

博士論文

論文題目

人間・エージェント間の協調的
コミュニケーション制御に関する研究

速水 達也

論文概要

本論文は、筆者が大阪工業大学大学院情報科学研究科に在籍中に行った「人間・エージェント間の協調的コミュニケーション制御に関する研究」に関する成果の総括である。

本論文は、5章で構成されており、以下にその概要を述べる。

1章では、現在の人間・ロボット間のコミュニケーション制御に関する問題点に触れ、本研究の意義について述べる。本研究では、人と密接にかかわるロボットあるいはエージェントが、人間と持続的かつ自然にコミュニケーションをするための制御方式を実現することを旨とする。本論文で扱う内容は主に以下の3つであり、それぞれ2章、3章、4章が対応する。

1) 自然な対話の流れをモデル化し、従来方式より自然でストレスの少ない親和性のあるコミュニケーション制御方式の実現方法

2) ロボットと人間のグループとのコミュニケーションを想定し、初対面時や社会的スキルに違いに応じた協調的コミュニケーション制御方式の実現方法。また、人間が初対面者を仲介する方法をモデル化し、初対面者同士を仲介するエージェントへの実装

3) これまでに十分検討されていなかったグループの感情を推定する手法についての研究
次に、各章ごとの概要について述べる。

2章では、交替潜時と韻律情報に基づく非線形振動子による連続的な引き込み制御とインタラクションルールによる離散的な引き込み制御により、人とエージェントが協調的にコミュニケーションを行うための会話制御方式を提案した。また、情報収集を目的としたエージェントへ実装し、会話同調制御の評価実験を行った。評価実験の結果からは、従来手

法に較べて、エージェントがどの程度親しみやすいかの評価尺度である親身度に対して有意差を示すことができ、同時にエージェントの会話の自然さ（ストレスが少ない）に対しても有意差が認められた。

3章では、ロボットと人間のグループとのコミュニケーションを想定し、これまで十分に検討されていなかった初対面時や社会的スキルの違いに応じた協調的コミュニケーション制御についての研究を行い、人間が初対面者を仲介する方法をモデル化した。また、初対面者の仲介モデルをエージェントへ実装し、その有効性を検証した。具体的には、このエージェントと2人の初対面者がコミュニケーションをする実験と、エージェントを用いず2人の初対面者がコミュニケーションをする2種類の実験を行い、アンケート結果から、学生の実験参加者を用いた実験室での限定された環境ではあるが、提案モデルの有効性を検証することができた。また、社会的スキルの観点からも検証を行った。実験を行い、具体的には、実験参加者として高齢者を対象とし、社会的スキルの高低によりグループを分け実験を構成したところ、社会的スキルの高い実験参加者に比べ社会的スキルの低い実験参加者のほうがロボットの仲介行動が有効に働き、社会的スキルが低い人にとって、5つの仲介行動の身体的振る舞いが有効に働くことが確認できた。

4章では、これまでに十分検討されていなかったグループの感情を推定する手法についての研究を行った。ロボットが人間と円滑にコミュニケーションを行うためには、ロボットが人間の状態に応じて適応的に振る舞うことが必要不可欠である。人と人がコミュニケーションを行うときには、言語情報よりも、視線や顔表情・ジェスチャ・韻律情報などの非言語情報がより重要な役割を担うことが明らかにされている。特に、音声や顔表情は豊富

な情報を含んでおり、コミュニケーションの状態理解において極めて重要な要素である。

したがって、人間の感情を推定する技術は、ロボットと人間の円滑なコミュニケーションを支える技術として、重要であると言える。これまでに、個人の感情を推定する手法は多く研究されている。しかし、グループの感情を推定する手法については十分には検討されていないのが現状である。そこで、複数人の人間から構成されるグループの感情をモデル化し、グループコミュニケーションの映像情報からグループ感情を推定する手法について提案した。提案する手法は、個人の顔表情と韻律情報から個人感情を推定し、個人感情と観測されたパラメータを基にベイジアンネットワークにより統合し、グループの感情を推定する手法である。実験の結果、個人感情については認識率が約 70%と良好な値を得た。一方、グループ感情についても、笑いの感情については認識率が約 70%、全体の認識率は約 60%であった。グループがどのような感情であるかどうかを推定することは、人間でも認識困難な問題であり、実験からは十分な結果が得られたと思われる。今後、個人感情の認識については、発話の有無による認識率低下を防ぐため、音声認識による言語的情報の組み込み等を検討する。グループ感情の認識についても、より人間の直感に即した認識ができるよう、検討を進める必要がある。

5章では、1章から4章までに述べた研究結果について述べるとともに、本研究の研究成果を総括した。

今後の課題としては、提案した人間・ロボット間のコミュニケーション制御方式を、人とコミュニケーションを取ることを目的としたロボットやエージェントへ実装し、本制御方式の実用性について検証することが考えられる。具体的な応用先としては、遠隔（医師

と患者間) で問診可能な「遠隔問診エージェント」や介護を目的とした介護ロボットなどが挙げられる。今後は、これらの実現に向けて取り組みたい。

目次

1 章 緒言	9
1. 背景	9
2. 本研究の目的	10
2 章 交替潜時と韻律情報に基づく会話同調制御方式と情報収集を目的としたエージェント への実装	13
1. はじめに	13
2. 提案する同調制御方式と会話エージェント	16
2.1 非線形振動子の引きこみ現象とパラ言語	16
2.2 提案する同調制御方式の概要	17
2.3 コミュニケーション同調制御	19
2.3.1 連続的引き込み制御	19
2.3.2 離散的引き込み制御	22
2.4 情報収集を目的とした会話構造モデルと引き込み制御	23
3. 提案手法の実装	25
3.1 インタラクション状態量算出部	25
3.2 エージェントインタラクション制御部	26
3.2.1 音声合成モジュール	27
3.2.2 動作生成モジュール	27
4. 評価実験	28

4.1	実験目的	28
4.2	比較制御方式.....	28
4.3	実験方法	30
4.4	実験結果	31
4.4.1	親身度の項目における分析.....	31
4.4.2	ストレスの項目における分析.....	34
4.5	考察	36
5.	おわりに	38
3章	初対面期の会話を促す仲介エージェント.....	41
1.	はじめに	41
2.	初対面者の仲介モデル.....	43
2.1	初対面者の仲介モデルと仲介者の振る舞い.....	43
2.2	初対面仲介における4つの状態とその遷移モデル.....	44
2.3	初対面における5つの仲介行動.....	45
2.4	仲介行動と場の遷移.....	48
3.	エージェントへの実装.....	50
3.1	4層構造モデルのエージェント.....	50
3.2	エージェントの9つの動作.....	52
3.3	身体的引き込み制御におけるタイミング生成.....	53
4.	初対面における仲介行動の分析.....	55

4.1	仲介行動の観察実験.....	55
4.2	仲介行動の分析と分類.....	55
5.	評価実験.....	56
5.1	実験目的.....	56
5.2	実験手順.....	57
5.3	実験結果.....	61
6.	社会的スキルの観点から見たエージェントの分析.....	63
6.1	実験目的.....	63
6.2	実験手順.....	64
6.3	実験結果.....	65
7.	おわりに.....	66
4章	表情と韻律情報に基づくグループ感情.....	69
1.	はじめに.....	69
2.	グループ感情推定.....	70
2.1	システム概要.....	70
2.2	グループ感情場モデル.....	71
3.	個人感情の推定.....	72
3.1	表情特徴量の抽出.....	73
3.1.1	顔領域の検出.....	73
3.1.2	表情特徴量の抽出.....	74

3.2	音声特徴量の抽出.....	75
3.2.1	音声特徴量.....	75
3.2.2	音声パワーとピッチの抽出.....	76
3.3	表情特徴量と音声特徴の統合.....	76
4.	グループ感情の推定.....	77
5.	実験.....	78
5.1	実験方法.....	78
5.2	実験結果.....	79
6.	考察.....	81
7.	おわりに.....	83
5 章	結言.....	85
	外部発表リスト.....	89
	謝辞.....	90
	付録 A.....	91

1章 緒言

1. 背景

近年、ロボティクスの分野では、人とのコミュニケーションや介護を目的としたコミュニケーションロボットや介護ロボットが盛んに研究開発されており、ロボットが一般の人々にも身近な存在となりつつある。コミュニケーションロボットや介護ロボットのように、人々の生活により密接にかかわる場合、人と持続的かつ自然にコミュニケーションをとらなければならない。しかし、従来の1対1のエージェントコミュニケーションでは、会話の盛り上がりや自然な会話の流れが十分に考慮されていない。人とロボットが自然にコミュニケーションを進めるためには、人とロボットが協調的に対話するためのダイナミクスを解明し、より自然でストレスの少ない親和性のあるコミュニケーション制御が必要である。

一方で、我が国は高齢化率が21%を超える超高齢化社会である。したがって、今後、エージェントによる老人の介護や、孤独な高齢者とのコミュニケーションを対象としたエージェントの技術開発が期待されている。特に孤独な高齢者は、コミュニティへ参加することが、高齢者のQoL向上につながると言われている。しかし、高齢者と近所の人たちとの交流の弱まりや、約半数の高齢者がグループ活動などのコミュニティに参加していないといった孤独な高齢者の存在が問題視されている。エージェントが、孤独な高齢者をコミュニティへ参加させることを支援することができれば、高齢者のQoL向上につながることを期待される。

また、人間社会の中で活躍するエージェントにとって、エージェントが人間の感情や心

理を読み取り，人間の状態に応じて適応的に振る舞うことが望ましい．人間同士のコミュニケーションにおいては，言語情報のほかに非言語情報とパラ言語情報がやりとりされる．伝達される情報の多くは，非言語情報とパラ言語情報である．特に，表情・ジェスチャ・イントネーション・アクセント，といった情報は会話において重要である．したがって，これらの情報から，人間の感情を推定することは，人間とエージェントの円滑なコミュニケーションを実現するために重要であると言える．また，個人の感情だけでなく，人間の集団的な感情を推定することができれば，エージェントが複数人の人間と適切なグループでコミュニケーションを実現することができる．応用先として，人間の集団的な感情を推定する機能をエージェントに搭載すれば，個人感情の遷移とグループ感情の遷移の状況を考慮して人間同士の会話のタイミングを見計らって会話に参入することなどが可能となり，より自然なコミュニケーションが可能なロボットの実現などが見込まれる．

2. 本研究の目的

前節で述べたように，従来の 1 対 1 のエージェントコミュニケーションでは，会話の盛り上がりや自然な状態のモデル化が十分に行われていなかった．筆者はまず，1 対 1 の対話の流れをモデル化し，情報収集を目的とした会話エージェントへ実装を行った．

2 章では，交替潜時と韻律情報に基づく非線形振動子による連続的な引き込み制御とインタラクションルールによる離散的な引き込み制御により，人とエージェントが協調的にコミュニケーションを行うための会話制御方式を提案する．また，提案方式を情報収集を目的とする会話エージェントへ実装し，会話同調制御の評価実験を行った．評価実験の結果

から、直前の状態量の局所的制御を行う方式や交替潜時のみの単純な同調方式の従来手法に較べて、エージェントがどの程度親しみやすいかの評価尺度である親身度に対して有意差を示すことができ、同時にエージェントの会話の自然さ（ストレスが少ない）に対しても有意差が認められた。

また、ロボットと人間のグループがコミュニケーションをすることを想定し、これまで十分に検討されていなかったグループコミュニケーションの制御方式として、初対面時や社会的スキルの違いに応じた協調的コミュニケーション制御方式を提案する。特に、孤独な高齢者がコミュニティへ参加することが、高齢者の QoL 向上につながると言われている。しかし、高齢者と近所の人たちとの交流の弱まりや、約半数の高齢者がグループ活動などのコミュニティに参加していないといった孤独な高齢者の存在が問題視されている。孤独な高齢者が新しくコミュニティに参加するときに、心理的障壁の高さなどの要因から初対面者同士が十分に情報交換できないことも多い。

3 章では、初対面者同士の会話を仲介するエージェントの実現を目指す。そのために、初対面期の仲介行動を分析し、適切な仲介行動を 5 カテゴリに分類した。また、提案するエージェントの制御モデルとして、(1) 4 つの状態、(2) 5 つの仲介行動、(3) 9 つのエージェント動作、(4) タイミング制御により構成された、4 層構造モデルの提案をする。提案モデルは、大学生の実験参加者 20 名による初対面者紹介実験において、学生に被験者による限定された環境ではあるが、提案モデルの有効性を検証することが出来た。また、社会的スキルに着目し、実験参加者を高齢者とした場合の実験を行った。

続いて 4 章では、ロボットが複数の人間と円滑にコミュニケーションするためにはコミュニケーション場の雰囲気把握することを目指したグループ感情場モデルを提案する。顔表情が同じでも、発話の有無によって推定すべき感情は異なるため、顔表情のみを用いた感情の推定結果は、人間の直感とは大きく異なる。そこで本章では、顔表情の特徴量と音声特徴量を統合することにより、ロバストな感情認識を目指す。また、提案するグループ感情場モデルを実装し評価実験を行った。

2章 交替潜時と韻律情報に基づく会話同調制御方式と情報収集を

目的としたエージェントへの実装

～会話のゆらぎをモデル化した1対1のエージェントコミュニケーションデザイン～

1. はじめに

近年、ロボティクスの分野では、人とのコミュニケーションや介護を目的としたホームロボットや介護ロボットなどが開発され、ロボットが一般の人々にも身近な存在となりつつある。ホームロボットや介護ロボットのように、ロボットが人々の生活により密接にかかわる場合、人と持続的かつ自然にコミュニケーションをとる必要がある。そのためには、人とロボットが協調的に対話するための対話の流れを解明し、より自然でストレスの少ない親和性のあるコミュニケーション制御が必要である。

人間同士の会話では、言語情報とともに非言語情報とパラ言語情報がやりとりされる。伝達される情報の多くは非言語情報およびパラ言語情報であるが、その中でも発話時間・交替潜時・イントネーション・アクセント・音圧レベル・発話速度などのパラ言語情報が、話者に対する印象や対話の流れに大きな影響を与えていることが確認されている。同時に、パラ言語情報は、話者間で互いに同調することも確認されている[1][2][3][4][5][6][7]。

長岡ら[2][3]は、協調的な音声対話では、対話者同士がほぼ同じような発言長や交替潜時で発言や相槌をする引き込み現象を確認している。また、対話者の交替潜時が円滑なコミュニケーションを促すことも示唆している。交替潜時とは、一方の話者が話し終わって次の

話者が話し始めるまでの時間間隔のことである[2]. 三宅ら[4]は、機械の発話特徴（発話長・発話間隔）が一定であるならば、人間は自らの発話特徴を機械に同調させ、機械が途中で発話特徴を変化させた場合には人間は変化後の発話特徴に向けて同調させていく、ことを確認している. 小松ら[5]は、人間同士の音声対話において、発話速度の引き込み現象が存在することを示し、人間と人工物との対話においても同様に観察されることを示した. 垣田ら[6]は、質問応答形式の実験にて、一方の話者の基本周波数が高ければもう片方の話者の基本周波数が高くなることを指摘している. 西村ら[7]は、協調的な音声対話システムを実現するために、人間同士の対話の印象と韻律変化との間にどのような関係性があるのかを分析し、人間同士の対話における韻律変化には、同調傾向の相互作用があることを示した. また、同調傾向が強い対話は盛り上がっている印象を持っており、フランクで親しみのある盛り上った対話では、対話の韻律特徴（F0、パワー、話速）が同調していることを示した.

パラ言語に基づく引き込み制御による会話システムの実現方式としては、(1) 最初に最適解を求めておき会話に対する状態量を静的に適用するもの、(2) 直前の状態量に着目し動的に局所的制御を行うもの、(3) 会話全体のリズムに着目し同調させ動的に制御を行うもの、に分類ができる. (1) の問題点としては、状態量が固定されているため会話中のコンテキストの変化に対して対応ができない[8]. (2) については、頷きや身体動作タイミングの予測モデルに基づき引き込み制御を行い、局所的なコンテキストの変化には対応することはできる[9]が、会話全体のリズムに関わるような大局的な状態量の変化には対応できない. (3) の会話全体にリズムがあるような会話例としては、聞き取りのように会話を

テンポよく進行させるようなケースや応答・共感などが繰り返され会話に一定以上のダイナミクスがあるようなケースが挙げられる。(3)の会話全体のリズムに着目し同調させながら引き込み制御を行う方式としては、林ら[10]の研究が挙げられる。しかし、交替潜時や発話速度の時間的リズムに基づく引き込み制御を行うに留まっており、韻律情報による同調効果[7]については考慮されていない。会話全体をより自然で親しみやすくするためには、会話全体の「間」のゆらぎを反映した時間的リズムに加え、会話全体の「トーン」のゆらぎを反映した韻律的なリズムを同時に制御する必要があるが、このような会話全体のゆらぎに着目し引き込み制御を行い、会話システムをデザインしている研究例はない。本研究では、図 2.1 に示すように、会話の「間」の変動と韻律の変動を対話の流れにおける「ゆらぎ」と捉えモデル化を行う。ゆらぎのリズムが共振したときに同調現象が起こることが知られている[11]。コミュニケーションにおける同調現象は、協調的会話の状態や、会話が自然に盛り上がっているときなどに観察される[3][7]。本研究では、このような会話全体のゆらぎに基づいた引き込み制御モデルを特徴とする会話エージェントを提案する。会話エージェントの効果としては、「ストレスが少なく自然」で、より「協調的である」ことが期待される。具体的には、非線形振動子に基づき引き込み制御モデルの定式化を行い、交替潜時と韻律情報に基づいた大局的な同調制御を行う。この会話全体の時間的ゆらぎと韻律的ゆらぎに着目した同調制御方式を情報収集会話エージェントとして実装し、提案する同調制御方式の有効性の検証を行う。

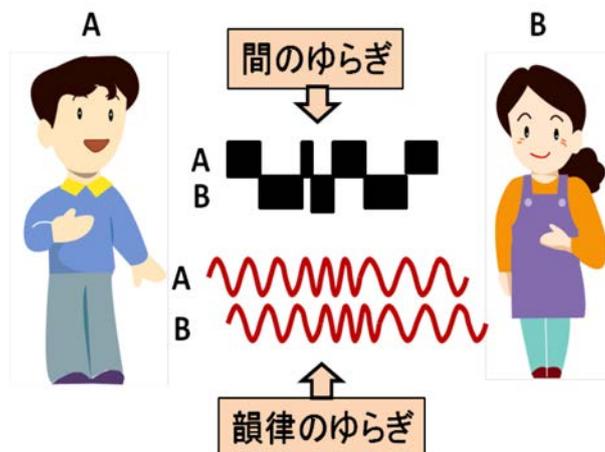


図 2.1 会話における状態量のゆらぎとダイナミクス

2. 提案する同調制御方式と会話エージェント

2.1 非線形振動子の引きこみ現象とパラ言語

非線形振動子には、外部の振動に合わせて振動数を同調させる引き込み現象が見られる。引き込み現象は、コミュニケーションを行っている二者間でも起こり、自らの表出するパラ言語と相手のパラ言語が同調する現象が観察されている。本研究では、この引き込み現象を Van der Pol 方程式を拡張することで、実現する[11].

Van der Pol 方程式は、式 (1) で与えられる。

$$\ddot{x} + a(x^2 - 1)\dot{x} + x = 0 \quad (a > 0) \quad (1)$$

ここで、 a は粘性係数であり、振動子の挙動の強度を制御できる。式 (1) は、2つの振動子が相互に影響を与えるように、式 (2) に拡張可能であり、これにより、相互引き込みを発生させることができる[11].

$$\begin{aligned} \ddot{x}_1 &= \varepsilon (1 - x_1^2)\dot{x}_1 - x_1 + \lambda (x_2 - x_1) \\ \ddot{x}_2 &= \varepsilon (1 - x_2^2)\dot{x}_2 - x_2 + \lambda (x_1 - x_2) \end{aligned} \quad (2)$$

2.2 提案する同調制御方式の概要

人とエージェントが協調的にコミュニケーションを実現するためには、対話者とエージェント間の対話の流れを考慮してコミュニケーション制御方式を設計しなければならない。この対話の流れをモデル化する一般的な方法論として、制御すべき状態量を、状態方程式で表現する方法が考えられる。

本研究で検証するコミュニケーションの状態量は、先行研究で人間同士の同調傾向が確認されている交替潜時（時間的引き込み制御パラメータ）、および基本周波数 F_0 （韻律的引き込み制御パラメータ）とする。本論文では、基本周波数 F_0 を以下、ピッチと呼ぶ。これらの状態量の同調現象を非線形振動子に基づく状態方程式によりモデル化し、対話者とエージェント間の連続的で大局的な引き込み制御を行う。

しかし、状態方程式で記述される連続的な引き込み現象は、状態量差が大きいと発現しない。したがって、エージェントと人間の会話がまったく成立していない状況では、このような引き込み制御を行うことは困難であることに留意しなければならない。本研究では、共感を引き起こす会話構造モデルの下で、頷きなどの身体的インタラクションなどの離散的な引き込み制御を行うことによって、応答可能な会話を成立させ、同調現象を引き起こす条件を調整する戦略をとる。すなわち、図 2.2 に示すように、離散的引き込み制御によって、同調現象を発現するだけの状態量差まで調整し、同調現象を発現させる。一旦、同調現象が発現されると、安定したコミュニケーションを維持することが可能となる。

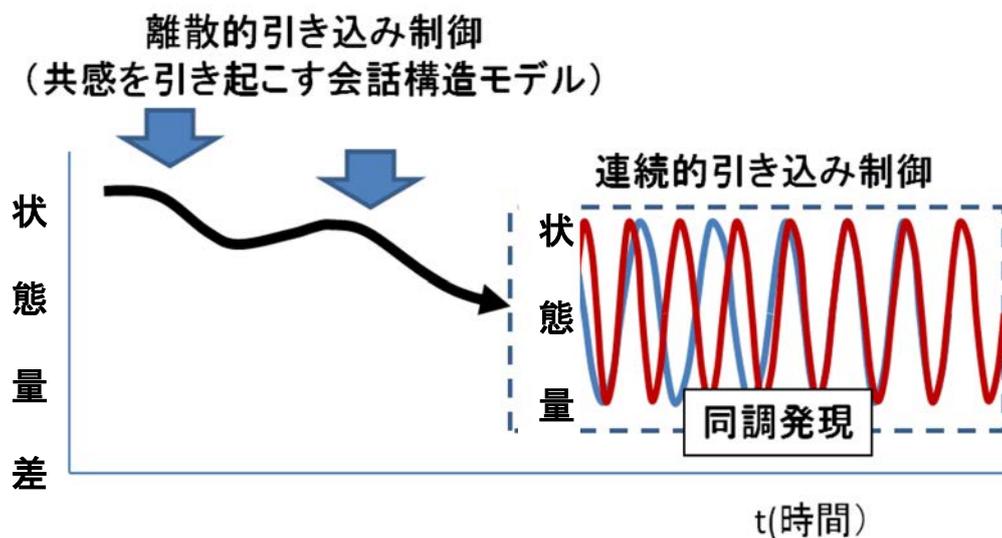


図 2.2 同調現象発現のための引き込み制御

図 2.3 に、本研究で提案する制御方式を示す。人間の発話や動作をセンシングし算出されたユーザ状態量と、インタラクション制御部から生成されるエージェントの状態量に対して、状態方程式による連続的で大局的な引き込み制御とインタラクションルールによる離散的で局所的な引き込み制御に基づきコミュニケーション同調制御を行い、制御情報をエージェントのインタラクション制御（バーバル制御、ノンバーバル制御）に反映させ、共感を引き起こす会話構造モデルのもとで会話状態を動的に遷移させ、コミュニケーションを促進する。2.3 節及び 2.4 節でその詳細を示す。このように、会話エージェントによる同調発現を系統的に捉えデザインした研究例はない。

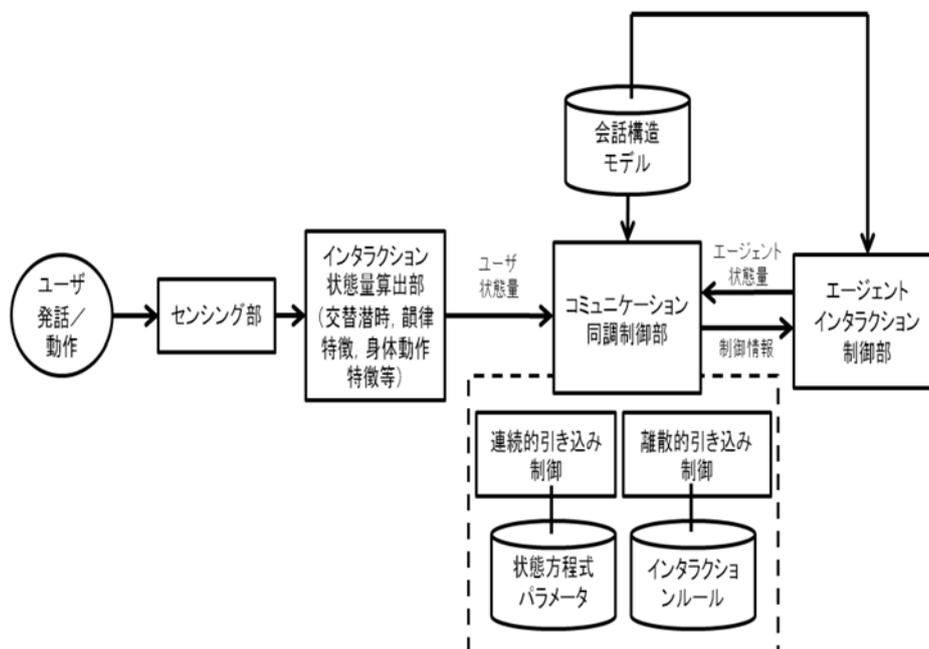


図 2.3 コミュニケーション同調制御方式の構成図

2.3 コミュニケーション同調制御

2.3.1 連続的引き込み制御

本研究では、引き込み制御の数理モデルとして、Van der Pol 方程式を用いる。Van der Pol 方程式は、引き込み特性を持つ非線型振動子である。非線形振動子は、ここで、大局的な時間的リズムを示す交替潜時の引き込み制御と、大局的な韻律的リズムを示すピッチの引き込み制御は独立であると仮定し、それぞれ定式化を行う。エージェントの時刻 (t) における交替潜時の変位を式 (3)、ピッチの変位を式 (4) で定義する。

$$\ddot{x}(t) = \mu(1 - x(t)^2)\dot{x}(t) - \omega^2x(t) + \alpha c(t) \quad (3)$$

$$\ddot{y}(t) = \mu(1 - y(t)^2)\dot{y}(t) - \omega^2y(t) + \beta p(t) \quad (4)$$

ここで、 ω 、 μ はそれぞれエネルギー保存係数、粘性係数である。 $c(t)$ 、 $p(t)$ はエージェントに同調させる対話者のパラ言語情報であり、対話者の交替潜時およびピッチのリズムを近似した \sin 波上の時刻 t におけるそれぞれの変位を表す。 α ($0 \leq \alpha \leq 1$)、 β ($0 \leq \beta \leq 1$) はその際に用いる引き込み係数で、これらを変化させることで対話者のパラ言語に対する同調のしやすさを調節する。 μ 、 ω 、 α 、 β のパラメータの値は、パラメータの値を刻み幅 0.1 で変更していくことで Van der Pol 方程式においてリミットサイクルが発現する値を求めた。そして、同調を通して変化した周期情報を用いてエージェントが次に表出する交替潜時やピッチを制御する。

一般的に、人間の脳波や心拍・呼吸なども一定の生体リズムを有しており、調和振動（非線形振動）が起こることが確認されている[12][13]。調和振動（非線形振動）は、微分方程式で表され、一般解は、正弦波となり振動する。筆者は、これらの知見から、交替潜時や韻律情報などの会話パラメータも、大局的には、非線形振動の挙動の中に支配されていると仮説をたてた。そこで、対話者の時刻 (t) における交替潜時の変位 $c(t)$ を式 (5)、ピッチの変位 $p(t)$ を式 (6) に定義する。

$$c(t) = \sin\left(\frac{2\pi t}{ks}\right) \quad (5)$$

$$p(t) = \sin\left(\frac{2\pi t}{kr}\right) \quad (6)$$

ここで、 s は対話者から観測される交替潜時長である。 r はピッチの変化率であり現在のターンの発話区間の基本周波数と 1 ターン前の発話区間の変動値として定義する。 k は波

の位相のスケーリングファクタで定数である。k の値は、事前実験の結果より決定する。対話者の交替潜時やピッチの正弦波は、交替潜時長 s やピッチ変化率 r を k 倍した周期をもつ波として表現できる。

交替潜時の同調制御のフローを図 2.4 に示す。人の交替潜時を周期とした正弦波とエージェントの交替潜時を周期とした正弦波を同調モデルを介して同調させる方式をとる。

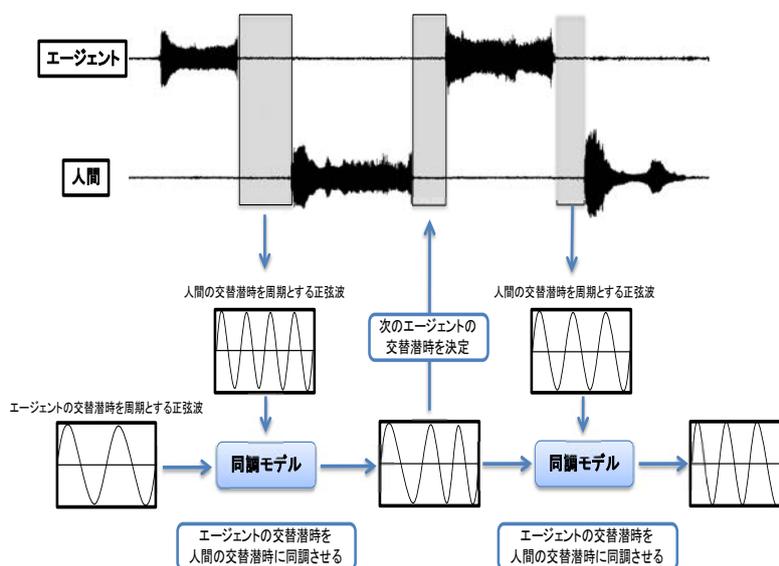


図 2.4 交替潜時の同調制御のフロー

次に、同調モデルによるピッチの制御方法を図 2.5 に示す。ピッチの制御は、交替潜時の制御と同じく人のピッチ変化率とエージェントのピッチ変化率を正弦波で近似し同調モデルを介して同調させる方式をとる。

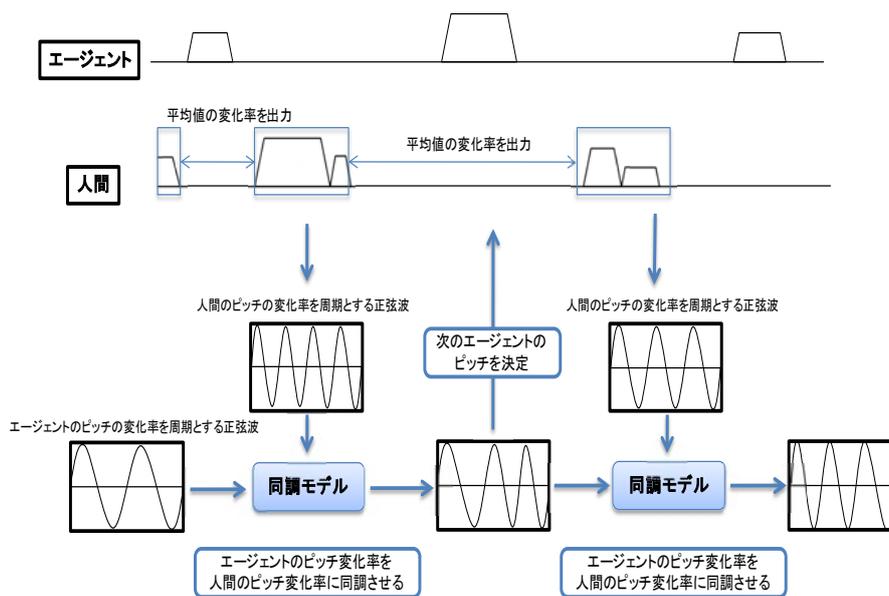


図 2.5 ピッチの同調制御のフロー

2.3.2 離散的引き込み制御

人間同士の対話では、適切なタイミングで相槌や頷きを返している。聞き手に相槌や頷きなどの反応がなければ、話し手は相手が自分の発話を理解しているのか不安になってしまう。したがって、相槌や頷きなどの現象は自然対話によるマンマシンインタフェースを考える上で非常に重要なインタフェースであると考えられる[14]。そこで、本研究ではエージェントに、発話開始および発話終了を意味する身体的インタラクションとして頷きを導入する。エージェントは、話者の発話内容について「話者に了解を求める」あるいは「話者の発話に対して賛同する」場合において頷くとする。エージェントには、人間の発話終了後に浅い頷きを2回行い、エージェントの発話終了後に深い頷きを、行うようにインタラクションルールを設定する。これにより、会話を成立させ、同調現象を引き起こす条件を調整する。

2.4 情報収集を目的とした会話構造モデルと引き込み制御

会話を成立させるには、適切な会話構造モデルが必要となる。本研究では、対話を通して問診を行う情報収集タスクを想定して、会話構造モデルを設計した。図 2.6 に示すように、エージェントは、まず相手から情報を聞き出し（聞き出しモード）、聞き出した情報に対して何らかの応答を行い必要に応じて共感し（応答・共感モード）、共感した情報に対して深く掘り下げる（エピソード展開モード）ことを繰り返す。このように、対話者の行動やエピソードに共感し、エージェントへの引き込みを促進しながら、情報収集を円滑に進めていく。各モードの発話内容例を表 2.1、表 2.2、表 2.3 にそれぞれ示す。先行研究[7]では、会話の共感時に話者間で韻律パラメータの相関があることが確認されており、本研究では、ここで提案する会話構造が、韻律パラメータにとって非線形振動子による同調現象を引き起こす前提条件であると仮定する。また、共感し、情報をさらにテンポよく聞き出すことを繰り返すことで、交替潜時についても同調現象を引き起こすことに繋がると想定する。なお、応答内容は人間の発話に関係なく質問内容に応じてランダムに決定される。

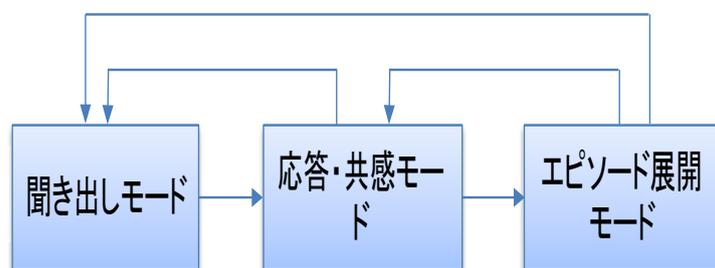


図 2.6 エージェントの会話構造モデルと引き込みの促進

表 2.1 聞き出しモードの発話内容

聞き出しモード	
今日の気分はどうですか？	最近、急激な体重の増減はありましたか？
定期的な運動をしていますか？	医師に高血圧と診断されたことはありますか？
規則正しい生活をおくっていますか？	アレルギーがありますか？
自分に適切な睡眠時間を取っていますか？	朝起きるのが辛いと思いますか？
お酒を飲みますか？	最近、イライラすることが増えましたか？
たばこを吸いますか？	肩の凝りはありますか？
定期検診を受けていますか？	手足のしびれを感じたことはありますか？
体力に自信がありますか？	食欲が無いことがありますか？
体は軽いですか？	腰の調子は悪くないですか？
食生活に気を付けていますか？	最近、身体冷えを感じることはありませんか？
現在、身体に関する悩みはありますか？	疲れを慢性的に感じていませんか？
急に気分が沈んだりすることはありますか？	視力が落ちてきたと感じますか？
現在服用している薬はありますか？	生活においてストレスを感じる事が多いですか？
過去に大きな病気になったことがありますか？	冷え性ですか？
生活環境に不安はありませんか？	毎日30分以上は歩いていますか？

表 2.2 応答・共感モードの発話内容

応答・共感モード
私も～に、チャレンジしたくなってきました。
やっぱり、自分のペースでやるのが一番いいですね。
嫌いなものはなかなか好きになれませんよね。
私も応援しています。これから続けていってくださいね。
最近は特に、色々考えさせられることが多いですね。
私もやってみたいと思います。
私も一緒にチャレンジさせてください。
私も好きです。好みが一緒ですね。食事の中に好きなものが入っているととても嬉しくなりますよね。
いいですね。是非私も行ってみたいです。
私も、～に行く事があれば、楽しみたいです。
わかります。私もこれから見てみたいです。
そうですか。

表 2.3 エピソード展開モードの発話内容

エピソード展開モード
～さんは～が趣味だとお伺いしましたが、～の良さを教えてください。
趣味の～は、どれくらいの頻度でされているのですか？
嫌いな食べ物～は～ですが、なぜ嫌いなんですか？
～さん、日課の～の楽しさを教えてください。
～の気になるニュースは、～ですが、なぜか教えて頂けますか？
～さんは何かチャレンジしたい事はありますか？
～さんは～が好きだということですが、～がなぜ好きなのか理由を教えてください。
～さんは～に旅行に行きたいとのことですが、～に行きたい理由を教えてください。
～さんは出身地が～ということですが、～の良いところを教えてください。
～さんが好きなテレビ番組は～ということですが、どこが好きなのか教えてください。

3. 提案手法の実装

提案方式は、図 2.3 に示すようなモジュール構成として実装した。インタラクション状態量算出部，コミュニケーション同調制御部，エージェントインタラクション制御部の各部の動作について示す。

3.1 インタラクション状態量算出部

インタラクション状態量算出モジュールは，マイクから入力された音声のパワーとピッチを用いて，人間の発話開始/終了時間とエージェントのピッチを決定する。インタラクション状態量算出モジュールは，マイクから入力された音声のパワーとピッチを用いて，人間の発話開始/終了時間とピッチの平均値を求め，その値を用いてエージェントの発話開始時間，ピッチの平均値を算出する。発話末の変化が多い日本語では，発話末の制御が必要であるが，本研究では詳細な引き込み制御ではなく大局的な引き込み制御を主眼に置いている。したがって，発話区間の代表値を用いて制御を行う。発話区間のピッチの代表値としては平均値と中央値が考えられる。ここで，事前実験でピッチの平均値と中央値の残差平均は $\pm 3.0 \sim \pm 3.5\%$ に収まっており，ピッチの変動が4%以内であれば，実験参加者は変化を感じないことが確認されている。本研究では，中央値より演算コストの低い平均値で十分であると判断した。また，音声パワーの基準値を $20 \times 10^{-6} \text{ Pa}$ とした。次式を用いて実験参加者の発話音声のパワーを算出する。

$$L_p = 20 \times \log_{10} (P/p_0) \quad (7)$$

(p : 音圧, p0 : 基準値)

発話開始・終了判定にはパワーを用いた。約 0.02 秒ごとにパワーを算出し記録する。そのデータを用いて約 1 秒間の単純移動平均を算出しその時点でのパワーとする。閾値を越えると発話とみなし、約 0.25 秒間発話し続けると発話開始とした。閾値を下回ると無音とみなし、約 0.75 秒間無音が続くと発話終了とした。これらの発話開始および発話終了の時間間隔は、高杉らの研究[15]を基に事前実験を行い、0.75 秒とした。また、発話音声から基本周波数を求めピッチとして算出した。算出方法には音声の基本周期を音声データから読み取り、その逆数を取る方法を用いた。発話開始から発話終了までの区間の音声データを用いて、約 0.02 秒ごとにピッチを算出し、区間内のその値の平均をその発話のピッチとした。算出した現在のターンのピッチと 1 ターン前のピッチの変動値をピッチ変化率とし、引き込みモジュールへは人間の発話開始/終了時間とピッチの変化率を送信する。ピッチ変化率の算出式を式 (8) に示す。

$$\text{ピッチの変化率} = \frac{\text{今回の発話ピッチ}}{\text{前回の発話ピッチ}} \quad (8)$$

3.2 エージェントインタラクション制御部

エージェントインタラクション制御部は、音声合成モジュールと動作生成モジュールから構成される。

3.2.1 音声合成モジュール

音声合成モジュールは、コミュニケーション制御同調部から受信したパラメータをもとに発話内容やピッチを制御し、再生を行うモジュールである。音声の再生終了時には、その終了時刻をコミュニケーション同調制御部へ送信する。音声合成エンジンとしては、ピッチ、パワー、スピード等の韻律情報、アクセント等が調整可能な「Voice Text」(HOYA 社)の女性音声版 (HARUKA) を用いた。ピッチの変動幅については、違和感を与えないという条件下で事前実験により算出し、基準値の±20%とした。また、音声合成による遅延は約30ms であり対話タイミングに影響がないことを確認した。会話内容は、会話構造モデル (図 2.6) の各モードに対応した発話パターン (表 2.1~表 2.3) に基づき生成する。エピソード展開モードの発話内容は、事前に獲得済みのプロフィールデータを参照することにより生成する。

3.2.2 動作生成モジュール

会話エージェントの表示制御を行うとともに、引き込みモジュールから受信した動作 ID をもとにエージェントの動作を制御するモジュールである。動作の種類は以下の 4 つとした。

- (1) 各動作が終了した後、停止する。
- (2) 音声合成モジュールが発話を開始した際に口を開閉させる。
- (3) 人間が発話終了した際に浅く 2 回頷く。
- (4) 音声合成モジュールが音声の再生を終了した際に深く 1 回頷く。

会話エージェントの動作例を図 2.7 に示す。

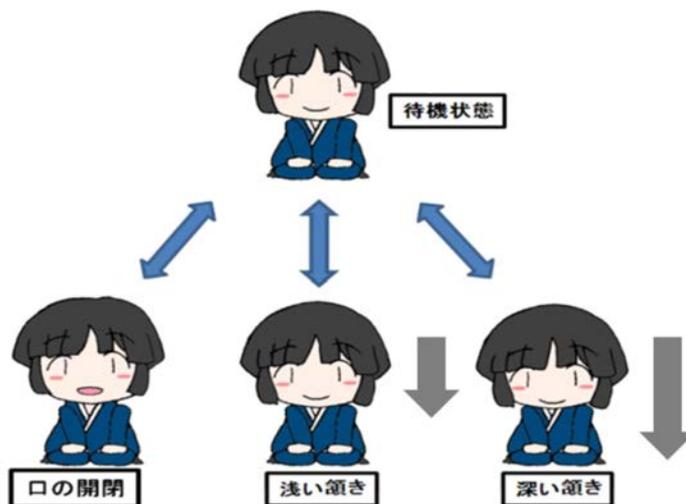


図 2.7 会話エージェントの動作例

4. 評価実験

4.1 実験目的

本実験では、提案する会話同調制御方式を会話エージェントへ実装し、エージェントの交替潜時及びピッチを、人の交替潜時及びピッチと同調させることにより、従来方式よりも、実験参加者が、「自然でストレスが少なく」かつ「協調的である」ことが言えるかどうかを検証することを目的とする。

4.2 比較制御方式

本研究では、提案する同調制御方式と比較検証する制御方式として、引き込み制御方式や引き込み制御パラメータが異なる 3 種類のエージェントを作成した。実験参加者とのこの

3種類のエージェントを会話させることで、提案方式に関する評価を行った。作成した3種類の引き込み制御方式の詳細を表 2.4 に示す。

表 2.4 実装した3種類の引き込み制御方式

方式	種類	パラメータ		説明
①	非同調方式	交替潜時	ピッチ	人間の交替潜時・ピッチを模倣したエージェント(比較方式)
②	同調方式	交替潜時	ピッチ	人間の交替潜時・ピッチと同調したエージェント(提案方式)
③	同調方式	交替潜時		人間の交替潜時と同調したエージェント(比較方式)

インタラクション状態量算出、会話構造モデル、離散的な引き込み制御などの条件は3方式とも同じとした。提案方式は②である。

比較方式①は、模倣という最も基本的な引き込みモデルで、人間の直前の交替潜時・ピッチという局所的な状態量に着目し直前の状態量を模倣するという制御方式とした。具体的には、人のピッチ・交替潜時をそれぞれ $P_H \cdot C_H$ 、エージェントのピッチと交替潜時をそれぞれ $P_A \cdot C_A$ とし、現在のターンを t とすると、エージェントの状態量 $P_A \cdot C_A$ は、式 (9) で算出できる。

$$P_A(t) = P_H(t-1)$$

$$C_A(t) = C_H(t-1)$$
(9)

コミュニケーション全体のダイナミクスを反映しておらず、直前の状態量のみに着目しているということで非同調方式と命名した。このように引き込み制御方式の全く異なる非

同調方式と対話の流れを反映した提案方式の効果の差を検証する。

比較方式③は、交替潜時のみの同調制御方式であり、提案方式②とは、同調に適用するパラメータが、交替潜時のみであるという違いがある。交替潜時のような時間的引き込み制御パラメータのみで同調させるという考え方[10]に基づく。時間的引き込み制御に韻律的引き込み制御を付加する効果を検証する。

4.3 実験方法

大学生 10 名を実験参加者とした。実験参加者は、対話エージェントが表示されるモニターに向かい合うように着席させた。実験中、実験参加者らはマイクを装着し、側面と後方に設置したビデオカメラにより実験風景を撮影した。本実験では音声認識は用いておらず、音声はピッチの平均値を算出するために使う。実験参加者からは、事前に、趣味や特技などエピソードに関する情報を取得し、会話エージェントシステムに実験参加者のプロフィールデータとして組み込んだ。実験参加者には、問診システムであるので質問されたことにのみ答えるように指示をした。エージェントの実験を行う順序はランダムとし、実験時間は 3 分とした。実験終了後、表 2.5 に示すようなアンケートを実施した。会話エージェントに期待される効果である「協調的であるか」という評価項目については、エージェントの親身度と捉え、「個人的な話をすることができたか？」と「エージェントが私に興味を持っていると感じたか？」の 2 項目を設定した。また、「ストレスが少ない」（会話そのものに対するストレス）という項目に関しては、直接、「ストレスなく話すことができたか？」という問いかけとした。以上のアンケート項目に対して、7 段階（7 が最もよい）のレベル

を設定した.

表 2.5 アンケートの項目

分類項目	質問内容	質問番号
親身度	個人的な話ができた	1
親身度	エージェントが私に興味を持っていると感じた	2
ストレス	ストレスなく話すことができた	3

4.4 実験結果

上記3項目（質問1，質問2，質問3）のアンケート結果を分析し，多重比較検定を行った結果，3つの項目ともに有意差が認められた．以下，分析結果の詳細を示す．

4.4.1 親身度の項目における分析

質問1「個人的な話ができた」の項目に対する各制御方式の平均値と標準偏差を，図2.8と表2.6に示す．図2.8と表2.6から，最もアンケート評価の平均値が高いのは提案方式②である．

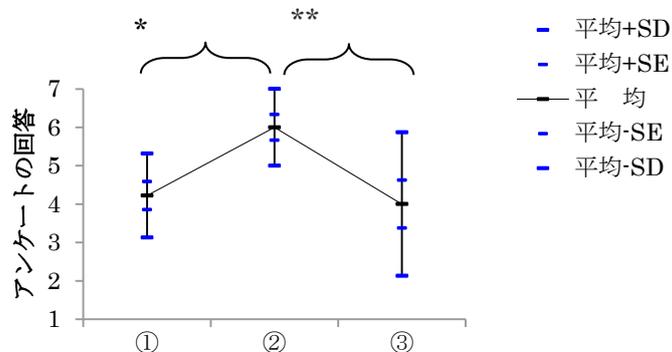


図 2.8 質問1の項目に対する各方式の平均値と標準偏差

表 2.6 質問 1 の項目に対する各方式の平均値と標準偏差

因子A	平均	標準偏差(SD)	平均-SD	平均+SD	標準誤差(平均-SE)	平均+SE
①	4.22	1.09	3.13	5.32	0.36	3.86 4.59
②	6.00	1.00	5.00	7.00	0.33	5.67 6.33
③	4.00	1.87	2.13	5.87	0.62	3.38 4.62

次に、質問 1 に対する分散分析表を表 2.7 に示す。分散分析の結果から提案方式②は、比較方式①と③に対して危険率 1%で有意差があった。さらに、提案方式②と比較方式①、および提案方式②と比較方式③との多重比較検定を Fisher の LSD（最小有意差検定）を用いて実施した結果を、表 2.8 に示す。提案方式②と比較方式①の間に危険率 5%で有意差があり、提案方式②と比較方式③の間に危険率 1%で有意差があった。以上のことから各方式間の評価は、②>①、②>③であると言える。これにより、親身度に関しては、比較方式よりも提案方式の方が、優位性があると言える。

表 2.7 質問 1 の項目に対する分散分析表

**:危険率1%有意 *:危険率5%有意

因子	平方和	自由度	平均平方	F 値	P 値	判定
因子A	21.63	2.00	10.81	5.70	0.01	**
誤差	45.56	24.00	1.90			
全体	67.19	26.00				

表 2.8 質問 1 の項目に対する多重比較検定の結果

**:危険率1%有意 *:危険率5%有意

水準1	水準2	平均1	平均2	差	統計量	P 値	判定
②	①	6.00	4.22	1.78	2.74	0.01	*
②	③	6.00	4.00	2.00	3.08	0.01	**

質問 2 の項目に対する各制御方式の平均値と標準偏差を、図 2.9 と表 2.9 に示す。図 2.9

と表 2.9 から、最もアンケート評価の平均値が高いのは提案方式である②である。

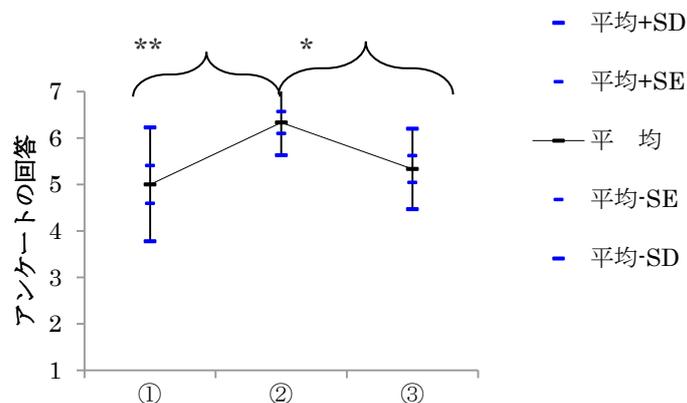


図 2.9 質問 2 の項目に対する各方式の平均値と標準偏差

表 2.9 質問 2 の項目に対する各方式の平均値と標準偏差

因子A	平均	標準偏差(SD)	平均-SD	平均+SD	標準誤差(SE)	平均-SE	平均+SE
①	5.00	1.22	3.78	6.22	0.41	4.59	5.41
②	6.33	0.71	5.63	7.04	0.24	6.10	6.57
③	5.33	0.87	4.47	6.20	0.29	5.04	5.62

次に、質問 2 に対する分散分析表を表 2.10 に示す。分散分析の結果から提案方式②は、比較方式①と③に対して危険率 5%で有意差があった。さらに、提案方式②と比較方式①、および提案方式②と比較方式③との多重比較検定を Fisher の LSD を用いて実施した結果を表 2.11 に示す。提案方式②と比較方式①の間に危険率 1%で有意差があり、提案方式②と比較方式③の間に危険率 5%で有意差があった。以上のことから各方式間の評価は、②>①、②>③であると言える。これにより、親身度に関しては、比較方式よりも提案方式の方が、

優位性があると言える。

表 2.10 質問2の項目に対する分散分析表

**:危険率1%有意 *:危険率5%有意

因子	平方和	自由度	平均平方	F 値	P 値	判定
因子A	8.67	2.00	4.33	4.73	0.02*	
誤差	22.00	24.00	0.92			
全体	30.67	26.00				

表 2.11 質問2の項目に対する多重比較検定の結果

**:危険率1%有意 *:危険率5%有意

水準1	水準2	平均1	平均2	差	統計量	P 値	判定
②	①	6.33	5.00	1.33	2.95	0.01**	
②	③	6.33	5.33	1.00	2.22	0.04*	

4.4.2 ストレスの項目における分析

質問3の項目に対する各制御方式の平均値と標準偏差を図 2.10 と表 2.12 に示す。図 2.10

と表 2.12 から、アンケート評価の平均が最も高いのは提案方式②である。

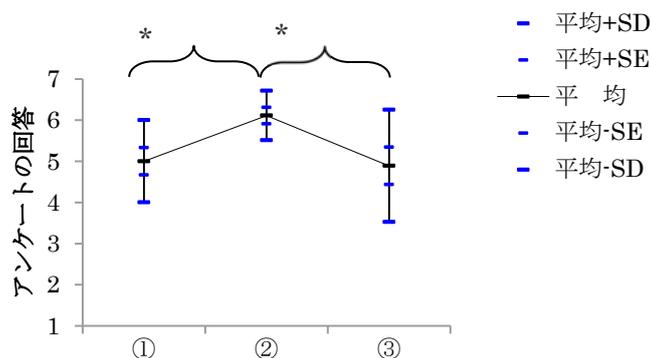


図 2.10 質問3の項目に対する各方式の平均値と標準偏差

表 2.12 質問3の項目に対する各方式の平均値と標準偏差

因子A	平均	標準偏差(SD)	平均-SD	平均+SD	標準誤差(SE)	平均-SE	平均+SE
①	5.00	1.00	4.00	6.00	0.33	4.67	5.33
②	6.11	0.60	5.51	6.71	0.20	5.91	6.31
③	4.89	1.36	3.52	6.25	0.45	4.43	5.34

次に、質問3の項目の分散分析表を表 2.13 に示す。分散分析の結果から提案方式②は、比較方式①と③に対して危険率5%で有意差があった。さらに、提案方式②と比較方式①、および提案方式②と比較方式③との多重比較検定を、FisherのLSDを用いた結果を表 2.14 に示す。表 2.14 から、提案方式②と比較方式①の間に危険率5%で有意差があり、提案②と比較方式③の間に危険率5%で有意差が確認された。以上のことから各方式間の評価は、②>①、②>③であると言える。これにより、ストレス性に関しても、比較方式よりも提案方式の方が、優位性があると言える。

表 2.13 質問3の項目に対する分散分析表

**:危険率1%有意 *:危険率5%有意

因子	平方和	自由度	平均平方	F 値	P 値	判定
因子A	8.22	2.00	4.11	3.83	0.04	*
誤差	25.78	24.00	1.07			
全体	34.00	26.00				

表 2.14 質問3の項目に対する多重比較検定の結果

**:危険率1%有意 *:危険率5%有意

水準1	水準2	平均1	平均2	差	統計量	P 値	判定
②	①	6.111	5.000	1.111	2.274	0.032	*
②	③	6.111	4.889	1.222	2.502	0.020	*

4.5 考察

実験結果（提案方式②と比較方式①の比較）より，対話者の直前のコミュニケーション状態量を模倣する方式（非同調方式）と較べて，提案の同調方式の方が，親身度を向上させ，ストレスを軽減させる効果があることが確認された．提案の同調方式は，対話者のコミュニケーション状態量を模倣するのではなく，会話全体のコミュニケーションのゆらぎに着目し，そのゆらぎの中で引き込みを起こし同調現象を発現させることにより，より協調的で自然でストレスのない会話を実現できたと解釈できる．図 2.11 および図 2.12 は，エージェントと実験参加者の会話開始時点から 110 秒までの交替潜時とピッチ変化の推移を図示したものである．図 2.11 と図 2.12 から，エージェントの交替潜時とピッチ変化が，実験参加者の交替潜時とピッチ変化に合わせて同調しているのが確認された．また，図 2.11 の 20 秒付近と 80 秒付近は実験参加者が考え込んでしまった場面で，交替潜時が長くなっているが，交替潜時が長くなってもエージェントは人の交替潜時に動的に追従していることが確認できる

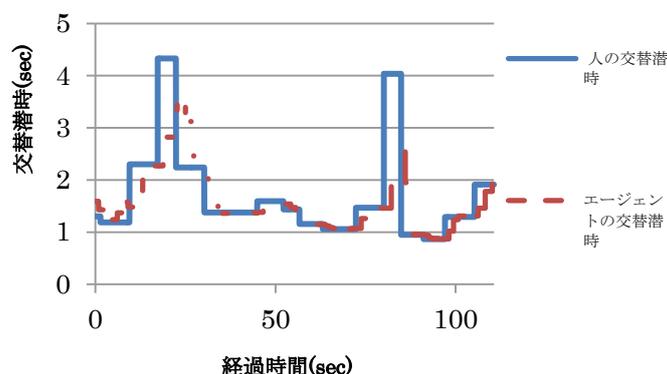


図 2.11 人とエージェントの交替潜時の変化.

また，図 2.12 のピッチ変化率では，50-60 秒付近にかけて，ピッチ変化が大きくなってお

り、この付近では、ビデオ観察により人とエージェントの会話が盛り上がっていた。

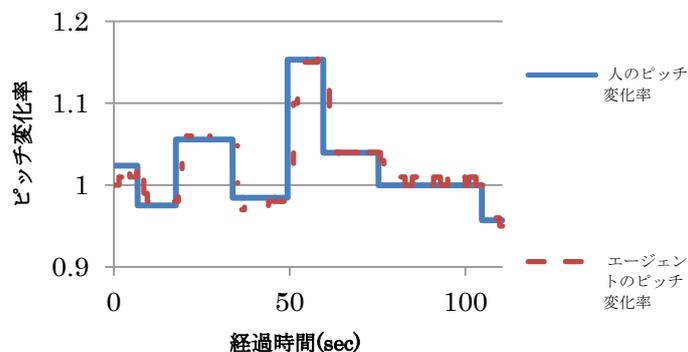


図 2.12 人とエージェントのピッチ変化

このように、本提案方式により、人とエージェント間の対話の流れの中で、エージェントが人のコミュニケーション状態量の変化に適応的に振る舞うことができ、かつ、一旦同調を発現すると、多少の変動があっても、安定的にコミュニケーション状態を維持することが可能となる。これは、システムの挙動が一旦、アトラクタの中に入れば、多少のノイズがあっても頑健に振る舞うことができるという非線形振動子の挙動特性を裏付けている [12][16].

さらに、提案方式②と比較方式③を比較した実験結果により、本同調方式の引き込み制御パラメータとして、交替潜時だけでなく、韻律情報（ピッチ）が、協調的で自然でストレスのない会話の実現に寄与していることが裏付けられた。このように、韻律的なリズムの同調は会話全体の調整に大きな役割を果たしていると言える。

親身度およびストレス軽減の評価について、提案方式に基づく会話エージェントとして、

両方とも平均 6 点（7 点満点）以上と高い評価を得ることができた。このことは、図 2.3 で示した同調制御方式に基づく会話エージェントシステムが有効に機能したと言える。さらに、3 分間の会話において、平均 10 個程度の間診情報を安定的に獲得できており、情報収集エージェントとしても有効に動作した。

5. おわりに

本論文では、非線形振動子による連続的で大局的な引き込み制御とインタラクションルールによる離散的で局所的な引き込み制御に基づきコミュニケーション同調制御を行い、共感を引き起こす会話構造モデルのもとで会話状態をダイナミックに遷移させ、コミュニケーションを促進するエージェント会話制御方式を提案した。さらに、提案した会話同調制御方式を情報収集を目的とした会話エージェントへ実装し、有効性の検証を行った。その結果、

- (1) 提案方式は、対話者のコミュニケーション状態量を模倣する制御方式と較べて、より協調的でストレスのない引き込み制御ができる
- (2) 会話エージェントは、人のコミュニケーション状態量の変化に適応的に振る舞うことができる
- (3) 交替潜時だけでなく、韻律情報（ピッチ）は、協調的でストレスの少ない会話の実現に寄与し、韻律的なリズムの同調は会話全体の調整に大きな役割を果たしていることが明らかとなり、これらを実現するシステム仕様について明確化した。

また、本実験を通して、より親しみが持てストレスの少ない会話のもとで持続的な情報収

集が行える会話システムの実現可能性を示すことができた。

参考文献

- [1] 木村昌紀, 余語真夫, 大坊郁夫: 感情エピソードの会話場面における同調傾向の検討ー擬似同調傾向実験パラダイムによる測定ー, 対人社会心理学研究, Vol.4, pp.92-99(2004).
- [2] 長岡千賀, Maria Draguna, 小森政嗣ほか: 音声対話における交替潜時が対人認知に及ぼす影響, ヒューマンインタフェースシンポジウム予稿集, Vol.2002, pp.171-174(2002).
- [3] 長岡千賀, 小森政嗣, Draguna Raluca Maria ほか: 協調的対話における音声行動の2者間の一致-意見固持型対話と聞き入れ型対話の比較-, ヒューマンインタフェースシンポジウム 2003 論文, Vol.2003, pp.167-170 (2003).
- [4] 三宅美博, 辰巳勇臣, 杉原史郎: 交互発話における発話長と発話間隔の時間的階層性, 計測自動制御学会論文集, Vol.40, No.6, pp.670-678 (2004).
- [5] 小松孝徳, 森川幸治: 人間と人工物との対話コミュニケーションにおける発話速度の引き込み現象, 情報処理学会研究報告-知能と複雑系, Vol.2004, No.105, pp.71-78 (2004).
- [6] 垣田邦子: 簡単な“質問一答”形式の対話における F0 の話者間相互作用, 日本音響学会研究発表会, Vol.1995, No.2, pp.305-306 (1995).
- [7] 西村良太, 北岡教英, 中川聖一: 音声対話における韻律変化をもたらす要因分析, Journal of the Phonetic Society of Japan, Vol.13, No.3, pp.66-84 (2009).
- [8] 志和敏之, 神田崇行, 今井倫太ほか: 対話ロボットの反応時間と反応遅延時における間投詞の効果, 日本ロボット学会誌, Vol.27, No.1, pp.87-95(2009).
- [9] 渡辺富夫: 身体的コミュニケーションにおける引き込みと身体性ー心が通う身体的コ

コミュニケーションシステム E-COSMIC の開発を通してー, ベビーサイエンス, Vol.2,
pp.4-12(2003).

[10] 林貴宣, 加藤昇平, 伊藤英則: パラ言語に基づいた会話ロボットの精神リズム同調モデル, 人工知能学会全国大会論文集, Vol.23, pp.ROMBUNNO.1H2-4(2009).

[11] 伊藤宏司: 身体知システム論, 共立出版(2005).

[12] 蔵本由紀: リズム現象の世界, 非線形・非平衡現象の数理①, 東京大学出版会(2005).

[13] ハーマン・ハーケン: 脳機能の原理を探るー非平衡協同現象としての脳神経活動・行動・認識ー, シュプリンガー・フェアラーク東京 (2000).

[14] 市川熹, 堀内靖雄, 西田昌史: マルチモーダル音声対話コーパスの収録とうなずきの分析, 広瀬・峯松研究室 (オンライン), 入手先
(http://www.gavo.t.u-tokyo.ac.jp/tokutei_pub/houkoku/corpus/ichikawa.pdf)(参照 2013-01-06).

[15] 高杉 将司, 山本知仁, 武藤ゆみ子ほか: コミュニケーションロボットとの対話を用いた発話と身振りのタイミング機構の分析, 計測自動制御学会論文集, Vol.45, No.4,
pp.215-223(2009).

[16] 多賀巖太郎: 脳と身体の動的デザインー運動・知覚の非線形力学と発達ー, 金子書房 (2002).

3章 初対面期の会話を促す仲介エージェント

～社会的スキルを考慮した身体的引き込み制御に基づくグループコミュニケーションのデザイン～

1. はじめに

孤独な高齢者がコミュニティへ参加することが、高齢者の QoL 向上につながると言われている[1]。しかしながら、高齢者と近所の人たちとの交流の弱まりや、約半数の高齢者がグループ活動などのコミュニティに参加していないといった孤独な高齢者の存在が問題視されている[2]。

Levinger らは、未知の人間が出会い知りあう過程において、お互いの新密度を高める段階の前に、お互いの形式的な情報を交換する段階があるとしている[3]。つまり、図 3.1 に示すように、孤独な高齢者がコミュニティへ初めて参加し溶け込むまでの過程として、初対面期と親密度向上期の 2 段階がある。初対面期は、コミュニティへ参加しようとする高齢者とコミュニティに所属する人間がお互いの情報を交換し合い、対話者について知っている情報量を増やす段階である。一方、親密度向上期は、初対面期の後、コミュニティへ参加しようとする高齢者とコミュニティに属する人間が親密度を高め高齢者がコミュニティに溶け込む段階である。高齢者をコミュニティに参加させるためには、最初の段階として、初対面期の情報交換を促進しお互いの情報量を増やすことが重要となる。

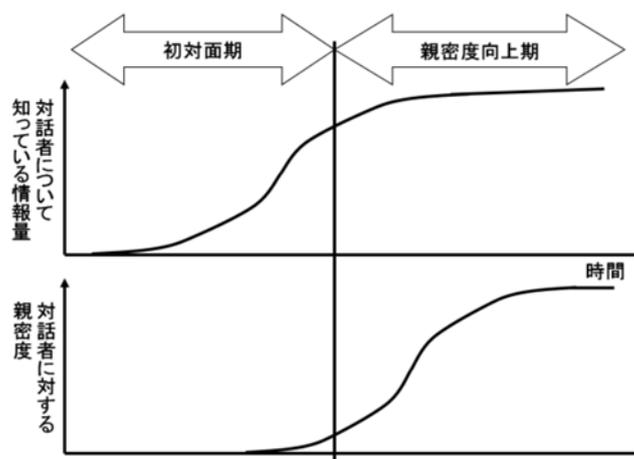


図 3.1 初対面期と親密度向上期

しかしながら、初対面期においては、お互いの情報交換が十分にできない問題がある。

この問題の原因は、初対面者がお互いを知らないためや、会話するには心理的障壁が高いこと、会話のきっかけがない、ためである。そこで、初対面期における問題の解決策を以下の3つに分類した。

- (1) **他人による仲介** 仲介者が初対面者の人間を仲介する
- (2) **環境による支援** 音楽や食事などによって環境を作る
- (3) **自助努力** 人間の初対面期における情報交換スキルを高める

本章では、初対面期の情報交換の促進する (1) の解決策として、初対面者同士を仲介するエージェントの実現方法およびコミュニケーション制御方式について述べる。これらが実現すれば、例えば老人福祉施設などで仲介者として期待されている介護職員の負荷を軽減することが可能になる。

エージェントが初対面者同士を円滑に仲介するためには、自己紹介などの話題提供に加

えて身体動作生成やタイミング制御による引き込み制御も重要となる。人間と円滑なコミュニケーションを行うロボットの代表的な例として神田らや渡辺らの研究があげられる。神田らは、視線や指差しなどの身体動作に着目し身体表現のために人間と同様に動く頭部や腕を持つ日常活動型ロボットの Robovie を開発している[4]。渡辺らは頷きなどのノンバーバル情報によるコミュニケーション時のリズム共有・相互引き込みに焦点をあて、発話者の音声情報に基づき頷き等の動作タイミングを推定する機構をアバターやロボットの制御に組み込んだ InterActor, InterRobot をコミュニケーション支援に適用している[5]。初対面コミュニケーション支援に関する研究例としては、中西らによる異文化コミュニケーション支援エージェント[6]や、栗山らによる共通の知人に関する話題を提供するエージェントがある[7]。

本研究では、初対面でのコミュニケーション支援が可能なエージェントの実現を目指し、初対面コミュニケーションにおけるコミュニケーション制御方式の提案と初対面仲介エージェントへの実装を行う。また、エージェントが初対面コミュニケーションにおいて有効であるかを検証するために評価実験を行った。

2. 初対面者の仲介モデル

2.1 初対面者の仲介モデルと仲介者の振る舞い

本章では、複数人の初対面仲介の基本として考えられる1人の仲介者による2人の初対面仲介をとり上げる。3人以上の会話における人間の役割に関する社会学の知見[8]に基づきコミュニケーション場の状態を分類し、この状態に応じて仲介者が異なる振る舞いを実施

する4層構造モデルを提案する。この4層構造モデルの第1層は、コミュニケーション場の4つの状態遷移モデルである(図3.2(a))。コミュニケーション場において、仲介者および2人の初対面者は5要素の仲介行動(図3.2(b))の組み合わせで、コミュニケーションを実行するとモデル化できる。4状態の各状態は5つの仲介行動の組み合わせとして記述でき、さらに5つの仲介行動はエージェントの動作9つの動作の組み合わせで記述できる(図3.2(c))。また、動作制御はタイミング制御により実行される(図3.2(d))。

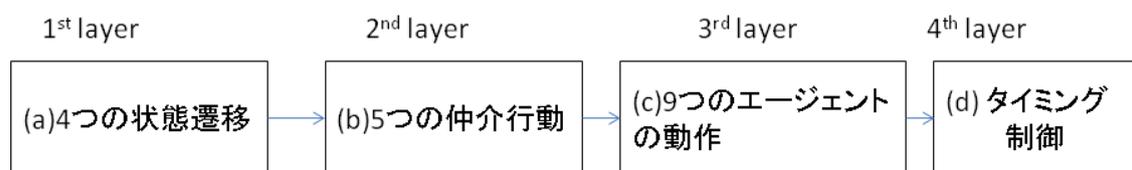


図3.2 4層構造モデル

2.2 初対面仲介における4つの状態とその遷移モデル

3人以上の会話における人間の役割に関する社会学的知見を基に、人間の仲介者が2人の初対面者を仲介するコミュニケーション場の状態を分類する[9]。Ervingらは、3人以上の会話において、人間は話し手、受け手、傍参与者の3種類の役割を動的に交代しながら担っている。傍参与者は会話を聞いているが、発話しておらず、発話を受けてもいない人間の事を指す。

筆者は、まず人間の仲介者が話し手、受け手、傍参与者のどの役割を担おうとするかによってコミュニケーション場を話題提供状態、話題掘り下げ状態、会話促し状態に分類し

た。そして、上記の3状態の初期状態として挨拶状態を導入し、初対面コミュニケーション場の状態を4状態に分類した(図3.3)。

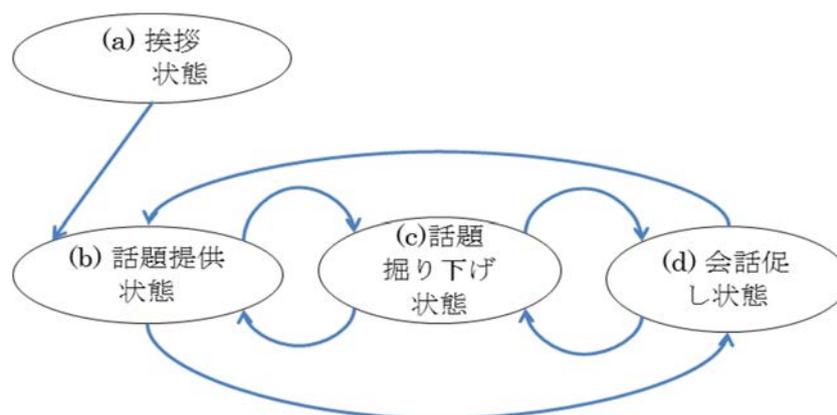


図 3.3 コミュニケーション場における4つの状態遷移モデル

2.3 初対面における5つの仲介行動

人間同士の対話において聞き手は適切なタイミングで相槌や頷きを返している。聞き手に相槌や頷きなどの反応がなければ、話し手は相手が自分の会話を理解しているのか不安になるからである。したがって、協調的な動作はマンマシンインタフェースを考える上で非常に重要なインタフェースであると考えられる[9]。また、初対面期におけるコミュニケーションは社会的相互関係を築く段階のコミュニケーションであり、初対面者同士のコミュニケーションを円滑にするため適切な仲介行動を導入する必要がある。

まず、2人の初対面者と1人の仲介者による、初対面仲介は、2人の初対面者についてある程度情報を持っている仲介者が初対面者に働きかけることで1人の初対面者と仲介者の2項関係を築くプロセスと、2人の初対面者と仲介者の3項関係を築くプロセスから成り立つ

ている。

仲介の初期段階では、仲介者が2人の初対面者に趣味や特技といった話題を提供（発話誘導）することで、話しやすいきっかけを作り初対面者と仲介者の2項関係を築く。

次に、初対面者同士の会話の話し手あるいは聞き手と同調（話し手同調、聞き手同調）して相槌や頷きを行うことで、会話を盛り上げることで話者と2人の初対面者の3項関係に押し進めていく。また、会話していても視線（顔向き）が一致していなければ、発話を促しつつ顔向きを初対面者に向けさせることで、視線の誘導を行う（視線誘導）。

初対面者同士の会話が盛り上がれば視線を2人の初心者に配分するだけとし聞き役に徹する（視線配分）。会話が途切れた場合には、発話誘導からやりなおす。

筆者は、仲介行動を以下の5つに分類した。

- (1) **発話誘導** 発話誘導は、会話が途切れたときに、仲介者が最もよく行う仲介行動であり、最も頻度が高かった。4状態が挨拶状態から話題提供状態、話題掘り下げ状態、会話促し状態とコミュニケーションが活性化するにつれて減少する。話者のどちらか、もしくは両方が、社会的スキルが低い場合は、状態の進展が遅く、発話誘導の場面が多くなる。
- (2) **視線誘導** 視線誘導は、一方の参加者に関連する話題を仲介者が提供をしたときに、もう一方の参加者にその話題を介して、参加者同士が会話をするように仕向ける仲介行動であり、話題提供状態と話題掘り下げ状態に限定される。
- (3) **視線配分** 視線配分は、2人の会話が進んで、聞き手状態（会話促し状態）にな

ったときに、話が途切れないように注意を配る仲介行動であり、あまり頻繁に注意を配ると不自然さになるので、不自然さがでないよううまく視線配分をしている様子が観察された。

- (4) **聞き手同調** 聞き手同調は、聞き手の頷きや相槌のタイミングに同調する仲介行動で、話し手同調に較べれば受動的であり、コミュニケーションが活性化してくるに従って頻度が増加する。
- (5) **話し手同調** 話し手同調は、聞き手に視線を向けながら、話し手のリズムに同調し、聞き手の同調を誘う仲介行動で、話題提供を行って、話題に乗ってきた話し手に、聞き手の注意を促す場面によく見受けられ、頻度が多い。

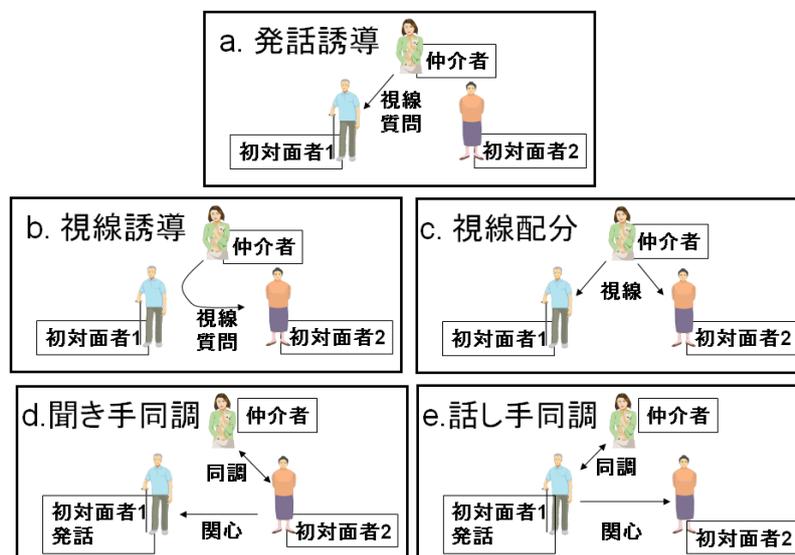


図 3.4 5つの仲介行動

図 3.4 において、被仲介者に対する仲介行動の働きかけはそれぞれ異なる。これら 5つ

の仲介行動の働きかけを図 3.5 に示す。図 3.5 では、仲介者を R、2 人の初対面者をそれぞれ H1・H2 と表記している。また、仲介行動の働きかけは矢印の向きで表記している。

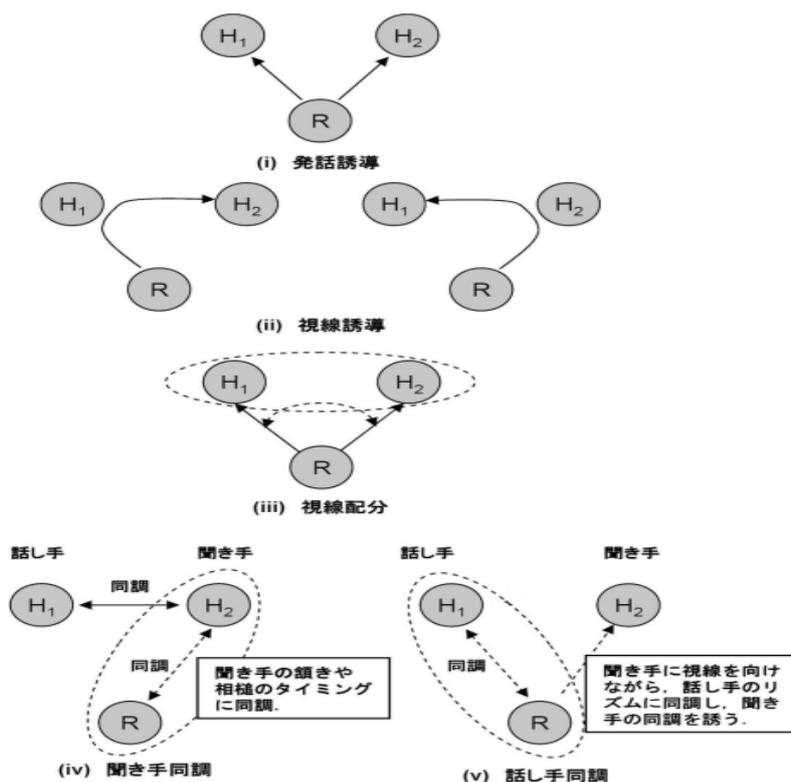


図 3.5 5つの仲介行動の働きかけ

この仲介行動の分類が妥当であるかどうかは、4章で検証する。

2.4 仲介行動と場の遷移

仲介者および2人の初対面者は、コミュニケーション場の4つの状態に応じて5要素の仲介行動を実施する。コミュニケーション場の4つの状態を次に示す。また、仲介者と2人の初対面者がコミュニケーション場の状態と実施される仲介行動の対応を表 3.1 に示す。

- (1) **挨拶状態** 挨拶状態では、仲介者と初対面者 H1 および仲介者と初対面者 H2 との会話の発生をめざし、仲介者と 2 人の初対面者の仲介行動(1)を行う。仲介者は、初対面者のプロフィールに基づき、初対面者 H1 および H2 に対して別々に発話誘導を行う。また、エージェントと各被紹介者との視線一致率、および平均発話量が、ともに閾値以上であれば話題提供状態に遷移する。
- (2) **話題提供状態** 話題提供状態では、2 人の初対面者が向き合って会話することを目指し、仲介者と 2 人の初対面者の仲介行動 (3) 以外を実施する。仲介者は、一方の初対面者の情報を他方に伝える。そして、共通の話題について 2 人に意見を述べさせる。最後に、視線誘導を行い向かい合って対話するように誘導する。次の状態への遷移は、被紹介者間の視線一致率が閾値以上であれば、話題掘り下げ状態へ遷移する。
- (3) **話題掘り下げ状態** 話題掘り下げ状態では、仲介者と 2 人の初対面者との 3 者の会話の発生をめざし、仲介者と 2 人の初対面者の仲介行動(3)以外を実施する。仲介者は、質問を投げかけて 2 人の初対面者のコミュニケーションへ参入する。また、被紹介者間の視線一致率が閾値以上であり、かつ、平均発話量が閾値以上であれば会話促し状態へ遷移する。
- (4) **会話促し状態** 会話促し状態では、仲介者が聞き役として初対面者の会話に同調して会話を盛り上げることを目指し、仲介者と 2 人の仲介行動 (3) ~ (5) を実施する。仲介者は適切なタイミングで頷きと相槌を行い、2 人の初対面者の会話に同調する。また、視線一致率や平均発話量が減少し、閾値を下回った場合は、前の状態に遷移する。

各状態において状態遷移を起こす状態量の閾値は、事前実験により決定した。

表 3.1 コミュニケーション場の各状態における仲介行動

	(a) 挨拶状態	(b) 話題提供状態	(c) 話題掘り下げ状態	(d) 会話促し状態
(1) 発話誘導	○	○	○	x
(2) 視線誘導	x	○	○	x
(3) 視線配分	x	x	x	○
(4) 聞き手同調	x	△	○	○
(5) 話し手同調	x	○	○	○

3. エージェントへの実装

3.1 4層構造モデルのエージェント

仲介者と2人の初対面者の仲介行動とコミュニケーション場の4つの状態遷移をエージェントを用いて実現する。そのために、仲介者及び2人の初対面者の仲介行動の基となるエージェントの9つの動作とコミュニケーション状態量を定義する。本研究では、エージェント初対面者間のコミュニケーション場の状態量として、時刻 t における過去 Δt 時間の被紹介者同士の視線一致率 $R_{\text{gazeShare}}$ と、時刻 t における過去 Δt 時間の被紹介者の平均発話量 $Average_{Pu}$ 、2者間の視線方向を定義した。視線一致率および平均発話量は、それぞれ (1) および式 (2) として定義する。視線方向は、カメラから推定した視線方向（顔向き）であり整数値で与えられる。

(1) 2者間の視線方向

(2) 視線一致率

$$R_{\text{gazeShare}}(x \leftrightarrow y, t, \Delta t) = \frac{\sum_{t_i=t-\Delta t}^t \text{Gaze}(x \leftrightarrow y, t_i)}{\Delta t} \quad (1)$$

(3) 平均発話量

$$Average_{Pu}(x \leftrightarrow y, t, \Delta t) = \frac{\sum_{t_i=t-\Delta t}^t \text{UtterancePower}(x \leftrightarrow y, t_i)}{\Delta t} + P_n \quad (2)$$

式 (2) において P_n は環境雑音を示しており，実験環境に応じて事前に観測した数値より決定する．

コミュニケーション状態量の変動により，コミュニケーション場の状態を遷移させる．

図 3.6 は，提案する初対面仲介エージェントを実現するシステムの概要である．本システムは 4 層構造となっており，それぞれのレイヤは上から順に 3 章で述べた 4 層構造モデルの (a) ~ (d) に対応する．また，コミュニケーションの状態量は，カメラとマイクを用いて初対面者の視線一致率と平均発話量を算出することで取得する．インタラクション制御は，コミュニケーション状態量より，推定したコミュニケーション場の状態に応じてエージェントから初対面者への適切なインタラクションを決定する．

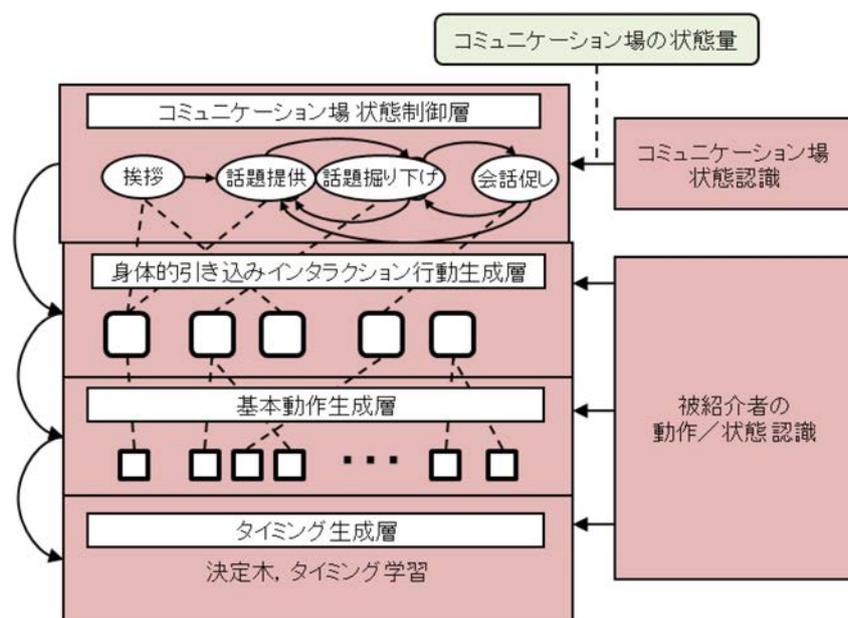


図 3.6 システム概要

3.2 エージェントの9つの動作

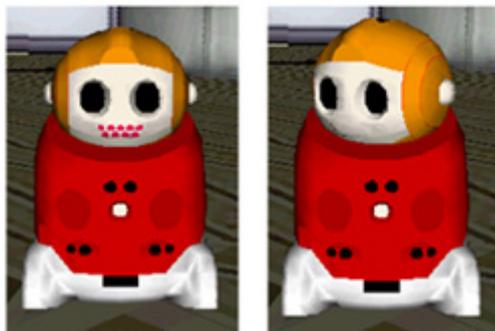
仲介者および2人の初対面者の5つの仲介行動を実現するための9つの動作を次に示す。

- (1) 頷く
- (2) 1の初対面者に対して発話する
- (3) 2人の初対面者に対して発話する
- (4) 発話を終了する
- (5) 1人の初対面者へ視線を向ける
- (6) エージェントと1人の初対面社者が同時に視線を向ける（視線一致状態）
- (7) 1人の初対面者へ顔を向ける
- (8) エージェントと1人の初対面者が同時に顔を向ける（対面状態）
- (9) 1人の対面者へ方向を指示する

本研究では、エージェントが1人の初対面者へ視線を向けるときは、顔も向けるとして
いる。図 3.7 に、エージェントの動作である、発話する（図 3.7 (a)）および視線を向ける
（図 3.7 (b)）様子を示す。図 3.7 に示すように、エージェントが発話する際にはエージェ
ントの口部分の赤色に点滅する。

また、エージェントがエージェントの右側にいる初対面者へ視線を向ける際には、顔を
右に回転する。例えば、エージェントと2人の初対面者の仲介行動である(1) 発話誘導は、
発話する、および視線を向ける 2 つの動作の組み合わせで実現する。エージェントは、初
対面者に対して発話を促すために、まず、初対面者に対して視線を向ける。そして、あら
かじめ入力された初対面者の趣味や特技といった情報を基に初対面者に対して質問を発話さ

せた.



(a) 発話

(b) 注視

図 3.7 エージェントの動作例

3.3 身体的引き込み制御におけるタイミング生成

エージェントの頷きと発話タイミング制御について述べる. 紹介エージェントは, 身体的引き込みを促進するために適切な動作を行うが, 動作タイミングの制御は引き込み制御に大きく影響する. そこで本研究のタイミング生成には, 初対面コミュニケーションにおける人間同士のインタラクションデータから, 図 3.8 に示す決定木を再帰的に構成して適用する方法を取る. 決定木から, 頷きと発話タイミングについて, 下記の 3 種類のルールが導出された. 各ルールにおけるパラメータは, 人間同士が初対面紹介をする事前実験で得られた結果を用いて決定した.

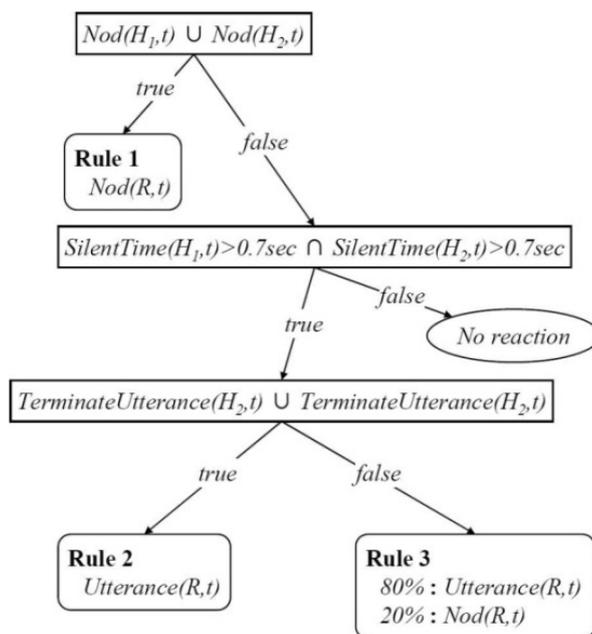


図 3.8 タイミング制御の決定木

Rule 1: 頷きタイミングルール 被紹介者の頷きに対する即応的な頷きを表している。

Rule 2: 発話タイミングルール 無音区間が一定時間 (0.7 秒)以上続き、文末として判断されるならば発話する。

Rule 3: 発話タイミングルール 相手の発話に応じた頷き・発話タイミングルール無音区間が一定時間 (0.7 秒)以上続き、文末ではない場合には 20 %の確率で頷く。または、無音区間が一定時間 (0.7 秒)以上続き、文末ではない場合でも 80 %の確率で発話する。

文末か否かの判定は、音声データに対して形態素解析を実行し、発話が、助詞、終助詞など文末に頻出する品詞で終了したかどうかで決定する。

4. 初対面における仲介行動の分析

4.1 仲介行動の観察実験

初対面仲介における仲介行動を分析するために、人間同士の初対面コミュニケーションにおける仲介行動の観察実験を行った。実験参加者は大学生 23 名とした。実験は、被仲介者 2 名と仲介者 1 名からなる 3 者間で実施した。実験時間は 10 - 15 分程度とし、実験の終了は仲介者の判断に任せた。なお、本実験では、実験参加者間の距離は、西出の対人距離モデル[10]に基づき、会話を強制する圧力を強く与える距離ではなく自然な会話が生まれる距離として 1.5m と設定した。

4.2 仲介行動の分析と分類

仲介行動の観察実験の結果をビデオ分析し、仲介行動の分類を行った。まず、初対面仲介での仲介行動を動作単位でみると動作パターンは、(1) 発話、(2) 話を聞く、(3) ジェスチャ、の 3 つの基本動作と (1) 相槌、(2) うなずき、の 2 つの補助動作から構成されており、これらの動作の組み合わせで仲介行動が行われていることが分かった。

そして、観察された仲介行動を 2 章で述べた 5 つに分類し、各仲介行動の発生頻度を調べた。図 3.9 に仲介行動のヒストグラムを示す。

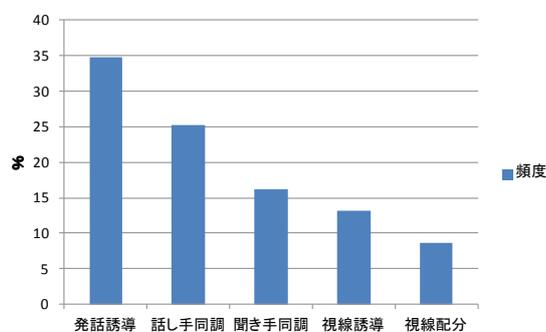


図 3.9 仲介行動のヒストグラム

図 3.9 のヒストグラムにおいて、発生頻度が高い順に発話誘導、話し手同調、聞き手同調、視線誘導、視線配分の順であった。発話誘導の頻度が最も高い理由は、会話が途切れた時点で発話誘導からやり直したことで、話し手同調・聞き手同調・視線誘導においても発話を誘発する要素があるからである。次に頻度が高い話し手同調は、仲介者が定期的に話題を提供し、少なからず発話内容への応答があるためである。聞き手同調が話し手同調より少ないのは、仲介者が聞き手に対して同調する行動であり、被仲介者に対して積極的に働きかける行動ではないためだと考えられる。視線誘導の頻度が少ないのは、初対面者同士が会話しているが向かい合っていない状態の時の仲介行動であり、初対面者同士ではあまり会話が行われていないからである。同様に視線配分が最も少ないのは、被仲介者同士が会話している場合であり、初対面仲介の場において初対面者同士が面と向かって発話する場面が少なかったからだと考えられる。

5. 評価実験

5.1 実験目的

初対面期の情報交換の促進を目指し、初対面者を仲介するモデルを実装したエージェントを用いた初対面仲介実験を行う。初対面仲介のコミュニケーションモデルの有効性を検証するため、本モデルを実装したエージェントと 2 人の初対面がコミュニケーションする実験（エージェント条件有り）と、エージェントを用いず 2 人の初対面者がコミュニケーションをする実験の 2 種類を実施した。初対面期における問題は、高齢者に限ったことではなく、実験参加者を高齢者とすると、音声の聞き取りなどの別の課題があるため、本実

験では高齢者以外を実験参加者とした。

5.2 実験手順

本実験におけるセンシング環境を図 3.10 に示す。図 3.10 に示すように、2 人の初対面者の視線（顔向き）を複数台のカメラによりリアルタイムで検出する（図 3.10 (e) ~ (f)）。本実験においてエージェントはディスプレイに表示される。図 3.10 (g) に示すように、初対面者の傾きは頭部に装着した加速度センサにより検出し、初対面者の発話は実験参加者に装着したマイクにより検出する（図 3.11）。

本実験は、大阪工業大学の実験室で実施した。実験参加者は、10 組の 2 人の初対面者であり、大阪工業大学の在學生 20 名とし、高齢者は実験対象から除外した。高齢者を除外したのは、エージェントの韻律の変化を弁別できなかったためである。エージェント条件ありとエージェント条件無しの 2 種類の実験をそれぞれ実験参加者の組み合わせを変えて 10 回ずつ計 20 回実施した。人間の仲介者が 2 人の初対面者を仲介する予備実験から、本実験の時間は、数回の話題の切り替えが起こる 6 分間とした。実験におけるエージェントの話題提供は、あらかじめエージェントに入力した初対面者の趣味や特技といった情報を用いた。

本実験で得られたコミュニケーション場の状態遷移の例を図 3.12 に示す。

このコミュニケーション場の状態が遷移した際のエージェントと対話者の発話内容及び動作例を以下に述べる。図 3.12 より、コミュニケーション場の盛り上がりに応じてコミュニケーション場が状態遷移していることが確認できる。図 3.12 の時間帯 A のコミュニケー

シヨンの状態は、挨拶状態であり、エージェント初対面者 1 と視線を一致させ会話することができた。時間帯 B のコミュニケーション場は話題提供状態であり、無言の時間が 5 秒程度続いた後、エージェントが初対面者 2 に対して新たな話題提供することができた。時間帯 C のコミュニケーション場の状態は会話促し状態であり、エージェントは聞き役として初対面者 1 と 2 の発話の後に、それぞれに視線を向けることができた。時間帯 A, B, C におけるエージェントと 2 人の会話例を以下に述べる。会話例において、エージェントは R, 2 人の初対面者はそれぞれ H1 および H2 と表記している。

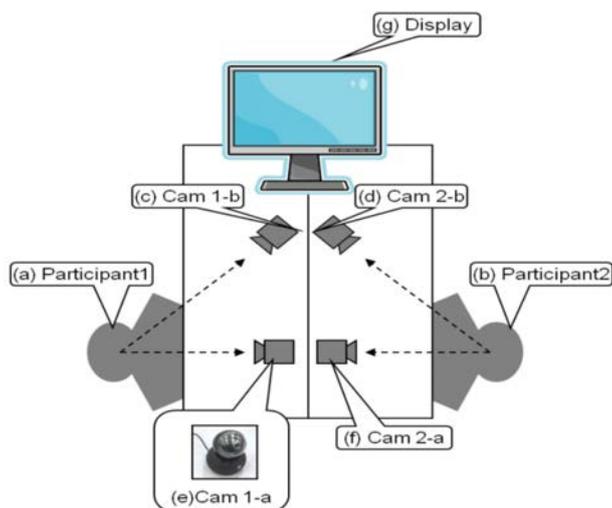


図 3.10 視線（顔向き）検出



図 3.11 加速度センサとマイク

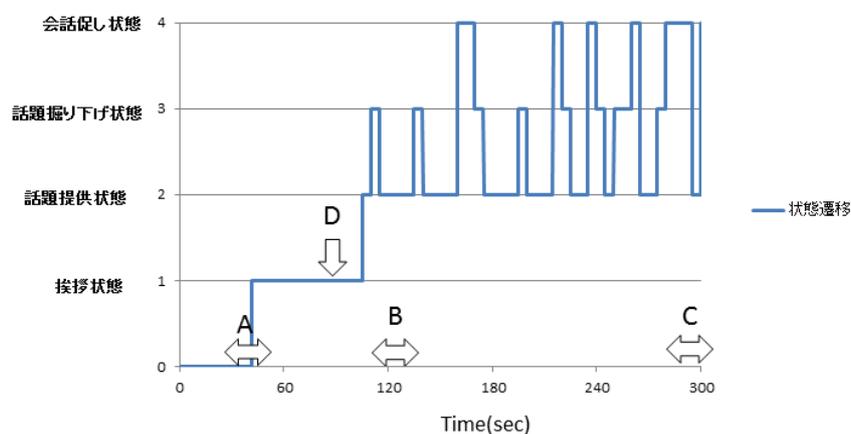


図 3.12 コミュニケーション場の状態遷移

時間帯 A

R:こちらの方は, (H1 の氏名) さんと言います.

H1:はじめまして.

R:こちらの方は, (H2 の氏名) さんと言います.

(エージェントが H1 に視線を向ける)

R:僕は枚方から来たんだけど, (H1 の氏名) さんは

今日どこからきたの？

(H1 が R に視線を向ける)

H1:大阪の平野区から来ました.

時間帯 B

H2:めっちゃ防寒着, 着まくって来てます.

H1:その服の上から?

H2:ええ.

(無言の発話が 5 秒程度続く)

(エージェントが H2 に視線を向ける)

(H2 がエージェントに視線を向ける)

R: (H2 の氏名) さんはテニスが得意なんだって

H1:テニスですか. 結構長いことやってるんですか?

H2:小学校から 10 年くらいやっていますね.

時間帯 C

H2:三重は田舎すぎてそんなことないですね.

全然関係ない.

(R が H2 に視線を向ける)

H1:犯人すぐ捕まるんじゃないですか?

H2:すぐ捕まる.

5.3 実験結果

エージェント条件有りとはエージェント条件無しとの2種類の実験終了後に実験参加者から取ったアンケートの集計結果を表3.2に示す。エージェントが初対面者の良好な関係を構築しつつ情報交換を促進できたかを検証するため、アンケートの項目は、「相手のことがよくわかったか」、「相手に好意が持てたか」、「会話は盛り上がったか」の3項目とした。実験参加者は、この3つのアンケート項目に対して5段階で回答した。

エージェント条件有りとはエージェント無しとの2種類の実験のアンケート結果をt検定を用いて比較した。t検定で比較した結果、「相手のことがよくわかったか」、「相手に好意が持てたか」の2つのアンケート項目に対して、有意水準5%の有意差で、エージェント条件有りの方が高い点数となった。

表 3.2 アンケート結果の統計情報

エージェント条件	統計	盛り上がった	相手のことがよくわかった	相手に好意を持てた
無し	平均値	3.9	3.5	4.05
	中央値	4	4	4.5
	標準偏差	1.165287	1.192079	1.316894
有り	平均値	4.1	4.1	4.6
	中央値	4	4	5
	標準偏差	0.788069	0.788069	0.502625

実験結果から、学生の実験参加者による実験室での限定された実験環境ではあるが、初対面仲介モデルの有効性を検証できたといえる。

本実験で観察されたコミュニケーション場の状態遷移、視線一致率、平均発話量を、それぞれ図3.12、図3.13、図3.14にそれぞれ示す。本実験において、話題提供状態と話題

掘り下げ状態を交互に状態遷移する様子が観察された。図 3.12 の 200sec から 240sec では、エージェントが話題提供するが、会話が盛り上がらなかったため再度話題提供することを繰り返している。

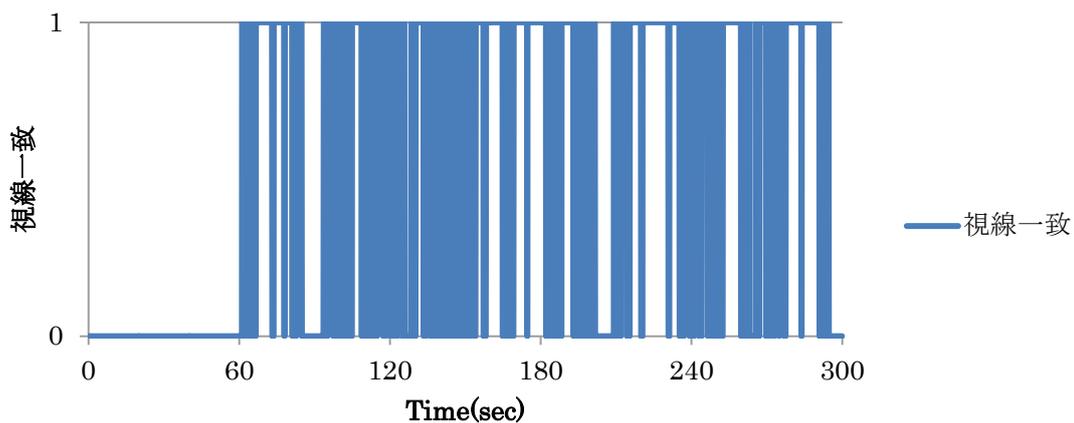


図 3.13 視線一致率

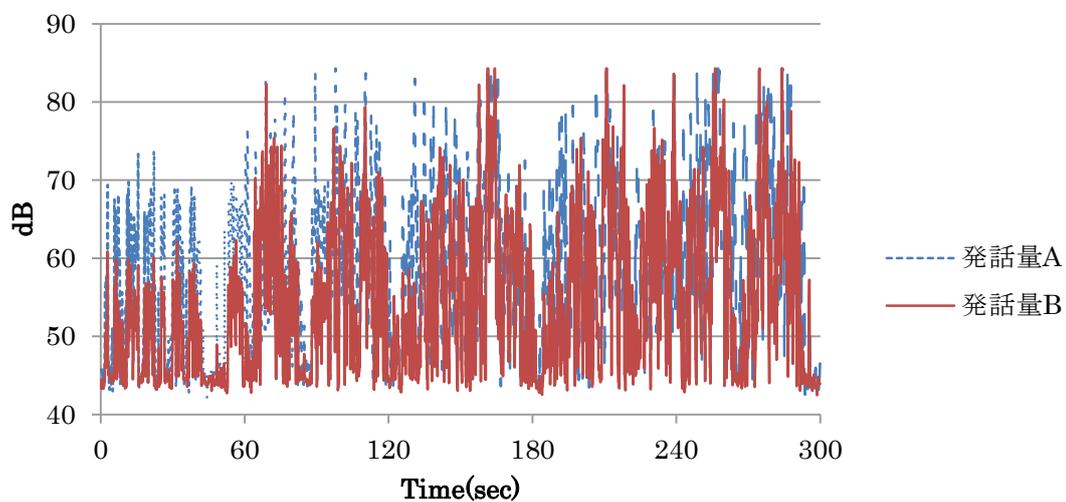


図 3.14 初対面者の平均発話量

したがって、初対面者の会話のセマンティクスを利用することなく、発話音量、初対面

者間の領きの同調度合いによって、本システムはコミュニケーションの盛り上がりの程度を推定できていると考えられる。

また、図 3.12 (D) あたりでは初対面者がエージェントを頼る様子が観察された。初対面者がエージェントを頼った理由は、エージェントが会話を促す機能をもっており、エージェントによって初対面者という状況の心理的負荷を軽減できると考えたためと考えられる。特に、話題掘り下げ状態および会話促し状態において、初対面者同士の話話が尽きかけると、初対面者がエージェントに話題提供を求めるような様子が観察された。この際、初対面者が視線をエージェントへ向けており、話題提供を求めているようであった。これより、より円滑にコミュニケーションを活性化するためには、初対面者のエージェントへの視線情報を用いて話題提供状態へ状態遷移することが有効であると考えられる。

6. 社会的スキルの観点から見たエージェントの分析

6.1 実験目的

筆者は、前章で5つの仲介行動の有用性を基本効果の検証を通じて明らかにした。本節では、社会的スキルの観点から「仲介行動における身体的引き込み要素が、社会的スキルが高い人にとっては影響を及ぼさないが、社会的スキルが低い人にとっては影響を及ぼし有効に働く」という仮説を立て、検証を行う。

社会的スキルが低い人は、社会的スキルが高い人よりも、話すタイミングを逃すことが多く、社会的スキルが高い人が一方的に話してしまうことがある。また、そのような状況を打開するために、発話誘導・視線配分・話し手同調・聞き手同調など5つの仲介行動の

身体的振る舞いが、社会的スキルの低い人にとって有効なコミュニケーション手段になっている、と考えられる。その検証を行うために、実験を行った。

6.2 実験手順

実験は、初対面の高齢者 2 人一組と、初対面者を仲介する仲介ロボットで行った。また、5 つの仲介行動において、視線一致（姿勢変化）や傾きがある場合とない場合の 2 種類の実験を行った。前章では、大学生に対する実験を行ったが、本章では、社会的スキルが低い人に対するコミュニケーション支援として、高齢者福祉施設（デイケアセンター）でのコミュニケーション支援への適用可能性を検証するため、実験参加者を高齢者とした。実験参加者は、高齢者の男性 18 名および女性 20 名の計 38 名である。なお、実験環境は前節の実験環境と同じである。

実験を行う前にアンケートを行い、アンケートから得た情報をロボットの質問内容に組み込んだ。実験参加者にはロボットが 2 人の仲介をすることと、質問内容は真ん中のディスプレイに表示されることを教示した。また、質問をきっかけとして、2 人で会話をするよう指示した。

各実験参加者は事前に社会的スキルの評価尺度である Kiss-18 指標を受けてもらい、社会的スキルの高いスキル高群と社会的スキルの低いスキル低群に分類した。各実験参加者の社会的スキルの統計量を表 3.3 に、社会的スキルの全国平均を表 3.4 に、それぞれ示す。

表 3.3 各実験参加者の社会的スキルの統計量

	社会的スキル（低群）	社会的スキル（高群）
サンプル数	19	19
平均	57.26	72.47
標準偏差	9.95	3.55

表 3.4 社会的スキルの全国平均

	全国平均	
男性	61.82	(n=45, SD=9.41)
女性	60.1	(n=121, SD=10.5)

6.3 実験結果

実験終了後、被験者にモーションなりの場合とモーションなしの場合についてアンケート(付録 A)を答えさせた。

また、アンケートに対して t 検定を実施した結果、アンケート項目の「有能であるか」の項目に対して、社会的スキルの高いグループに対しては、モーションありとモーションなしに有意差は見られなかった。表 3.5 にスキル低群の t 検定の結果を、表 3.6 にスキル高群の t 検定の結果をそれぞれ示す。

表 3.5 スキル低群の t 検定

*: 危険率5%有意差 ** : 危険率1%有意差	
t検定	
統計量:t	1.9509
自由度	16
両側P値	0.0688
片側P値	0.03441*

表 3.6 スキル高群の t 検定

*: 危険率5%有意差 **: 危険率1%有意差

t検定	
統計量:t	1.2865
自由度	18
両側P値	0.2146
片側P値	0.1073

しかし、社会的スキルの低いグループに対しては、モーションありのほうがモーションなしに比べて有意差があることがわかった。

これは、社会的スキルの高い実験参加者に比べ社会的スキルの低い実験参加者のほうがロボットの仲介行動が有効に働いたためだと考えられる。以上の結果から、社会的スキルが低い人にとって、5つの仲介行動の身体的振る舞い（視線制御や頷き）が有効に働くことが明らかとなった。

7. おわりに

本章では、人間が初対面仲介をする方法をモデル化し、このモデルをエージェントへ実装して、その有効性を検証した。実験では、このエージェントと2人の初対面者がコミュニケーションをする実験と、エージェントを用いず2人の初対面者がコミュニケーションをする2種類の実験を行った。実験のアンケート結果から、学生の実験参加者を用いた実験室での限定された実験結果ではあるが、提案モデルの有効性を検証することが出来た。

参考文献

- [1] 山崎勝男, 柿木昇治, 藤沢清, 宮田洋:新生理学 3 卷,新しい生理心理学への展望, 北大路書(1998).
- [2] 内閣府:高齢社会白書 平成 19 年度版, 内閣府(2007).
- [3] Levinger G., Snoek D.J.:Attraction in relationship: A new look at interpersonal attraction, General.Learning Press(1972).
- [4] 神田崇行, 石黒 浩, 小野哲雄, 今井倫太, 前田武志, 中津良平: 研究用プラットホームとしての日常活動型ロボット Robovie の開発, 電子情報通信学会論文誌. D-I, 情報・システム, I-情報処理, Vol.85, No.4, pp.380-389(2002).
- [5] 渡辺富夫: 心が通う身体的コミュニケーションシステム E-COSMIC, 電子情報通信学会技術研究報告. HCS, ヒューマンコミュニケーション基礎, Vol.101, No.114, pp.49-54(2001).
- [6] 中西英之, キャサリンイズビスタ, 石田 亨, クリフォードナス: 仮想空間内でのコミュニケーションを補助する社会的エージェントの設計, 情報処理学会論文誌, Vol.42, No.6, pp.1368-1376 (2001).
- [7] 栗山 進, 大平雅雄, 井垣 宏, 大杉直樹, 松本健一: Social Context-Aware Information System: 初対面時の「きまずさ」解消を目的とした実空間ソーシャルネットワークワーキングシステム, インタラクション 2006 論文集, 情報処理学会シンポジウムシリーズ, Vol.2006, No.4, pp.195-196 (2006).
- [8] 西出 和彦, "人と人との間の距離", 人間の心理・生態からの建築計画 (1), 建築と実務, No. 5, pp. 95-pp.99(1985).

[9] Erving G.: Forms of talk, University of Pennsylvania Press(1981).

[10] 市川熹, 堀内靖雄, 西田昌史:マルチモーダル音声対話コーパスの収録とうなずきの分析

4章 表情と韻律情報に基づくグループ感情

～グループコミュニケーションにおける感情パラメータの推定～

1. はじめに

近年、ロボットが生活に入り込み、支援を行う機会が増えてきている。ロボットが円滑な支援を行うためには、人間の状態に応じて適応的に振る舞うコミュニケーション機能が必要不可欠である。人と人がコミュニケーションを行うとき、文字や言葉などの言語情報よりも、視線や顔表情、ジェスチャ、音響情報などの非言語情報がより重要な役割を担うことが明らかにされている[1]。特に、音声や顔表情は豊富な情報を含んでおり、コミュニケーションの状態理解において極めて重要な要素であると言える。また、人間との円滑なコミュニケーションを支える基盤技術として、人間の感情を推定する技術は、心理学や社会学、人工知能、ロボット技術などへの応用が期待されており、個々人の感情を音声や画像系列から推定する方式に関する研究が数多くなされている[2][3]。一方、人間社会の中で活躍するロボットにとっては、個々の人間の感情を推定するだけではなく、人間の集団的な感情を推定することも、人間との円滑なコミュニケーションを実現する上で重要である。例えば、人間の集団的な感情を推定する機能をコミュニケーションロボットに搭載すれば、個人感情の遷移とグループ感情の遷移の状況を考慮して人間同士の会話のタイミングを見計らって会話に参入することなどが可能となり、より自然なコミュニケーションが可能なロボットの実現などが見込まれる。

しかしながら、複数の人間が構成するコミュニケーションの場全体の感情推定に関しては、まだ十分な検討がなされていない。例えば[4]では、複数人からなる集団の感情を推定するとしているが、個人の感情を介さずに、聴衆の視線量や注視点の量などから、直接集団感情を推定するものとなっている。このため、個人感情がグループ感情に及ぼす因果関係を扱っておらず、細やかな感情推定ができない。

そこで本章では、コミュニケーションロボットと複数の人間が存在する空間において、それぞれの人間の感情に由来する空間全体の感情をグループ感情モデルとして表現し、個人感情からグループ感情をこのモデルを用いて推定する手法について検討する。

2. グループ感情推定

2.1 システム概要

本節では、提案する感情認識システムの概略について述べる。図 4.1 に、感情認識システムの概略図を示す。

センシング入力部は、カメラから表情を観測するための映像情報、マイクから韻律感情を観測するために音声情報を入力する。センシング情報処理部は、映像情報から顔の表情を示す特徴量を、韻律情報から声の感情を示す特徴量を算出し、個人感情認識部に出力結果を送る。個人感情認識部およびグループ感情認識部では、次節で説明するグループ感情場モデルを用い、時間推移の中で個人感情およびグループ感情を認識する。認識結果出力部では、認識された感情が確率分布の形で出力される。また、学習機構として、個人環境状態遷移モデル生成部、グループ感情状態遷移モデル生成部を備えている。

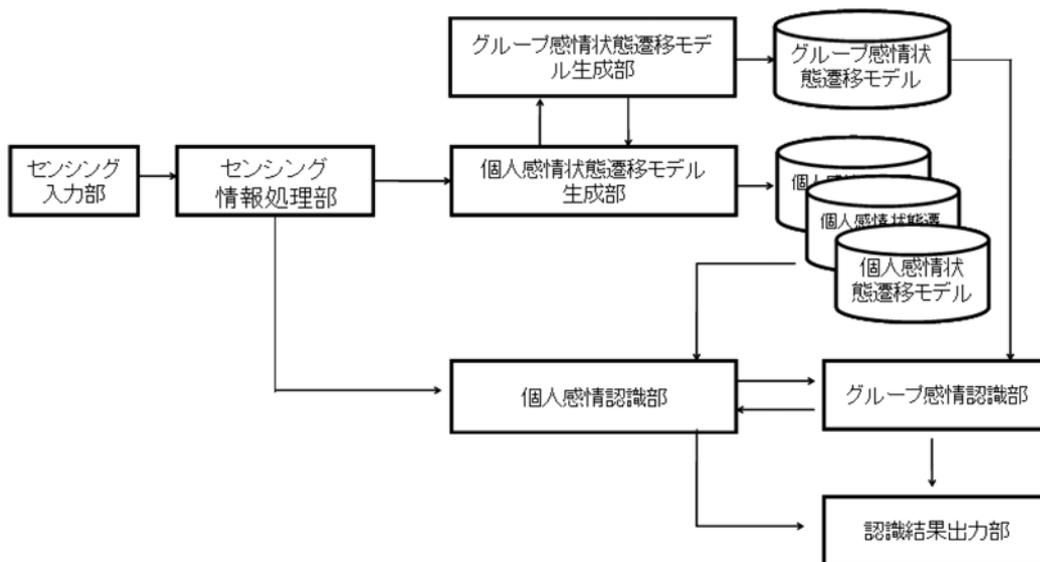


図 4.1 提案する感情認識手法の概略図

2.2 グループ感情場モデル

本節では、提案するグループ感情場モデルについて述べる。図 4.2 に 2 人(A 及び B と表記する)で構成されるグループ G のグループ感情場モデルの例を示す。図 4.2 に示すように、グループ感情場モデルはグループ構成員、およびグループ全体の感情分布とセンサによる観測量をノードとするグラフ構造として表現できる。

図 4.2 において $E_A(t)$, $E_B(t)$ は、それぞれ、 A と B の個人感情を表す確率変数であり、 $E_G(t)$ はグループ感情を表す確率変数である。時刻 t における A, B, G の感情は、それぞれの確率変数の分布で表現する。このとき、時刻 t におけるグループ感情場 $G(t)$ は、以下のように、それぞれの感情の組で表現する。グループ感情は式 (1) から推定できる。

$$G(t) = \{ E_A(t), E_B(t), G(t - 1) \} \quad (1)$$

また、モデルへの外部からの入力は、顔表情推定・音声獲得器 $F_A(t)$, $F_B(t)$ で表記してある。グループ感情場モデルにおいては、時刻 t におけるグループ感情 $E_G(t)$ は、時刻 $t-1$ におけるグループ感情 $E_G(t-1)$ および、A, B の個人感情 $E_A(t-1)$, $E_B(t-1)$ から推定する。

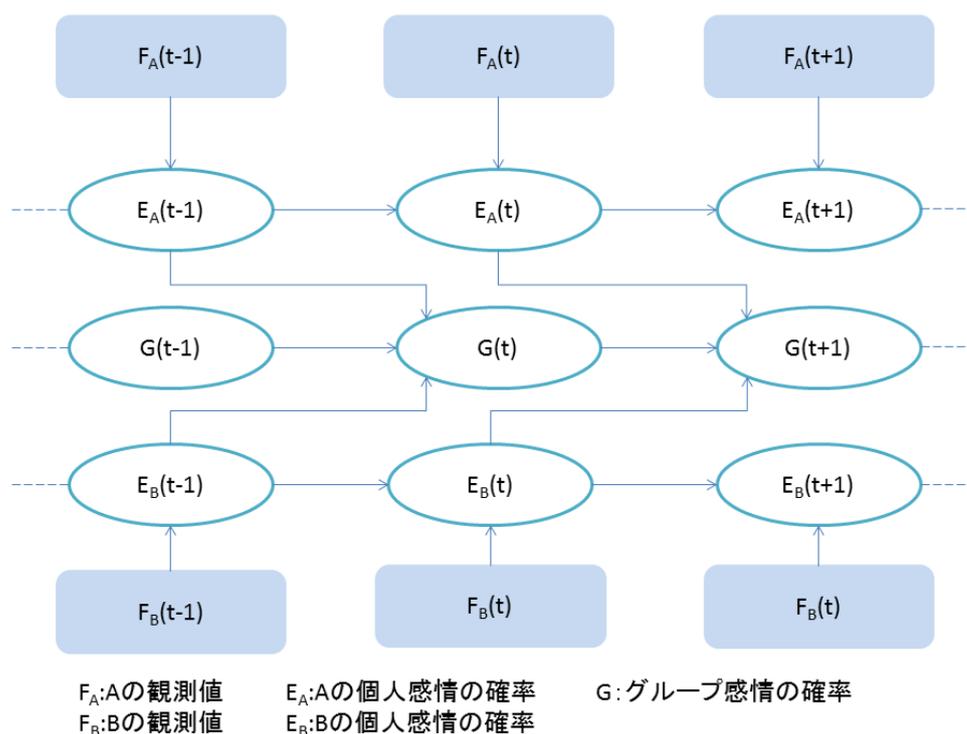


図 4.2 グループ感情場モデル

3. 個人感情の推定

本章では、グループ感情場モデルの構成要素である個人感情の推定方式を述べる。個人感情は、画像から得られる表情特徴量と、音声特徴量を用いて推定する。

3.1 表情特徴量の抽出

3.1.1 顔領域の検出

表情特徴量を抽出する前に、まず顔領域の検出が必要である。提案手法では、Haar-like 特徴量に基づいた Ada-boost 識別器を用いる。Haar-like 特徴量は、隣接した矩形領域の平均輝度差に基づいた特徴量である。図 4.3 に Viola & Jones の Haar-like 特徴量と Lienhart の Haar-like 特徴量を示す。

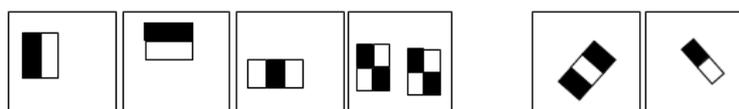


図 4.3 Viola & Jones の Haar-like 特徴量 (左) Lienhart の Haar-like 特徴量 (右)

Ada-boost 識別器は、弱い識別器を組み合わせ強い識別器を実現することができる。図 4.4 に、Ada-boost 識別器の学習アルゴリズムを示す。

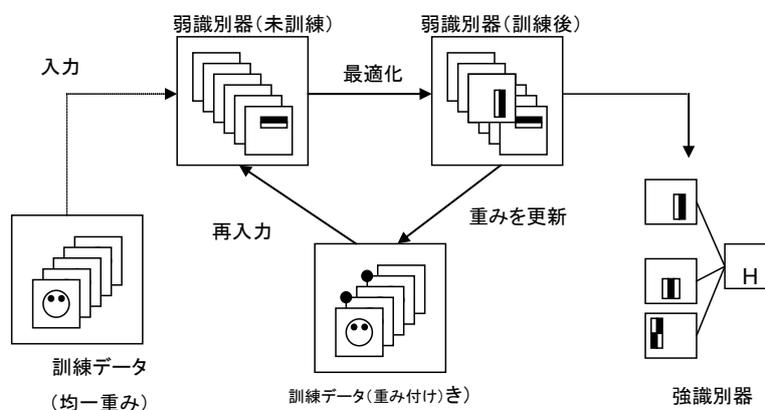


図 4.4 Ada-boost 識別器の学習アルゴリズム

3.1.2 表情特徴量の抽出

検出された顔領域から表情特徴量の抽出を行う。まず、検出された顔領域の複数から正面顔の位置と大きさ（半径）を求める。次に、ASM(Active Shape Model)[5]を用いて特徴点位置の抽出と追跡を行う。抽出する特徴点を、表 4.1 に示す。

表 4.1 表情特徴量

	表情推定のための特徴量
F1	両目の上端・下端の輝度変化
F2	口右側特徴点と目右側特徴点の距離
F3	口左側特徴点と目左側特徴点の距離
F4	口右側特徴点と口左側特徴点の距離
F5	口上側特徴点と口下側特徴点の距離

これらの特徴点は、5次元の特徴ベクトルに対応する、目、口の10点とし、FACS[6]のAction Unit(AU)の組み合わせに基づく構成とした。ただし、目に関しては検出領域が比較的小さい点、感情表出時に共起性がある点を考慮に入れ、両目上端、下端付近の輝度変化（平常時からの差）を特徴ベクトルとした。これらの特徴ベクトルを人物A、Bそれぞれについて求める。図4.2に、検出された顔領域に対するASMの顔モデルの例、および、特徴ベクトルF2～F5の様子を示す。

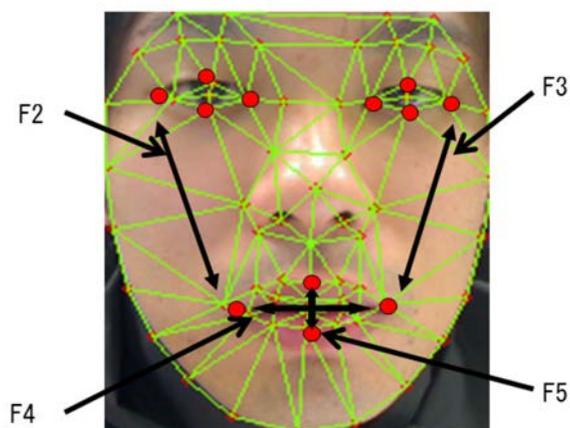


図 4.5 表情特徴量

そして、笑い、驚き、普通の3表情におけるこれらの特徴量をあらかじめ学習しておき、フレームごとに得られた特徴ベクトルから k-NN 法を用いて表情を推定する。推定結果は A と B それぞれに対して $F_A(t)$, $F_B(t)$ として確率分布の形で出力する。

3.2 音声特徴量の抽出

3.2.1 音声特徴量

音声は、韻律、音質、音韻の3つの要素により成り立っている。この中でも、韻律的特徴が人間の感情に最も影響を及ぼすことが明らかになっている[7]。そこで本研究では、発話音声から短時間パワー、ピッチを音声特徴量とした。パワーおよびピッチは、平均、最大、最小値を抽出した。このとき、短時間分析におけるフレーム長を 40ms、フレーム周期を 20ms とした。

3.2.2 音声パワーとピッチの抽出

まず、音声パワーの抽出について述べる。音声パワーは、音圧とも呼ばれ式 (2) から算出することができる。式 (2) において、 P 、 P_0 は、それぞれ音圧と基準値を表している。

$$L_p = 10 \log_{10} \frac{p}{p_0} \quad (2)$$

次に、ピッチ抽出について述べる。ピッチ抽出には、自己相関関数に基づいた手法により抽出する。自己相関関数は、任意の信号とそれ自身の相関関数である。自己相関関数は、式 (3) として表現できる。

$$r(k) = \frac{1}{N} \sum_{t=0}^{N-1-k} x_t x_{t+k} \quad (3)$$

3.3 表情特徴量と音声特徴の統合

抽出された表情特徴量と音声特徴量を統合し、個人感情を推定する。個人の感情はグループの感情を受けると考えられるため、式 (4) のように個人感情の推定結果に現在のグループ感情の推定結果を考慮し、現在の個人感情とした。また式 (4) において、係数 w は重み係数である。重み係数 w の値は、事前実験の結果により決定した。

$$E_A(t) = w \cdot F_A(t) + (1 - w) \cdot G(t) \quad (4)$$
$$(0 \leq w \leq 1)$$

4. グループ感情の推定

2章で定義したグループ感情場モデルと、3章で述べた個人感情を用いて、ベイジアンネットワークを構築し、グループ感情を推定する。図4.6に、2人から構成されるグループのグループ感情推定ベイジアンネットワークの例を示す。図4.6において、SNDは、音声特徴量の観測値であり、IMGは表情特徴量の観測値である。また、 $E_A(t)$ と $E_B(t)$ は、時刻 t における個人感情であり、 $G(t)$ は時刻 t におけるグループ感情である。

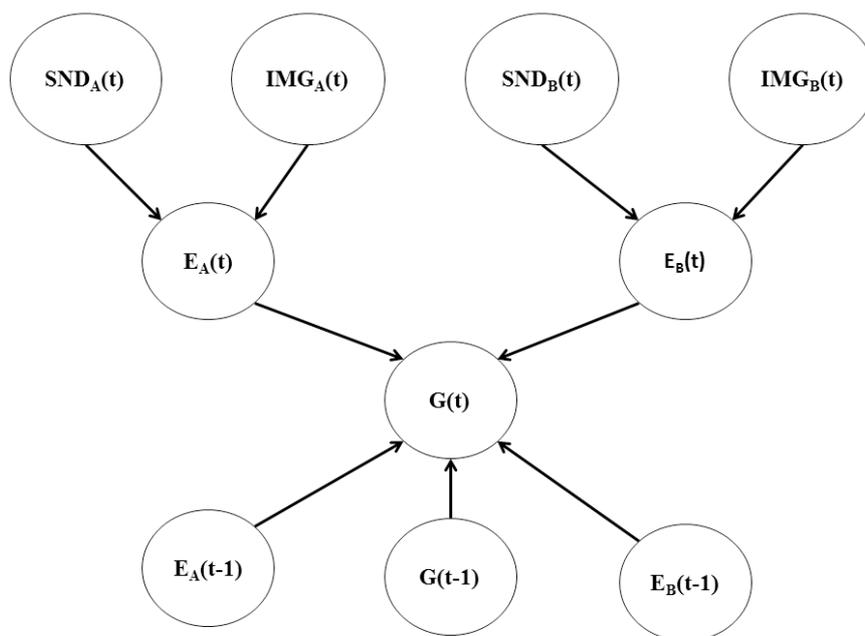


図4.6 グループ感情推定のためのベイジアンネットワーク

グループ感情のベイジアンネットワークは、グループ感情場モデルにおける時刻 $t-1$ での個人感情 $E_A(t-1)$ 、 $E_B(t-1)$ およびグループ感情 $E_G(t-1)$ を入力ノードし、時刻 t におけるグループ感情 $E_G(t)$ の確率分布を算出・推定する。

5. 実験

これまでに述べた手法を実装し、4人1組のグループを対象とした実験を行い、本稿で提案したグループ感情場モデルに基づく感情推定がどの程度機能するかについて検証した。

5.1 実験方法

4名によるグループでカードゲームの「UNO（ウノ）」を行ってもらい、ゲーム中に表出される自然な表情および会話音声を観測対象とした。4名の実験参加者には、それぞれマイクを装着し、各人に1台ずつ計4台のカメラでその様子を録画・録音した（画像サイズ=320×240画素，30fps，バストショット）。また、個人を撮影する4台のカメラの他に、全体を撮影する俯瞰カメラを1台用意し、ゲーム中の全員の状況も同時に撮影した。撮影環境および実験参加者の全体ショットを図4.7，図4.8にそれぞれ示す。また、実験機材を表4.2に示す。

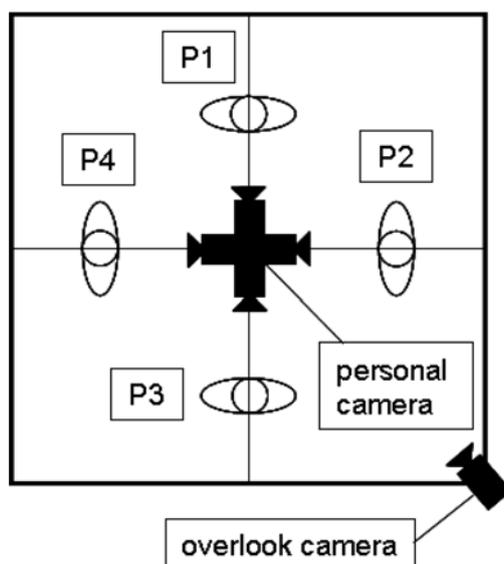


図 4.7 撮影環境図

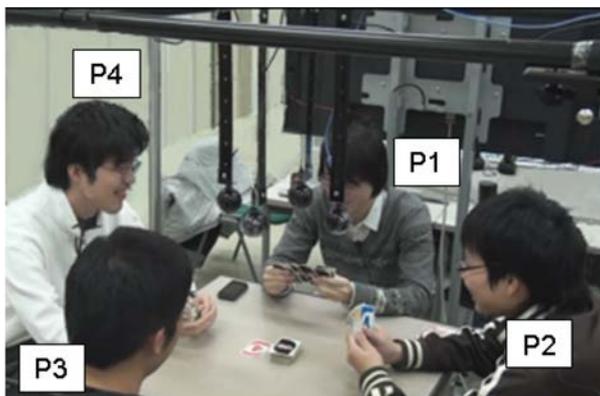


図 4.8 全体ショット

表 4.2 実験機材

OS	1,2: Microsoft Windows XP Professional
CPU	1: Intel® Xeon CPU X5450 3.00G
	2: Intel® Core™2Duo CPU T8100 2.10G
メモリ	1: 3.25GB RAM
	2: 2.00GB RAM
カメラ	Pcam: Logicool Qcam Orbit AF
	Ocam: SONY HDR-CX100
ビデオフォーマット	Pcam: AVI(30FPS,320x240px,RGB)
	Ocam: MPG(30FPS,Dolby AC3 6ch 448Kbps)
マイク	OLYMPUS ME52W
音声フォーマット	WMA(44100kHz, 16bit, 1channel)

5.2 実験結果

まず、個人感情の認識率を示す。認識率は、提案手法による個人の感情分布 $E_A(t)$, $E_B(t)$ において最大の確率を持つ感情が、正解と一致したかどうかを比較し、その一致率によって

評価した。なお、 $E_A(t)$ 、 $E_B(t)$ は映像のフレームごとに算出されるが、今回は被験者が表情を表出した区間の一致率を評価した。したがって、普通の場合は、映像全てを通じて普通の確率が最大値として出力され続ければ正解となる。実験参加者ごとの個人感情の認識率を表 4.3 に示す。

表 4.3 実験参加者ごとの認識率

実験参加者	普通	笑い	驚き	平均値
A	69.50%	74.50%	60.60%	68.20%
B	59.80%	80.10%	59.00%	66.30%
C	74.90%	78.50%	41.10%	64.83%
D	63.80%	58.30%	50.30%	57.47%
全体平均	67.00%	72.85%	52.75%	64.20%

普通、笑いの認識率は約 70%と高い精度を示す結果となった。しかし、驚きに関しては約 52%と他の感情と比べても認識率が低い結果となった。また、全体の平均認識率は 64.6%であった。

次に、グループ感情の実験結果について述べる。グループ感情の評価方法は、4名の評価者を被験者とは別に用意し、5.1節で述べた 30 パターンの映像クリップを見せてそれぞれの映像における普通、笑い、驚きの感情の分布を、それぞれの合計が 1.0 となるように評価してもらい、その平均値を正解分布とした。提案アルゴリズムによる出力との比較は、個人感情と同様に、感情の確率分布において最大の値を持つ感情が、出力分布と正解分布で一致しているかどうかによって求めた。グループ感情の認識率を表 4.4 に示す。

表 4.4 グループ感情の認識率

	普通	笑い	驚き	平均
認識率	63.80%	71.60%	45.70%	60.30%

この結果から、グループの感情推定については、普通、笑いの感情については比較的
正答率が高く、特に笑いの感情については、人間の直感に対して 70%以上の正答率で推定出
来ていることが分かる。一方、驚きの感情については、約 45%程度と正答率がかなり低い
結果となった。全体の認識率としては、60.30%であった。

6. 考察

まず、個人感情の推定結果については、普通と笑いの認識精度は比較的高い値を示して
いる。しかし、驚きの感情は他の感情と比べて認識精度が劣る結果となった。無表情の状
態からの発話は、表情特徴量が大きく変化することや、発話時の普通・驚きの表情は、発
話内容によって、見かけの変化に個人差が大きいためであると考えられる。また、実験内
容がゲーム中に発生する自然な感情の推定であるため、各実験参加者がゲームを楽しむこ
とで、普通や驚きの感情に笑いの感情が含まれた（含み笑いなど）ことが感情認識処理に
影響を与え、認識率の低下に繋がったと考えられる。

一方、グループ感情については、全体の正答率は約 59.3%と、個人感情より劣る結果とな
ったが、普通の感情が 65%程度、笑いの感情については 70%以上と、集団の感情分布の推

定問題が、人間でも誤りかねない非常に複雑な問題であることを考慮すると、良好な数値であると考えられる。驚きの正答率の低下については個人感情での考察で述べた通りである。

グループ感情分布の評価者に対し、感想を求めてみると、発話の有無が大きく影響しており、発話している被験者が表出している感情の分布を大きくしてしまうとのことであった。また、発声音量や声の高低、性別など被験者が持つ雰囲気や、グループ構成員の関係性によって影響の差異が生まれるようである。さらに、今回の実験参加者は比較的親しい関係性であったこともグループ感情分布に影響したと考えられる。

以上のことから、

- (1) グループ構成員の影響度を考慮したモデル化
- (2) グループ構成員の関係性を変えた分析・評価
- (3) 集団会話や討論など場面を変えた場合の分析

を今後行いたいと考えている。このような、様々な場面における感情表出の、より高度な特徴を探り、モデルに含めていくことは今後の重要な課題である。

加えて、コミュニケーションロボットから視覚情報として顔表情を観測できない状況も予想される。このような状況下においても、捉えることができない対話者の表情を推定することは、場の雰囲気を察したコミュニケーションを行うために重要である。この場合は、過去のグループやそのメンバーの感情分布を用いて、欠損感情を補完する等の工夫が考えられる。

7. おわりに

本章では、グループ感情をモデル化し、4人がグループを構成する映像からグループ感情を推定する手法について提案した。実験の結果、個人感情については認識率が約70%と良好な値を得た。また、グループ感情についても、笑いの感情については認識率約70%、全体でも約60%と、人間でも認識困難な問題であることを考慮すると、比較的良好な結果が得られたと思われる。今後、個人感情認識については、発話の有無による認識率低下を防ぐため、音声認識による言語的情報の組み込み等を検討する。グループ感情の認識についても、より人間の直感に即した認識ができるよう、検討を進めたい。

参考文献

- [1] Albert Mehrabian, *Silent messages: Implicit communication of emotions and attitudes* (2nd ed.), Wadsworth, CA (1981), (非言語コミュニケーション(西田司他共訳), 聖文社(1986))
- [2] 松本祥平, 山口健, 駒谷和範, 尾形哲也, 奥乃博, “ロボットでの利用を目的とした顔画像情報と音声情報の統合による感情認識,” 日本ロボット学会学術講演会予稿集, Vol.22nd, Page3D14, 2004.09.15.
- [3] 伊藤彰則, 王欽悦, 鈴木基之, 牧野正三, “音声処理と顔画像処理を統合した対話映像からの笑いの認識,” 情報処理学会研究報告, 音声言語情報処理(SLP), 2005(50), pp.41-46(2005)
- [4] 沖電気工業株式会社, “集団感情認識支援システム,” 特願 2007-116066 (2007.04.25), 特開 2008-272019 (2008).

[5] D.H. Cooper and T.F. Cootes and C.J. Taylor and J. Graham, "Active shape models - their training and application", *Computer Vision and Image Understanding*, 38-59, 1995.

[6] P. Ekman, W. V. Friesen "The Facial Action Coding Systems", Consulting Psychologists Press, 1978.

[7] 重永実, "感情の判別分析からみた感情音声の特性", *電子情報通信学会論文誌*, Vol.J83-A, No.6, pp.726.735(2000)

5 章 結言

本論文では、「人間・エージェント間の協調的コミュニケーション制御に関する研究」の成果について述べた。本章では、これまでの研究成果を総括し、本論文の結論を述べる。

近年、ロボティクスの分野では、人とのコミュニケーションや介護を目的としたコミュニケーションロボットや介護ロボットなどが開発され、ロボットが一般の人々にも身近な存在となりつつある。コミュニケーションロボットや介護ロボットのように、ロボットが人々の生活により密接にかかわる場合、人と持続的かつ自然にコミュニケーションをとる必要がある。そのためには、人とロボットが協調的に対話するための対話の流れを解明し、より自然でストレスの少ない親和性のあるコミュニケーション制御が必要である。

2章では、交替潜時と韻律情報に基づく非線形振動子による連続的な引き込み制御とインタラクションルールによる離散的な引き込み制御により、人とエージェントが協調的にコミュニケーションを行うための会話制御方式を提案した。また、情報収集を目的としたエージェントへ実装し、会話同調制御の評価実験を行った。評価実験の結果からは、従来手法に較べて、エージェントがどの程度親しみやすいかの評価尺度である親身度に対して有意差を示すことができ、同時にエージェントの会話の自然さ（ストレスが少ない）に対しても有意差が認められた。これらの結果から、

- (1) 提案方式は、対話者のコミュニケーション状態量を模倣する制御方式と較べて、より協調的でストレスのない引き込み制御ができる
- (2) 会話エージェントは、人のコミュニケーション状態量の変化に適応的に振る舞うこと

ができる

(3) 交替潜時だけでなく、韻律情報（ピッチ）は、協調的でストレスの少ない会話の実現

に寄与し、韻律的なリズムの同調は会話全体の調整に大きな役割を果たしている

ことが明らかとなり、これらを実現するシステム仕様について明確化した。

また、実験を通して、より親しみが持てストレスの少ない会話のもとで持続的な情報収集が行える会話システムの実現可能性を示すことができた。

3章では、ロボットが複数の人間から構成されるグループとのコミュニケーションを想定し、これまで十分に検討されていなかった初対面時や社会的スキルに違いに応じた協調的コミュニケーション制御についての研究を行い、人間が初対面仲介をする方法をモデル化した。また、初対面仲介モデルをエージェントへ実装して、その有効性を検証した。実験では、このエージェントと2人の初対面者がコミュニケーションをする実験と、エージェントを用いず2人の初対面者がコミュニケーションをする2種類の実験を行った。実験のアンケート結果から、学生の実験参加者を用いた実験室での限定された実験ではあるが、提案モデルの有効性を検証することが出来た。また、社会的スキルの観点から実験を行った。実験では、実験参加者として高齢者を対象とし、社会的スキルの高低によりグループを分けた。実験結果からは、社会的スキルの高い実験参加者に比べ社会的スキルの低い実験参加者のほうがロボットの仲介行動が有効に働き、社会的スキルが低い人にとって、5つの仲介行動の身体的振る舞い（視線制御や傾き）が有効に働くことが明らかにした。

次に4章では、これまでに十分検討されていなかったグループの感情を推定する手法についての研究を行った。

ロボットが人間と円滑にコミュニケーションを行うためには、ロボットが人間の状態に応じて適応的に振る舞うことが必要不可欠である。人と人がコミュニケーションを行うときには、言語情報よりも、視線や顔表情・ジェスチャ・韻律情報などの非言語情報がより重要な役割を担うことが明らかにされている。特に、音声や顔表情は豊富な情報を含んでおり、コミュニケーションの状態理解において極めて重要な要素である。したがって、人間の感情を推定する技術は、ロボットが人間との円滑なコミュニケーションを支える技術として、重要であると言える。これまでに、個人の感情を推定する手法は多く研究されている。しかし、グループの感情を推定する手法については十分には検討されていないのが現状である。

そこで、複数の人間から構成されるグループの感情をモデル化し、グループコミュニケーションの映像からグループ感情を推定する手法について提案した。提案する手法は、個人の顔表情と韻律情報から個人感情を推定し、個人感情と観測されたパラメータを基にベイジアンネットワークにより統合し、グループの感情を推定する。実験の結果、個人感情については認識率が約 70%と良好な値を得た。一方、グループ感情についても、笑いの感情については認識率約 70%、全体では約 60%であった。グループがどのような感情であるかどうかを推定することは、人間でも認識困難な問題であり、実験からは十分な結果が得られたと思われる。今後は、個人感情の認識については、発話の有無による認識率低下を防ぐため、音声認識による言語的情報の組み込み等を検討する。グループ感情の認識についても、より人間の直感に即した認識ができるよう、検討を進める必要がある。

以上の研究結果から、提案した人間とロボットが円滑にコミュニケーションするための制御方式について、十分な結果を得ることができたと考える。これにより、人間とロボットが円滑にコミュニケーションをすることが期待される。

今後の課題としては、提案した人間・ロボット間のコミュニケーション制御方式を、人とコミュニケーションを取ることを目的としたロボットやエージェントへ実装し、本制御方式の実用性について検証することが考えられる。具体的な応用先としては、遠隔（医師と患者間）で問診可能な「遠隔問診エージェント」や介護を目的とした介護ロボットなどが挙げられる。今後は、これらの実現に向けて取り組みたい。

外部発表リスト

< 学術論文 (査読有) >

- 速水達也, 佐野睦夫, 向井謙太郎, 神田智子, 宮脇健三郎, 笹間亮平, 山口智治,
山田敬嗣

“交替潜時と韻律情報に基づく会話同調制御方式と情報収集を目的とした会話エージェントへの実装”, 情報処理学会論文誌, Vol.54, No.8, pp.2109 - 2118 (2013)

< 国際会議 (査読有) >

- Tatsuya Hayamizu, Hiroaki Mori, Sano Mutsuo, Satoshi Nishiguchi, Nobuyuki Yamashita

“Group Emotion Estimation using Bayesian Network based on Face Expression and Prosodic Information”, 2012 IEEE International Conference on Control System, Computing and Engineering pp.181-186(2012)

- Tatsuya Hayamizu, Kenzaburo Miyawaki, Mutsuo Sano, Kentarou Mukai

“An Interactive Agent Supporting First Meetings Based on Adaptive Entrainment Control”,
The Six International Conference on Advances in Computer Human Interactions, pp.178-183 (2013)

< その他 >

2 件

謝辞

本研究の遂行と本論文を執筆するにあたり，終始適切なご指導とご鞭撻を頂きました大阪工業大学 情報科学部情報メディア学科 佐野睦夫教授に対し，心から感謝の意を表します。

論文の作成にあたり，懇切丁寧なご指導とご助言を頂きました，大阪工業大学 情報科学部情報ネットワーク学科 西口敏司准教授，並びに 大阪工業大学 情報科学部情報メディア学科 宮脇健三郎講師に深く感謝いたします。

大学院在学中に，親切なご指導，ご助言を頂きました大阪工業大学情報科学部の諸先生方に深く感謝いたします。

実験にあたり，快く実験参加者となって頂いた大阪工業大学 コンピュータインタラクション研究室の卒研究生に深く感謝いたします。

ここに，ご指導とご支援を賜りました関係者各位に厚く御礼を申し上げます。

付録 A

氏名： _____ 年 月 日

■以下の質問に答えてください。

ロボットに対して、あなたはどのような印象を持ちましたか？

下記の形容詞対がどれくらいあてはまるかを、1～7の中から選んで○をつけてください。

	非常に	かなり	やや	いど ちら ない とも	やや	かなり	非常に
1 有能	1	2	3	4	5	6	7 無能
2 助けになる	1	2	3	4	5	6	7 当てにならない
3 親切的な	1	2	3	4	5	6	7 不親切的な
4 親しくしたい	1	2	3	4	5	6	7 距離を置きたい
5 役に立つ	1	2	3	4	5	6	7 役に立たない