

ユーザの注視行動に基づく会話参加態度の推定 —会話エージェントにおける適応的会話制御に向けて

石井 亮^{†1,*1} 中野 有紀子^{†2}

対面会話において、聞き手が関心を持って会話に参加していることを、話し手は聞き手の動作や視線から察知し、積極的に参加していない様子であれば話題を変える等、会話の内容や方略を調整している。このような適応的な会話制御が可能な会話エージェントを目指し、本研究では、ユーザの注視行動から対話への参加態度を推定する機構を提案・実装する。まず、Wizard-of-Oz法により、ユーザの視線行動の計測データ、会話への関心低下に関するユーザの内観と他者の観察、発話情報を収集する。次に、ユーザの注視行動を分析し、理想的な会話参加態度から逸脱している視線遷移パターンを同定する。さらに、エージェントとの会話中にリアルタイムに取得される視線データから、個人差を考慮しながら逸脱度の高い状態を検出することができる個人適応型会話参加態度推定アルゴリズムを提案し、これを、視線計測装置を統合した会話参加態度推定機構として実装する。最後に、評価実験から、個人適用型アルゴリズムの有用性、ならびにユーザの会話参加態度に応じてエージェントが振舞いを変化させることのインタラクションにおける効果を示す。

Estimating the Degree of Conversational Engagement Based on User's Gaze Behaviors in Human-agent Interactions: Towards Adaptive Dialogue Management in Conversational Agents

RYO ISHII^{†1,*1} and YUKIKO NAKANO^{†2}

In face-to-face conversations, speakers are continuously checking whether the listener is engaged in the conversation. When the listener is not fully engaged in the conversation, the speaker changes the conversational contents or strategies. Aiming at building a conversational agent that can control conversations in such an adaptive way, this study proposes a method for predicting whether the user is engaged in the conversation or not based on the user's gaze transition 3-Gram patterns. First, we conducted a Wizard-of-Oz experiment to collect the user's gaze behaviors as well as the user's subjective reports and an

observer's judgment concerning the user's interest in the conversation. Then, by analyzing the user's gaze behaviors, disengaging gaze patterns will be identified. Based on these results, we propose an engagement estimation method that can take account of individual differences in gaze patterns. The algorithm is implemented as a real time engagement judgment mechanism, and the results of our evaluation experiment showed that our method can predict the user's conversational engagement quite well, and the users felt that the agent's conversational functions were improved.

1. はじめに

対面場面の会話では、話し手は聞き手が会話に注意を向け、適切に参加しているか否かを確認しながら発話を行っている。一方、聞き手は、言語・非言語行動を通して会話への参加意思を相手に伝えている。たとえば、視線や頷きといった非言語行動は、会話に注意を向けていることを話し手に伝える有効なシグナルである。

このような、会話への参加意思確認の過程は会話を成立させるうえで、基本的、かつ不可欠なプロセスであり、人間同士のコミュニケーションだけでなく、人と人工物とのコミュニケーションにおいても重要な課題である。なぜなら、ユーザとのコミュニケーションチャンネルが適切な状態になれば、システムからの情報はユーザには伝わらない可能性が高いからである。システムがユーザの会話参加態度をモニタし、それが望ましい状態から逸脱していることを検出できれば、会話中にユーザの注意を喚起したり、話題を変えたりする等、システムによる適応的な振舞いが可能になる。

しかし、会話参加態度が視線や頷き等の非言語行動に現れるという一般的な傾向については広く認知されているものの、望ましい会話参加態度、あるいは非積極的な態度を、計算機で判断するための精緻なモデルや手法については十分に研究されていない。

そこで本研究では、会話エージェント¹⁾と呼ばれる、音声言語と表情やジェスチャを用い

†1 東京農工大学大学院工学府情報工学専攻
Graduate School, Tokyo University of Agriculture and Technology

†2 成蹊大学理工学部情報科学科
Department of Computer and Information Science, Faculty of Science and Technology, Seikei University

*1 現在、日本電信電話株式会社 NTT サイバースペース研究所
Presently with Recently with NTT Cyber Space Laboratories, Nippon Telegraph and Telephone Corporation

て人とコミュニケーションできるアニメーションキャラクタやロボットにおいて、ユーザとのより円滑なコミュニケーションを目指し、ユーザの会話参加態度推定方法の提案、実装、評価を行う。特に、会話エージェントとインタラクションする際の、ユーザの注視行動に着目し、具体的には、以下の2つの課題に取り組む。

- (1) Wizard-of-Oz システムによる実験から得られた視線データと会話参加態度に関する主観的評価データとの相関関係から、理想的な会話参加態度から逸脱していると見なされる視線遷移パターンを同定する。
- (2) エージェントとの会話中にリアルタイムに取得される視線データから、個人差を考慮しながら逸脱度の高い状態を発見することができる、個人適応型会話参加態度推定アルゴリズムを提案する。

以下、2章では、関連研究をあげながら、本研究の位置づけを議論する。3章では、対話収録実験について述べ、4章では、データ分析の結果を報告する。さらに、5章で、会話参加態度推定アルゴリズムを提案し、6章では、その実装について述べる。7章では提案方式の有用性をユーザテストにより評価した結果を報告し、最後にまとめの議論を行う。

2. 関連研究

対面場面の会話に関するコミュニケーション科学や心理学の研究において、視線の機能について数多くの研究がある。Kendon²⁾ は、エスノメソドロジの視点から視線の機能について詳細な観察を行っている。また、心理学の分野では、視線は、聞き手が会話に注意を向け、会話が円滑に進んでいることを伝える話し手への肯定的なフィードバックであること^{3),4)}、円滑なターン交代に寄与していること等⁵⁾ が明らかにされている。一方、何かの対象物についての会話を行う際には、聞き手が、話し手ではなく、対象物に視線を向けた状態の共同注視 (joint attention) も会話に積極的に参加していることを示す行動であるといわれている^{6),7)}。共有された対象物は、会話において、暗黙の共有知識として利用され、相互理解構築の重要な基盤となるからである⁸⁾。

このような人間の注視行動に関する知見に基づき、会話エージェントにおけるコミュニケーション機能を実装した研究として、Nakano らは、ヘッドトラッカから推定されたユーザの視線方向から、ユーザがエージェントの発話を理解しているか否かを判定し、それを対話制御に利用する方法が提案されている⁹⁾。また、Gratch らは、エージェントが聞き手として注視行動によるフィードバックを行うことにより、ユーザとエージェントとの信頼関係 (rapport) を向上させる効果を持つと報告している¹⁰⁾。さらに、人对コミュニケーションロ

ボットの研究では、注視行動は、ユーザの会話参加態度 (engagement) の問題として、議論されている¹¹⁾。文献 11) ではユーザがロボットとの会話に積極的に取り組んでいる、つまり “engage” しているか否かを、ヘッドトラッカにより認識されたユーザの頭部姿勢情報を用いて判断している。しかし、これらの会話エージェント研究では、ヘッドトラッカを用いてユーザの注視行動を計測しているため、厳密なユーザの視線方向を測定することができず、頻繁に変化する視線に関する精緻なモデルはまだ提案されていない。

一方、情報提示システムの研究では、アイトラッカにより計測されたより詳細な視線データを用い、視線移行パターンからユーザの迷いや¹²⁾、興味対象を推定する方法も検討されている^{13),14)}。

以上、先行研究では、コミュニケーションにおける視線の様々な働きが明らかにされてきており、また、アイトラッカを組み込んだシステムが、いくつか構築され始めている。このような状況のなか、本研究は、より高度なマルチモーダル対話システムの実現を目指し、ユーザの詳細な注視行動から会話参加態度を推定する方法を確立し、それを組み込んだ会話エージェントを実装することを目的とする。

3. ユーザ対エージェントの会話収録実験

ユーザの会話参加態度を分析・推定するうえで有用な言語・非言語行動を収集するために、Wizard-of-Oz システムを用いて、人と会話エージェントとの対話を収録した。

3.1 実験環境

実験環境を図1に示す。図中のプロジェクタから、120インチリア型のスクリーンに、図2に示すアニメーションキャラクタが投影される。ユーザ役の被験者（以下、ユーザと呼ぶ）には、これが携帯電話の販売員エージェントであると伝えられた。ユーザは、パーティションで仕切られた個室スペースに入り、販売員エージェントと対面するように、スクリーンからおよそ1.5m離れた位置に立つ。さらに、このパーティションの外側からハーフミラー越しに、観察者役の被験者（以下、観察者と呼ぶ）がユーザの様子を観察する。

3.1.1 エージェントの動作作成

会話エージェントの言語・非言語行動は、エージェント動作自動生成システム CAST¹⁵⁾を用いて作成した。CASTは音声合成器と連携して、テキストから合成音声を生成し、ジェスチャ、リップシンクのタイミング計算を行い、エージェント動作のアニメーションスクリプトを生成する。会話エージェント描画には、Haptек アニメーションシステム¹⁶⁾を、また、音声合成エンジンには、日立のHitVoiceを用い¹⁷⁾、Haptекにおけるアニメーション

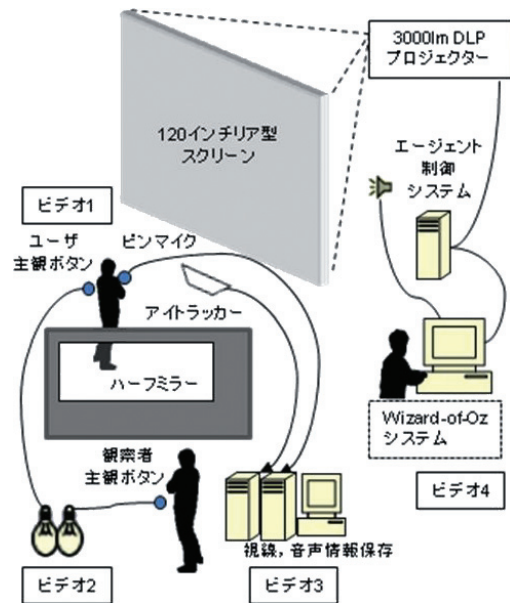


図1 対話収録実験のシステム構成
Fig.1 Experimental setting for data collection.

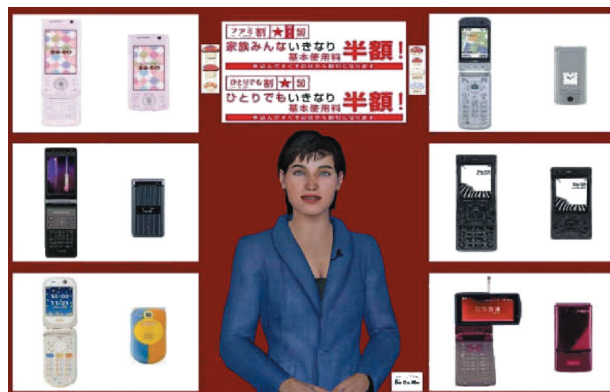


図2 スクリーン投影画面
Fig.2 Conversational agent.

スクリプトの実行と音声再生を同期して行う VB アプリケーションを Wizard-of-Oz システムとして開発した。これにより、実験者がシナリオに沿ってエージェントの発話を GUI から選択すると、音声発話とそれに同期したアニメーションがスクリーンに投影される。なお、エージェントの発話、動作画面はビデオカメラ (Sony HDR-HC1) で収録した。

3.1.2 ユーザ情報の取得

ユーザの行動データとして、以下を収集した。

- ユーザのビデオ映像：対話中のユーザの様子を収録するために、ビデオカメラ (Sony HDR-HC1) で上半身部分を録画した。
- 会話関心度の主観評価：ボタン押下式の装置をユーザと観察者に持たせ、ユーザには、エージェントの会話内容に関心がなくなったり、話題を変えたいと思ったりしたときに、押下式のボタン (以下、ユーザ関心低下ボタンと呼ぶ) を押し続けるように指示した。一方、観察者には、ユーザが会話への関心を失っているように見受けられたら、ボタン (以下、観察者関心判定ボタンと呼ぶ) を押し続けるよう指示した。その際、観察者は比較的近い距離からハーフミラー越しにユーザを観察していたので、会話音声は明確に聞こえ、ユーザの表情や視線も十分に見てとることができた。ボタンが押されると、別室に置かれたランプがそれぞれ点灯するようになっており、これらのランプの点灯/消灯状態をビデオカメラで収録した。
- ユーザの発話音声：ユーザの発話は、ピンマイク (Sony ECM-66B) ならびに、オーディオインタフェース (EDIROL UA-1000) を通して PC のハードディスクに記録された。
- ユーザの注視行動：ユーザの視線計測には、非接触型視線計測装置 (Tobii x50) を使用し、スクリーンに投影されたシステム画面上のユーザの注視位置を 50 fps で取得した。Tobii x50 は、非接触であるとともに、頭部の移動可能範囲が幅 30 × 高さ 15 × 奥行 20 cm と、比較的ユーザの自由度が確保され、ユーザへの負担が軽減されたシステムである。図 3 に、取得した注視データのプロット例を示す。注視時間長に応じて、注視位置の円形プロットが大きくなる。また、この視線データの時間を他のデータと同期させるため、画面出力されたシステム時間をビデオカメラに収めた。

3.2 実験手続き

ユーザとして、20 代の男性 9 人、女性 1 人の計 10 人 (情報系専攻学生 7 人、他の専攻 3 人)、また、ユーザが会話に関心を持っているか否かを評定する観察者として、各セッション 1 人、計男性 7 人が実験に参加した (複数回参加者を含む)。ユーザには、研修中の新人



図 3 ユーザの注視位置のプロット
Fig. 3 User's gaze plots.



図 4 携帯電話に向きながら説明中のエージェント
Fig. 4 Agent explaining and looking at a cell-phone.

販売員エージェントの接客を受ける客として、販売員の説明を最後までよく聞き、話題を変えたい場合には、エージェントにその旨を音声で伝えてもよいと指示した。これにより、ユーザが会話に関心を失っても、会話を続けざるをえない状況を作った。

また、ユーザのタスクとして、店頭の 6 種類の携帯電話の中から女子中高生に最も人気の

ある携帯、もしくはビジネスマンに最も人気のある携帯を言い当てる課題を課し、正解すれば千円の報酬があると伝え、実験への動機付けを行った。

ユーザへの提示刺激は 6 種類の携帯電話それぞれについて、説明をすべて聞いた場合、約 109 発話、16 分の説明である。また、視覚的刺激の新規性による影響を回避するため、エージェントの動きは全説明を通して単調で、規則的なものにした。たとえば、ジェスチャをとまなう発話としては、各携帯電話の説明開始時（発話「
にございますが、××でおなじみの（機種名）です。」と同時）に携帯電話をポインティングし、説明終了時にポインティングを終了するという定型を用いた。また、エージェントの視線行動として、各携帯電話の説明開始時に携帯を注視し、以後 10 発話ごとに、約 3 秒間ユーザに視線を向けるという規則的な行動を実装した。図 4 にエージェントが携帯電話を説明している様子を示す。

4. パターンの関連性

4.1 会話参加態度

本研究では、特に注視行動に着目し、ユーザが適切なタイミングで、エージェントからのフィードバック要求に応じて相互注視をしたり、エージェントの視線方向に従って説明対象に共同注視を行うことを「積極的で望ましい会話参加態度」と定義する。逆に、エージェントとの会話に関心・興味がなくなることにより、これらの行為が行われない状態を「望ましい会話参加態度からの逸脱」と見なす。たとえば頻繁によそ見をすることは望ましくない会話参加態度である。

以下の分析では、この考えに基づき、実測された注視行動から会話参加態度を推定することを目的とし、会話関心度の主観的評価と注視行動パターンの関係について分析を行う。

4.2 分析データ

3 章に示した実験環境、および実験手続きにより、平均約 16 分の会話を 10 会話収録し、その中から以下のデータを分析対象とした。

- ユーザならびにエージェントの発話テキスト（全データの発話総数はエージェントが 951 回、ユーザが 61 回）
- エージェントのジェスチャならびに視線行動
- ボタン押下式装置を用いた、ユーザ自身と観察者による、会話関心度に関する判定
- ユーザの視線データ。ただし、イトラッカによって取得されたデータのうち、半径 20 pixel 以内の円内に 20 ms 以上視線が停留した区間のみを分析対象とした。これらすべての情報をアノテーションツール anvil¹⁸⁾ を用いて統合し、言語・非言語情



図 5 作成した anvil ファイル
Fig. 5 Generated anvil file.

報の共起関係を視覚的にとらえられる 30 fps のビデオデータを作成した。図 5 にその一例を示す。たとえば、会話開始から 3 分 7 秒の時点では、エージェントはユーザを見ており (User), ユーザはエージェントの顔を見る (Agent_Head) という相互注視が起こっているが、3 分 8.5 秒の時点で、ユーザの注視対象が説明対象の携帯電話 (携帯 D) に移行している。

4.3 分析方法

収集された注視行動データは、最短視線停留時間を 20 ms に設定したため、非常に短時間の注視行動も計測されていたこと、また、まばたき等により、頻繁に注視行動が分断されていたことから、継続長の短いものが多かった。そのため、個々の注視行動を分析単位とするよりも、連続した複数の注視行動からなる注視行動のパターンを分析単位とするほうが適切であると考え、本研究では、ユーザの注視行動の遷移 3-Gram を作成し、エージェントが説明対象を注視しながら発話しているときの、注視行動遷移 3-Gram と関心低下/関心判定ボタン押下との関連性を分析した。

まず、視線データから、ユーザが画面上の何をしていたのかを算出し、注視対象を、T: エージェントの説明対象 (共同注視状態), AH: エージェントの頭部, AB: エージェントの頭部以外の上半身, F: それ以外の対象物 (F1≠F2≠F3) に分類した。このとき、200 ms

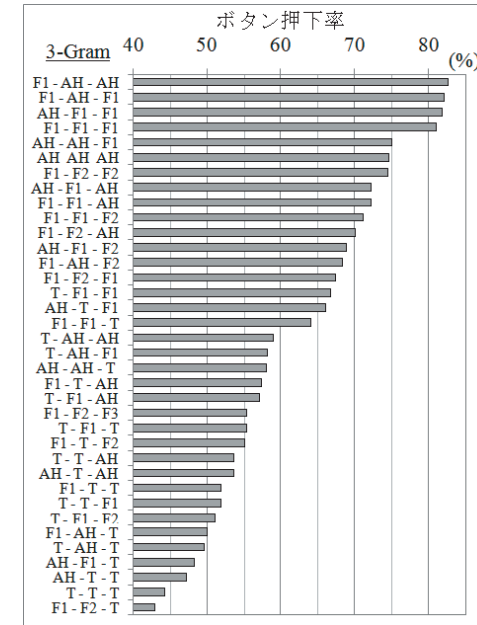


図 6 3-Gram と関心低下/判定ボタン押下率の関係
Fig. 6 Eye-gaze 3-Grams and probabilities of button pressing.

以内に、同じ対象が連続して再び注視されたときは、以前の注視行動が継続していると見なし、1 つの注視行動として扱い、注視行動を統合した。また、時間的に連続した 2 つの注視行動において、注視対象が変化するとき、もしくは注視行動の間に 200 ms 以上の間隔があった場合には、別の注視行動として扱い、連続した 3 つの注視行動を構成要素に持つ 3-Gram を作成した。たとえば、注視対象が T, (100 ms), AH, (50 ms), AH, (150 ms), F1, つまり、T と 1 つめの AH の間に 100 ms の間隔、1 つ目の AH と 2 つ目の AH の間には、50 ms、2 つ目の AH と F1 との間には、150 ms の間隔がある場合、50 ms 間隔で起こった 2 つの “AH” は統合され、“T-AH-F1” の 3-Gram が生成される。計測不良等により視線データが 1 秒以上存在しない場合は、未完の 3-Gram として破棄した。

また、視線行動の変化がユーザ関心低下ボタン/観察者関心判定ボタンの押下よりも先に表出するという仮定の下、3-Gram 開始から終了後 5 秒後までの間にボタンが押下された割合を、各 3-Gram ごとに計算した。なお、ユーザと観察者のボタン押下の一致率を調べた

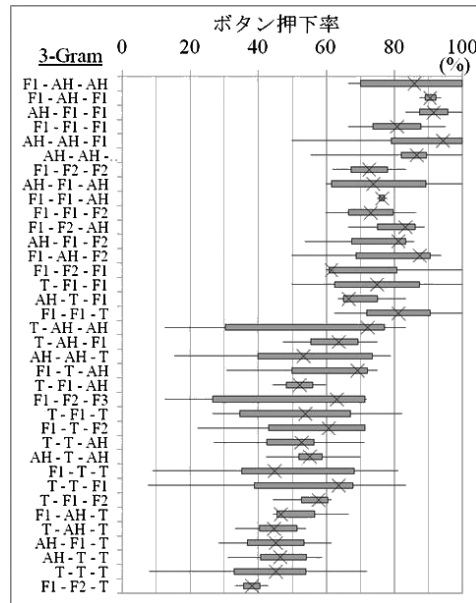


図 7 3-Gram と関心低下/関心判定ボタン押下率の個人分布
Fig. 7 Distribution of button pressing ratio with respect to eye gaze 3-Grams.

ところ、ユーザが関心低下ボタン押していたときに、観察者による関心判定ボタンも押されていた割合が 39.0%，観察者による関心判定ボタンが押されていたときに、ユーザの関心低下ボタンも押されていた割合は、54.4%と、どちらを基準にとっても一致率は低かったため、本分析ではどちらかのボタンが押下されている場合を、ボタン押下として扱った。

4.3.1 注視行動 3-Gram と会話関心度主観評価との関連性

図 6 に各 3-Gram のボタン押下率の平均値を示す。3-Gram によって、ボタン押下率に大きな差異が見られる。この結果は、ボタン押下率の高い 3-Gram は理想の会話参加態度からの逸脱行為であり、ボタン押下率の低い 3-Gram は望ましい参加態度に貢献する行為であることを示唆する。

さらに、3-Gram の押下率に個人差があるか否かを調べるために、各 3-Gram におけるボタン押下率の分布を箱ひげ図（ジョン・テューキーの方式）で表した（図 7）。図中の×の点が中央値、箱の左端が第 1 四分位点、右端が第 3 四分位点、棒の両端はそれぞれ最小値

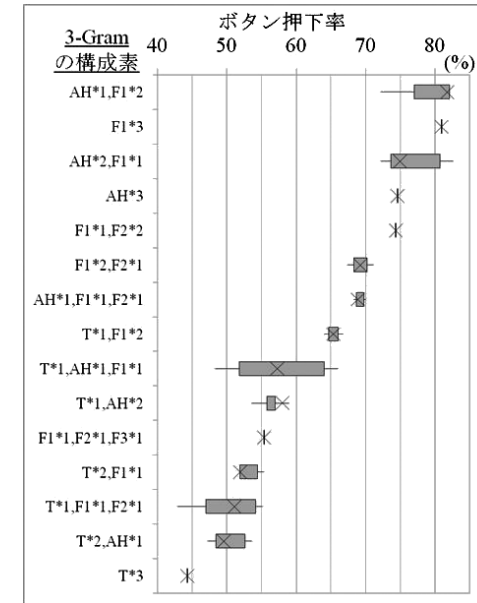


図 8 3-Gram の構成素の組と関心低下/関心判定ボタン押下率
Fig. 8 Distribution of button pressing ratio with respect to combinations of 3-Gram constituents.

と最大値を示す。押下率上位の 3-Gram と、下位の 3-Gram の間で、箱で示された範囲の重なりが大きくないことが分かる。よって、関心度が低下しているときの視線パターンと低下していないときの視線パターンは、比較的個人によらず共通していることが分かった。それゆえ、多くの人に共通する関心の有無を示す 3-Gram が存在するといえる。

4.3.2 注視行動 3-Gram の構成素と会話関心度主観評価との関連性

さらに、図 8 のように、図 6 に示した 3-Gram の中で、同じ構成素を持つ 3-Gram ごとに、押下率の分布を箱ひげ図で示した。たとえば、“F1-AH-F1”、“AH-F1-F1”、“F1-F1-AH”は、いずれも AH の注視行動 1 つと F1 の注視行動 2 つから構成されており（図中 “AH*1, F1*2” と記述）、ボタン押下率はそれぞれ、82.1, 81.8, 72.2 である。よって、“AH*1, F1*2”を構成素に持つ 3-Gram は、ボタン押下率が 72.2~82.1 の範囲に分布することになる。図に示すように、共通の構成素からなる 3-Gram の、ボタン押下率は狭い範囲にしか分散しておらず、このことから、3-Gram を構成する注視対象の組合せパターンをとってみても、

ユーザの会話参加態度の間に明確な相関関係があり、このことから、注視行動のパターンから会話参加態度を推定できることが示唆された。

たとえば、ボタン押下率上位の 3-Gram の要素に T が含まれておらず、ボタン押下率下位には多く含まれていることから、説明対象に対して共同注視を行わないのは会話参加態度として望ましくないと考えられる。一方、ボタン押下率の低かった 3-Gram の中には、“T*1, F1*1, F2*1” や “T*2, F1*1” のように説明対象と合わせて、説明対象外の対象物を見るものが含まれていることが分かった。これは一見、前の例と矛盾しているようにも思われるが、説明対象のみを注視するのではなく、説明対象と他のものを比較対照しながら説明を聞くことは、逸脱した会話参加態度ではなく、むしろ積極的な会話参加態度であると考えられる。

5. 会話参加態度の推定

5.1 望ましい参加態度からの逸脱の検出方法

4章の分析結果に基づき、注視対象遷移パターンから、理想の会話参加態度からの逸脱を検出する方法を提案する。まず、図 6 に示すボタン押下率を、「逸脱度」として各 3-Gram に割り当て、これを会話参加態度の指標とした。たとえば、ボタン押下率 75% の 3-Gram (AH-AH-F1) の逸脱度は 75 となる。

次に、図 9 のように、横軸に時間、縦軸に逸脱度をとり、各会話におけるユーザの 3-Gram をプロットした。また、ユーザ関心低下ボタン押下（斜線）、観察者関心判定ボタン押下（網掛け）をプロットし、ボタン押下（どちらかのボタンが押下）の有無と 3-Gram の分布との関係を調べた。なお、図 9 はユーザ 1 人分のデータである。

その結果、ボタンが押下されているときには、逸脱度の高い 3-Gram が頻出するが（図 9 の色つき部分）、ボタンが押下されていないときには、逸脱度の低い 3-Gram のみが観測された（図 9 の色なし部分）。このことから、逸脱度において閾値を定めることより、ボタンの押下、つまり、望ましい会話参加態度からの逸脱状態を判別できると考えた。

しかし、ユーザごとに図 9 のようなプロット図を描いてみると、逸脱度の分布幅には個人差があり、閾値を個人ごとに算出する必要があることが分かった。そこで、個人ごとに閾値を算出するために、まず、会話開始から最初の 120 秒間に現れた 3-Gram の逸脱度を、重心法を用いて 4 つにクラスタリングした。

以下に、重心法を用いたクラスタリング手法の手順を示す。

① 会話開始直後 120 秒間に表出した n 種類の 3-Gram について、逸脱度を重心座標として、

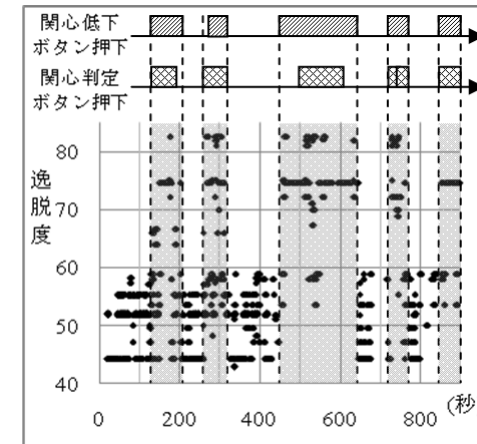


図 9 3-Gram と関心低下/判定ボタン押下の関係

Fig. 9 Plots of eye gaze behaviors and button pressing actions.

それぞれ個別のクラスタとする。さらに、同じクラスタに含まれる各 3-Gram の継続時間長の合計をそのクラスタの重みとして与える。たとえば、120 秒間に逸脱値 72.2 を持つ 3-Gram が、2.0 秒、3.0 秒の継続時間長で 2 回あったとき、これから重心値 72.2、重み 5.0 のクラスタができる。

② 次に、各クラスタの重心座標間の距離を求め、最も距離の近い 2 つのクラスタを統合する。新しいクラスタの重心値は、各 3-Gram の継続時間長の合計を重みとして、2 つのクラスタの重心座標の重みつき平均で算出する。

③ ② を、 $n = 4$ になるまで繰り返す。

④ 4 つのクラスタができただ後、最も逸脱度の高いクラスタと 2 番目に高いクラスタの重心座標の midpoint を閾値として決定する。

図 10 に、実際に実験で計測されたユーザ 1 人の、会話開始 120 秒間に表出した 3-Gram のプロットとクラスタリングの結果を示す。このとき、最も逸脱度（重心座標）の高いクラスタ 1 と 2 番目に高いクラスタ 2 の重心座標は、それぞれ 76.4, 61.8 であり、その midpoint 69.1 が閾値となる。実際にこのとき、ユーザ関心低下/観察者判定ボタンが押されていた（図中黒色ブロック部分）。このように、会話開始から 120 秒間のデータから閾値を算出し、その後のユーザの注視行動をこの閾値を用いて判別する方式をとることにより、個人適応が迅速

に行われるシステムを実現することが可能になる。

5.2 推定アルゴリズムの評価

個人適応型の会話参加態度推定アルゴリズムの評価として、収録実験で収集した10人分の視線データを用いて、会話開始後120秒後以降の、会話参加態度（関心低下/判定ボタンが押下された状態）の予測を行った。予測は10秒刻みで行い、10秒間に1度でも閾値を

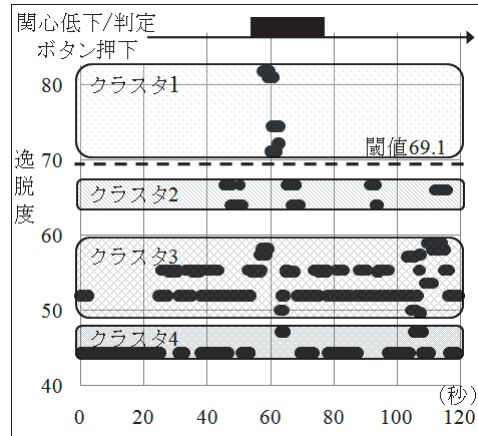


図 10 逸脱度の閾値の算出例

Fig. 10 Example of clustering eye gaze behaviors.

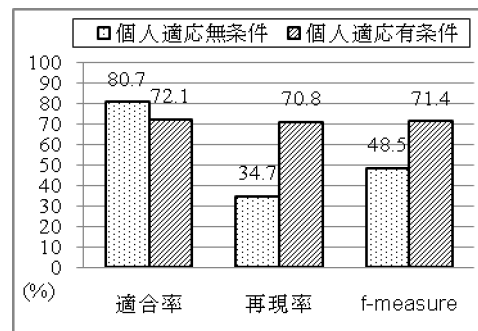


図 11 会話参加態度推定アルゴリズムの評価結果

Fig. 11 Evaluation of conversational engagement estimation algorithm.

超える 3-Gram が表出したとき、理想の会話参加態度からの逸脱と判定し、同一区間内でどちらかのボタンが押下されていれば正解とした。実験条件は以下の2種類である。

- 個人適応無条件：全員のデータから逸脱度の閾値を1つ算出し、それを全員に適用する。
- 個人適応有条件：個人ごとに閾値を算出し、それぞれのデータに適用する。

推定の結果を図 11 に示す。個人適応有条件は、個人適応無条件に比べて特に再現率が大幅に向上しており、適合率 72.1%、再現率 70.8%、F-measure 71.4%という結果が得られ、個人適応型推定アルゴリズムの有用性が示された。

6. 会話参加態度推定機構の実装

前章で提案した会話参加態度推定アルゴリズムを、視線計測装置を統合した会話参加態度推定機構として実装した。図 12 にシステム構成を示す。

(1) 視線計測装置制御部

ここでは、視線計測装置 Tobii の SDK に用意された API を利用して、視線計測装置を操作し、視線情報を取得する。ユーザ視線の計測が開始されると、視線位置座標（スクリーン上の座標）とシステム時間が 50 fps で視線計測装置制御部から会話参加態度推定部に送

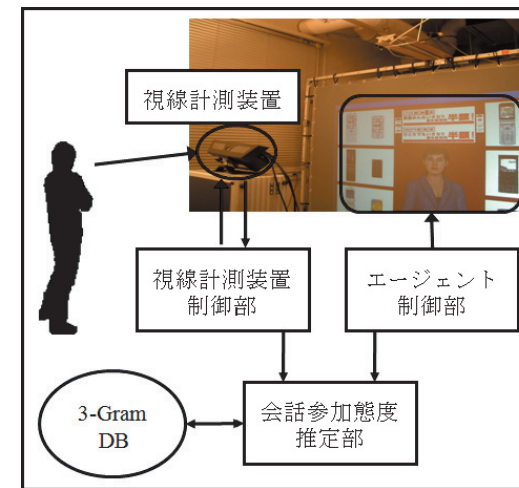


図 12 会話参加態度推定機構

Fig. 12 System architecture of conversational engagement estimation mechanism.

られる。

(2) 会話参加態度推定部

次に、視線計測装置制御部から送られてくる視線位置情報と、エージェント制御部から取得されるエージェントの現在の説明対象の情報を用いて、注視遷移 3-Gram を生成する。さらに、3-Gram の逸脱度を保有するデータベースから、生成された 3-Gram の逸脱度を取得する。エージェントの説明開始から 120 秒間に取得した逸脱度から閾値を算出し、その後、閾値を超える逸脱度が観測されると、ユーザの会話参加態度が逸脱したと判定し、そのメッセージをコマンドラインに出力する。

7. インタラクション評価実験

7.1 実験手続き

ユーザの会話参加態度に応じてエージェントの振舞いを変化させることの有用性を検証するために、評価実験を行った。被験者は、先の収録実験とは異なる 20 代の男性 7 人（情報系専攻学生）である。エージェントの動作ならびに発話は Wizard-of-Oz システムによって制御し、被験者には、会話データ収集実験と同じ教示を与えた（3.2 節参照）。

実験条件は以下の 2 条件であり、各被験者が両方を行う被験者内の実験デザインを採用した。このため、説明対象となる携帯電話は各条件 3 つずつとした。

- 実験条件：会話参加態度推定機構によってユーザが逸脱した会話参加態度であると推定されたとき、次発話でユーザの様子をうかがう発話を生成
- 対照条件：8 発話ごとに定期的にユーザの様子をうかがう発話を生成

なお、ユーザの様子をうかがう発話として、「ここまでよろしいでしょうか」、「他の携帯について説明をしましょうか」等 4 種類の発話をランダムに生成した。

評価方法は、リッカート法による 6 段階（1~6 点）の主観評価、ならびに客観的評価指標として、逸脱した会話参加態度の出現頻度を採用した。

主観評価における評価項目は、表 1 に示す 7 項目である。各評価項目に関し、4~5 個の質問に答えさせ、その被験者ごとの平均値を項目の評価値とした。

7.2 実験結果

7.2.1 主観評価結果

全ユーザの主観評価値の平均値を図 13 に示す。対照条件に比べて、実験条件がほとんどの評価項目で評価が高かった。特に、「態度への気づき」と、「対応の適切性」という項目で、実験条件の評価が大幅に高く、対応のある t 検定の結果、ともに有意差が認められた（両側

表 1 主観評価の評価項目

Table 1 Queries in questionnaire.

- 態度への気づき：エージェントがあなたの態度に気づいていたと感じたか
- 対応の適切性：あなたの態度に応じた適切なインタラクション（振る舞い）であったか
- 会話の円滑さ：会話が円滑であると感じたか
- 好感度：エージェントへの好感度が向上したか
- 動作の自然さ：エージェントの動作が自然であったか
- 人間らしさ：エージェントに人間らしさを感じたか
- 知能：エージェントに知能を感じたか

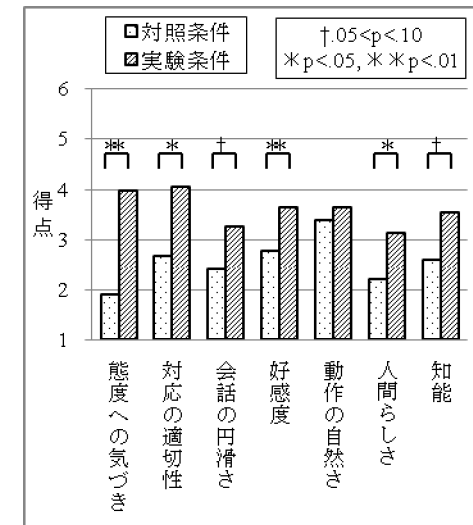


図 13 主観評価結果

Fig. 13 Results of subjective evaluation.

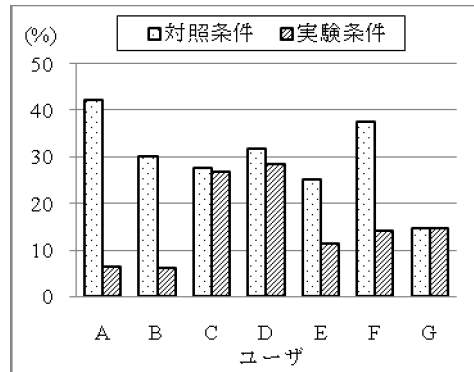


図 14 逸脱した会話参加態度の割合

Fig. 14 Frequency of disengaging gaze behaviors.

検定: $t(6) = 4.54; p < .01$, $t(6) = 3.62; p < .05$). この結果は、提案方式により推定された会話参加態度に基づき、エージェントの振舞いを変更することの有用性を実証するとともに、ユーザの会話参加態度に気づく会話エージェント実現可能性を示唆するものである。

また、関連する項目として、「会話の円滑さ」、「好感度」、「人間らしさ」、「知能」の評価項目においても、実験条件で評価が高く、有意差または有意傾向が認められた（両側検定: $t(6) = 2.26; .05 < p < .10$, $t(6) = 3.83; p < .01$, $t(6) = 3.34; p < .05$, $t(6) = 2.22; .05 < p < .10$ ）。

エージェントの「動作の自然さ」を問う項目では有意な差は認められなかったが、実際に、2条件間で、エージェントの動作には違いがないため、これは妥当な結果であると考えられる。

7.2.2 逸脱した会話参加態度の頻度

対話中に出現した全 3-Gram に対する、逸脱と判定された 3-Gram（閾値を超えた逸脱度を持つ 3-Gram）の割合をユーザごとに算出した結果を図 14 に示す。全ユーザにおいて、逸脱 3-Gram の割合は、対照条件より実験条件の方が少なく、対応のある t 検定の結果、有意差が認められた ($t(6) = 2.79; p < .05$)。よって、ユーザの態度に応じてインタラクションを行うことで、対話におけるユーザの逸脱した態度を減少させることができたといえる。

7.3 考 察

評価実験において、ユーザの会話参加態度の推定を行わずに振る舞うエージェントに比べ

て、態度の推定結果に応じて振る舞うエージェントの主観評価が、全体的に高くなる傾向を示した。「態度への気づき」と、「対応の適切性」という項目で、評価が大幅に高かったことは、ユーザが理想の会話参加態度から逸脱したことを、会話参加態度推定機構が正しく検出していたことを裏付ける結果であり、また逸脱時に、ユーザの様子をうかがう一連の制御がユーザへの対応として適切であったことを示唆している。

本評価実験では、会話参加態度の逸脱にエージェントが気づくことによる効果を調べることが目的であった。そのため、エージェントの発話は、逸脱に気づいていることを伝えるだけにとどめており、ユーザの注意をひきつけるための積極的な働きかけは行っていない。したがって、主観評価において、絶対評価として 3~4 点と十分に高い評定値が得られなかったことは事実であるが、ユーザの態度が逸脱していると見なされるときに、エージェントがユーザの態度に気づいていることを伝えるだけでも、統計的に有意な効果が得られることを本評価実験の結果は示している。ユーザによる逸脱した行動を検出した後に、どのようにして、ユーザの注意を再びひきつけ、積極的な参加態度に戻すのかという課題は、次のステップとして取り組まなければならない。

また、対話中に表出した逸脱 3-Gram の割合が減少したが、これは単にエージェントからの働きかけの回数が増えたからではない。ユーザの様子をうかがう発話の、セッションごとの平均頻度は、対照条件: 8.35 発話、実験条件: 7.74 発話であり、対応のある t 検定の結果、有意差は見られなかった ($t(6) = 0.598; p > .05$)。つまり、エージェントからの働きかけの頻度には有意差がなかったにもかかわらず、逸脱した会話参加態度が減少し、ユーザの主観評価も大きく向上したことは、エージェントからの働きかけのタイミングが適切であったことを示唆している。

「会話の円滑さ」、「好感度」、「知能」、といった関連項目の評価向上は、以上の結果の副作用として生じたとも考えられる。さらに、本システムの直接的な目的ではなかった「人間らしさ」についても評価が向上したことは興味深い。対話相手の態度に応じて、相互に適応的に会話を進めることは、人間の持つ重要なコミュニケーション能力であり、この機能をエージェントが持つことは、エージェントが人間にとってより身近な存在となるために重要な要素になると考えられる。

8. ま と め

本稿では、ユーザの会話参加態度に応じて適応的に振る舞う会話エージェントを目指し、ユーザの会話参加態度をユーザの視線、特に注視対象 3-Gram から推定するアルゴリズム

を提案し、システムを実装した。さらに、インタラクション評価実験により、本手法ならびにシステムの有用性が示された。本研究における一連の実験、分析から、コミュニケーション機能を有する人工物が、人間と共存を図るために人間らしさを目指すならば、ユーザの態度を理解し、適応的に振る舞うことが重要な要因になることが示唆された。

今後は、実装した会話参加態度推定機構を会話エージェントの対話制御機構に組み込むことにより、システムの全自動化を進める。また、本研究では、入力として視線情報からの注視遷移パターンのみを用いたが、頭の傾きや体全体の姿勢¹⁹⁾、瞳孔の大きさ²⁰⁾等もユーザの内的状態を表出しているという報告がある。これらの視線以外のモダリティを統合することにより、会話参加態度推定モデルをより詳細化していく必要がある。

謝辞 本研究の一部は、文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しいIT基盤技術の研究」によるものである。ここに記して感謝する。

参 考 文 献

- 1) Cassell, J., Sullivan, J., Prevost, S. and Churchill, E. (Eds.): *EMBODIED CONVERSATIONAL AGENTS*, The MIT Press (2000).
- 2) Kendon, A.: Some Functions of Gaze Direction in Social Interaction, *Acta Psychologica*, Vol.26, p.22-63 (1967).
- 3) Clark, H.H.: *Using Language*, Cambridge University Press, Cambridge (1996).
- 4) Argyle, M. and Cook, M.: *Gaze and Mutual Gaze*, Cambridge University Press, Cambridge (1976).
- 5) Duncan, S.: Some signals and rules for taking speaking turns in conversations, *Journal of Personality and Social Psychology*, Vol.23, No.2, pp.283-292 (1972).
- 6) Argyle, M. and Graham, J.: The Central Europe Experiment – looking at persons and looking at things, *Journal of Environmental Psychology and Nonverbal Behaviour*, Vol.1, pp.6-16 (1977).
- 7) Anderson, A.H., Bard, E., Sotillo, C., Doherty-Sneddon, G. and Newlands, A.: The effects of face-to-face communication on the intelligibility of speech, *Perception and Psychophysics*, Vol.59, pp.580-592 (1997).
- 8) Whittaker, S.: Theories and Methods in Mediated Communication, *The Handbook of Discourse Processes*, A. Graesser, Gernsbacher, M. and Goldman, S. (Eds.), pp.243-286, Erlbaum, NJ. (2003).
- 9) Nakano, Y.I., Reinstein, G., Stocky, T. and Cassell, J.: Towards a Model of Face-to-Face Grounding, *The 41st Annual Meeting of the Association for Computational Linguistics (ACL03)*, Sapporo, Japan (2003).
- 10) Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., Werf,

R.J.v.d. and Morency, L.-P.: *Virtual Rapport*, 6th International Conference on Intelligent Virtual Agents, Springer: Marina del Rey, CA (2006).

- 11) Sidner, C.L., Kidd, D.C., Lee, C. and Lesh, N.: Where to Look: A Study of Human-Robot Engagement, *ACM International Conference on Intelligent User Interfaces (IUI)*, pp.78-84 (2004)
- 12) 高木啓伸：視線の移動パターンに基づくユーザの迷いの検出—効果的な作業支援を目指して，情報処理学会論文誌，Vol.41, No.5 (2000).
- 13) 水口 充，浅野 哲，佐竹純二，小林亮博，平山高嗣，川嶋宏彰，小嶋秀樹，松山隆司：Mind Probing：システムからの積極的な働きかけによる視線パターンからの興味推定，情報処理学会ヒューマンコンピュータインタラクション研究会 (SIGHCI)，Vol.2007, No.99 (2007).
- 14) Qvarfordt, P. and Zhai, S.: Conversing with the user based on eye-gaze patterns, *Proc. CHI*, pp.221-230 (2005).
- 15) Nakano, Y., Okamoto, M., Kawahara, D., Li, Q. and Nishida, T.: Converting Text into Agent Animations: Assigning Gestures to Text, *Proc. Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2004)*, Companion Volume, pp.153-156 (2004).
- 16) Haptik, Inc. (online). available from <http://www.haptik.com/> (accessed 2008-03-24)
- 17) 汎用的知的音声合成システムボイスソムリエ，日立ビジネスソリューション (オンライン). 入手先 <http://www.b-sol.jp/voice/index.html> (参照 2008-03-24)
- 18) Kipp, M.: Anvil – A Generic Annotation Tool for Multimodal Dialogue, *7th European Conference on Speech Communication and Technology*, pp.1367-1370 (2001).
- 19) P. ブール：姿勢としぐさの心理学，北大路書房 (2001).
- 20) Hess, E.H.: Attitude and Pupil Size, *Scientific American*, pp.46-54 (1965).

(平成 20 年 3 月 24 日受付)

(平成 20 年 9 月 10 日採録)



石井 亮

2006年東京農工大学工学部情報コミュニケーション工学科卒業。2008年同大学院工学府情報工学専攻修士課程修了。同年日本電信電話株式会社入社，NTTサイバースペース研究所勤務。アバタコミュニケーションや人と会話エージェントの会話における，人の視線行動を利用した会話促進の研究に従事。日本バーチャルリアリティ学会会員。



中野有紀子（正会員）

1990年東京大学大学院教育学研究科修士課程修了。同年日本電信電話株式会社入社。2002年MIT Media Arts & Sciences 修士課程修了。2002～2005年（独）科学技術振興機構社会技術研究開発センター専門研究員，2005～2008年東京農工大学大学院工学府特任准教授を経て，2008年4月より成蹊大学理工学部情報科学科准教授。知的で自然なユーザインタフェースの実現に向けて，人との言語・非言語コミュニケーションが可能な会話エージェントの研究に従事。博士（情報理工学）。電子情報通信学会 HCG 第3種研究会ヴァーバル・ノンヴァーバル・コミュニケーション研究会副委員長。ACL，ACM，人工知能学会各会員。
