは身長約 33cm で体重約 1.5kg、有している機能は音声出力、LED 表示、内臓型 CCD カメラなどがあるが、本研究では動作に着目するため、サーボモータのみを用いる．サーボモータの数は合計 22 個あるが、ロボットのバランスを保つため、両足は固定し、上半身のみサーボモータ 7 個（両腕各 3 個、腰 1 個）を用いてデータを収集する．データは、サーボモータの数に応じた次元の時系列データからなり、等間隔で各サーボモータの角度が返ってくる．

図 4 に動作データの一例として、このロボットの右手を左右に振らせた時の教示データを示す．人間が右手を左右に振る場合肘を動かすのが普通であると考えられるが、このロボットの場合は肩のモータが左右に振られていることになる．なお、動作は教示中両腕を真っ直ぐ下に降ろした状態の基本姿勢に戻すことで区別する．

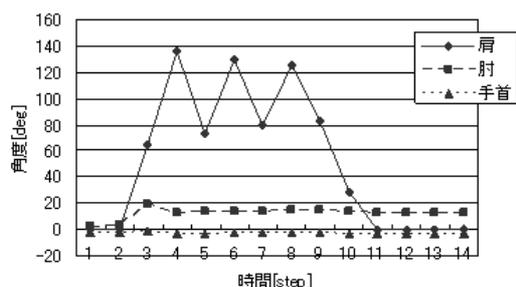


図 4: モータの時系列データ

4.2.1 被験者

被験者は、男子理系大学生 13 人、女子理系大学生 2 人、の計 15 人である．被験者はそれぞれ今回使用したロボットを見るのは初めてであり、ロボットに慣れていない学生がほとんどである．教示動作のイメージが

出来るように、事前にロボットに触れてもらい自由度を確認してから実験を始める．

4.2.2 タスク設定

人間は会話中無意識に動作をしているのであり、意識して意味のある動作を教示することは非常に難しい．本研究においても、動作の種類と動作数を調査するにあたって、多くの発話内容を仮定するのではなく、比較的動作が多いと考えられる挨拶の場面に限定した．動作をロボットに教示する際は、人間が行う動作を教示するのではなく、このロボットを仲介役として相手と対話を行うときに行ってほしい動作を教示する．

スクリプトを読む際は、相手の立場を仮想的に置き換え、今回は仮想的に研究室の先輩、後輩、同級生（同じ研究室員・他の研究室員）、とした．ここで、一般的な学校生活を仮想するので、同級生の他の研究室員とは顔見知りであるという条件をつけた．被験者には同一発話内容で、言い回しが異なるスクリプトを提示して動作を教示してもらった．被験者に提示したスクリプトを図5に示す．

```

Story1 (A : 年下, B : 年上)
A : おはようございます。
B : ああ、おはよう。元気だね。
A : はい、朝は得意ですので元気なんです。
B : そうか。私は朝が苦手だからまだそんなに元気が出ないだよ。
A : 元気出して今日も一日頑張らしましょう。
B : 本当元気だね。目が覚めるまでもう少し時間ちょうだいよ。

Story2 (同級生)
A : おはよう。
B : おお、おはよう。凄く元気だね。
A : まあ、朝は得意だからね。
B : そうか。僕は朝が苦手だからまだそんなに元気が出ません。
A : 元気出して今日も一日頑張らうぜ。
B : 本当元気だね。目が覚めるまでもう少し時間くれー。

Story3 (A : 年上, B : 年下)
A : おはよう。
B : おはようございます。凄く元気ですね。
A : まあ、朝は得意だからね。
B : どうですか。僕は朝が苦手なのでまだそんなに元気が出ません。
A : 元気出して今日も一日頑張らないと。
B : 本当元気ですね。でも、目が覚めるまでもう少し時間を下さい。
    
```

図 5: 被験者に提示したスクリプト

Story は事前いくつか用意した中から、動作の出やすいものを厳選した．Story の 1 から 3 はそれぞれ同じ発話内容で上下関係の役割が変わっている．Story1 と 3 は A と B の立場が入れ替わっているだけであるが、これは発話始めの立場による動作の違いを調査するためである．事前に被験者にはスクリプト中の話者の社会的立場の説明を行っている．

被験者へのタスクは次の 2 点となる．

- スクリプト中で、動作をしたいと思った箇所にチェックを付ける
- 動作を区切るため、動きの始点と終点は同様の基本姿勢に戻す

4.3 実験結果

被験者により教示された、話者間の上下関係と親密度の違いによる動作の種類と動作数の違いの評価を行う。

4.3.1 挨拶の場面における教示動作の割合

挨拶の場面についての、上下関係の有無、親密度を考慮した場合の動作の種類を述べる。なお、図5のスク립トにおいて、各 Story の始めの1ターンについてを挨拶とする。この挨拶の場面で、被験者より選択されたテキスト中の単語は「おはよう」であり、動作の種類と出力確率を表1に示す。

表 1: 上下関係の有無における動作の出力確率 [%]

関係 \ 動作	お辞儀	右手を上げる	他	無し
相手が年上	83	17	0	0
相手が年下	4	88	4	4
相手が同級生	23	61	4	12

話し始めがどちらの立場になるかは関係なく、高い確率で年上の立場では右手を上げ、年下の立場はお辞儀をするという結果になった。

次に、属するグループを分類して、親密度の少ない他の研究室員と親密度の高い同じ研究室員を比較した結果を表2に示す。

表 2: 親密度の違いにおける動作の出力確率 [%]

関係 \ 動作	お辞儀	右手を上げる	他	無し
他の研究室員	46	46	0	8
同じ研究室員	0	75	8	17

表1では、同級生同士では右手を上げる確率が一番高かった。しかし、表2では親密度の低い他の研究室員に対しては、お辞儀と右手を上げる確率が半々になった。同じ研究室員に対してはお辞儀をした者はおらず、右手を上げる確率が最も高かった。その他、同じ研究室員に対しては動作が無い場面も他に比べて多く見られた。

表 3: 右手を上げる動作の割合 [%]

関係 \ 動作	大きく上げる	少し上げる	少し振る
他の研究室員	16	78	6
同じ研究室員	30	54	16

右手を上げる動作において、小竹 [12] の手法を用いてクラスタリングしたところ、表3に示すように右手

を「大きく上げる」「少し上げる」「少し振る」という3種類の動作が分類された。これらの動作は、複数人から教示された代表教示動作として登録される。

4.3.2 挨拶の場面以外での動作数の違い

挨拶の場面以外における、上下関係の有無、親密度を考慮した場合の動作の種類を割合を表4に示す。

表 4: 挨拶以外の場面における動作の出力確率 [%]

関係 \ 動作	お辞儀	右手を上げる	他	無し
相手が年上	4	8	35	53
相手が年下	0	2	46	53
相手が同級生	4	13	7	76

どの場面においても動作が出ない確率が最も高かった。

挨拶の場面に比べ、挨拶以外の場面では動作数は少なかったものの、多くの種類の動作が教示された。挨拶の場面では、お辞儀もしくは右腕を上げるといった動作が多数であったが、左腕を上げたり、相手の手を上げる側に合わせて腕を上げる場合もあった。挨拶以外の場面で、最も多くの種類の動作が教示された場面は、スク립ト4行目の「苦手」という単語と「元気が出ない」「元気になれない」というところであった。一番多かったのは手を左右に振って拒否を示す動作であり、これは一般的に考えられる動作であった。しかし他に、お辞儀する人、両手を前に出して脱力感を表す人など、全部で6種類の動作が教示された。また、このスク립トを通じて全部で合計約20個の動作が教示された。

4.4 考察

挨拶の場面において、対話相手が年上になるとお辞儀をし、年下になると右手を上げる確率が最も高いという結果が得られた。そして、同級生同士の対話では右手を上げる確率が最も高く、何も動作をしない場合も見られた。ここで親密度を用いた結果、親密度が低いと考えられる集団に対しては、お辞儀と右手を上げる確率が同じであるが、親密度の高い集団と対話したときは、お辞儀を教示した者はいなかった。これにより、適応する集団の親密度によっても動作が変わることが分かった。また、同じ右腕を上げるという動作でも、親密度の違いにより変化が出た。同じスク립トでも教示する集団によって教示動作は異なり、その集団として親密度や上下関係が有効であると分かった。これにより、ロボットも発話相手に応じて動作を変える必要があると考えられる。

次に、挨拶以外の場面において、話者間に上下関係がある場面の方が動作数が多いという結果から、上下関係にある相手と対話を行い情報を伝達する場合はバーバル情報と同時に、動作が重要であると考えられる。同級生同士の場合では、同級生という集団意識から動作を省いて簡略化し、動作が難しい場面ではあえて教示しなかった。これは、動作に対する意識が薄いことが考えられる。これにより、ロボットの立場を考慮する場合は動作の数も重要になると考えられる。

最後に、各文章で動作の数が非常に多く見られた。発話内容に応じた動作は、一つだけではなく多数考えられる。同一動作においても人によって動作をするタイミングが異なり、実験者が考えていなかった動作も教示され、実験者が予め用意している動作以外にもロボットには有効な動作があることが示唆された。

5 おわりに

本稿では、同一発話内容における集団の違いによる教示動作の種類と動作数の相違を調査した。挨拶が行われる場面を想定したスクリプトを用意し、対話相手を仮想して仲介役であるロボットに発話内容の適した動作を直接教示した。

この実験より、同一発話内容でも属する集団により教示する動作が異なることが分かった。また、ロボットを作成する際、対話相手の違いにより動作を変える必要があることが示唆された。対話相手との年齢の上下関係がない場合よりも、上下関係がある場合の方が動作が多く教示されたことから、人間は上下関係のある人を相手にする方が、動作つまりジェスチャーを用いて情報伝達をしていることが分かった。

今後、現在までに収集した動作を用いて、多彩な動作を行い対話相手との社会的関係を考慮して動作を行うロボットを作成する。そして、単語と動作の一対一の組み合わせだけでなく、一対多の関係がロボットには重要であること調査する。

参考文献

- [1] <http://www.honda.co.jp/ASIMO/>
- [2] 村川, 十時: サービスロボットによる「振る舞い」の評価 - 商業施設での試験運用 -, *Human-Agent Interaction Symposium 2006*, (2006)
- [3] 黒川, ノンバーバルインターフェース, オーム社, (1994)
- [4] 中野, 村山, 西田: 会話エージェントによる情報提供 - 非言語による重要概念の強調 -, *社会技術研究論文集*, Vol.2, pp.159-166 (2004)
- [5] 稲邑, 中村, 戸嶋, 江崎: ミメシス理論に基づく見まね学習とシンボル創発の統合モデル, *日本ロボット学会誌*, Vol.22, No.2, pp.256-263 (2004)
- [6] 片上, 生和, 新田: 複数人直接教示によるロボットの適応的行動スキルの抽出とインタラクション解析, *The 22nd Annual Conference of the Japanese Society for Artificial Intelligence*, (2008)
- [7] 友野, 橋本: 対人場面におけるあいまいさへの非寛容が新入学生の適応過程に及ぼす影響, *パーソナリティ研究*, Vol.14, No.1, pp.132-134 (2005)
- [8] 大村, 片上, 新田: 異文化体験ゲームにおける社会適応エージェントの設計, *Human-Agent Interaction Symposium 2008*, (2008)
- [9] B.Endrass, M.Rehm, E.Andre: Culture-specific Communication Management for Virtual Agents, *AAMAS 2009*, (2009)
- [10] 山根, 中村: ヒューマンフィギュアの全身運動生成のための協応構造化インタフェース, *日本ロボット学会誌*, Vol.20, No.3, pp.335-343 (2002)
- [11] Schaal, S: Is imitation learning the route to humanoid robots?, *Trends in Cognitive Sciences*, Vol.3, No.6, pp.233-242 (1999)
- [12] M. Kotake, D. Katagami, K. Nitta: Acquisition of behavioral patterns depends on self-embodiment based on robot learning from multiple instructors, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.11, No.8, pp.989-997 (2007).
- [13] 大崎, 山本, 渡辺: 話し手における強調語句の発話音声に対する身体動作の分析と合成, *ヒューマンインターフェースシンポジウム 2005*, (2005)
- [14] <http://www.speechs.com/>