情報処理学会論文誌　Vol. 52　No. 12　3625–3636 (Dec. 2011)

視線と頭部動作に基づくユーザの会話参加態度の推定

石井　亮[†1,†2,†3]　大古亮太[†3]
中野有紀子[†3]　西田豊明[†1]

人間同士の対面会話において，話者は聞き手の視線や頭部動作から会話への参加態度をモニタリングし，聞き手の態度に応じて会話戦略を変更している．本研究では，ユーザとの会話において，ユーザの態度に適応的に振る舞うことのできる会話エージェントの実現を目指し，複数の視線情報と頭部動作の情報から会話参加態度を推定する手法を構築することを目的とする．我々はこれまでに注視対象の遷移パターンからユーザの会話参加態度を推定する手法を提案・実装してきた．その中で，推定精度向上の必要性，アイトラッカの眼球検出失敗時に推定が行えないことによる頑健性の問題が明らかになった．そこで本研究では，注視パターン，注視時間，注視位置の移動距離，瞳孔径に関する複数の視線情報，および眼球検出によらない情報としてヘッドトラッキングによる頭部姿勢の情報に関して，会話参加態度との関連性を検証する．さらに，これらのパラメータの様々な組合せによる会話参加態度推定モデルを比較し，視線情報と頭部動作情報を統合した推定モデルの性能が最も高いことを示す．

Estimation of User's Conversational Engagement Using Eye-gaze and Head Movement Information

Ryo Ishii,[†1,†2,†3] Ryota Ooko,[†3] Yukiko Nakano[†3]
and Toyoaki Nishida[†1]

In face-to-face conversations, speakers continuously check whether the listener is engaged in the conversation by monitoring the partner's eye-gaze behaviors and head movement. The goal of this study is to build an intelligent conversational agent that can recognize user's engagement from multiple types of eye-gaze and head movement information. In our previous work, we developed a method of estimating the user's conversational engagement using user's gaze transition patterns. However, it was not accurate enough. In this study, we looked at other factors, such as gaze duration, distance of eye movement, pupil size, and head motion. Then, based on the analysis for these factors, we used them as prediction parameters, and set up prediction models which consist of different combinations of parameters. By comparing the performance of these models, we discussed the useful parameter combinations. As the result, it was revealed that a model that integrates eye gaze and head motion information performed the best andwas able to predict the user's conversational engagement quite well.

1. はじめに

対話において，言語行動は発話の意味内容を伝えるために機能するが，非言語行動は言語行動を補助する機能を有する．その 1 つとして，対話を成立させるうえで必要不可欠な会話参加態度（engagement）がある．Sidner らは，会話参加態度とは 2 人以上の会話参加者間による知覚的なつながり（connection）が確立，維持，終結する過程であると定義している[1)]．たとえば，会話を維持するために話し手は聞き手が会話に注意を向け，適切に会話に参加しているか否かを確認しながら発話を行い，一方，聞き手は視線や頷きといった非言語行動によって会話に注意を向けていることを表出し，コミュニケーションが維持されることに肯定的であることを話し手に伝えている[2)]．

したがって，相手の会話参加態度を認識し，自分の会話参加態度を表出することは対面会話において不可欠であるが，人と人工物とのインタラクションのデザインにおいても，会話参加態度を考慮することは重要であると考えられる．ユーザがシステムとの会話に積極的に参加しているのか否かをシステム側が自動的に検知することができれば，会話に関心を失っているユーザに対して，会話への積極的な参加を促したり，話題を変えたりする等，システムが適応的に振る舞うための有用な情報となりうる．しかし，視線や頷き等が会話参加態度を示すシグナルであることは，すでにコミュニケーション科学により明らかにされているものの[3)]，それをシステムにより自動認識するための手法についてはまだ十分に研究されていない．

我々は，人対会話エージェントの対話において，ユーザの会話参加態度を察知し，適応的に振る舞うことができる会話エージェントの実現を目指し，アイトラッカから取得されるユーザの視線行動から，ユーザの会話参加態度を推定する方式の提案・実装に取り組んでき

†1 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University
†2 日本電信電話株式会社 NTT サイバースペース研究所
NTT Cyber Space Laboratories, NTT Corporation
†3 成蹊大学理工学部
Faculty of Science and Technology, Seikei University

た．ここでは，視線行動の中でも特に注視対象遷移パターンに着目し，パターンと会話参加態度に強い相関があることを示した．さらに，会話中にリアルタイムに計測される視線情報を用いて，ユーザの会話参加態度を推定可能なアルゴリズムを構築した．これらを対話システムに実装し被験者実験を行った結果，会話エージェントが会話参加態度を察知し，適応的に振る舞うことで，ユーザの会話への関心低下を防ぎ，エージェントに対する印象が向上することが確認された[4)–6)]．しかしながら，我々が構築した会話参加態度の推定アルゴリズムは，推定精度がまだ十分であるとはいえず，また会話中にユーザの頭部が大きく動くことによりアイトラッカでの眼球の検出ができなくなると，推定が不能になるため，実装システムの頑健性にも課題があった．

そこで本研究では，より頑健かつ高精度な会話参加態度推定機構を実現することを目指し，複数の視線情報と頭部情報を複合的に用いた推定手法を提案する．具体的には，以下の事項に取り組む．

- ユーザ対会話エージェントの会話コーパスデータを分析し，注視時間，視線の移動距離，瞳孔径等，複数の視線情報と頭部の動作情報について，会話参加態度との関連性を調査する．
- この分析に基づき，視線および頭部情報のパラメータを用いた新しい会話参加態度推定手法を構築する．

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行う．3 章では，対話収録実験および取得データについて述べ，4，5 章では，視線および頭部動作の各パラメータのデータ分析の結果を報告する．6 章では，分析で有用と判断されたパラメータの様々な組合せによる会話参加態度推定モデルを比較する．7，8 章では議論を行った後，最後に本研究のまとめを述べる．

2. 関 連 研 究

対面会話における視線や頭部動作による注視行動の機能について，コミュニケーション科学や社会心理学の分野において，数多くの研究が行われている．その中で，聞き手が話し手に視線を向けることは，聞き手の注意が会話に向いていること，そして会話が円滑に進んでいることを話し手に伝える肯定的なフィードバックとなり[2),7)]，円滑なターン交代に寄与していることが明らかにされている[8)]．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．これは，共有された対象物が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[9)–11)]．

このような人間の注視行動に関する知見に基づき，会話エージェントやロボットにおけるコミュニケーション機能を実装した研究が報告されている．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を提案している[12)]．また，人対コミュニケーションロボットの研究では，Miyauchi らはロボットがユーザの視線を認識し，ロボットを見ているユーザに対して顔を向ける等のコミュニケーション行動をとるロボットを開発している[13)]．さらに，Sinder ら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するコミュニケーションロボットを開発している[1)]．これらの研究はヘッドトラッカから得られる頭部姿勢・頭部動作の情報が，ユーザの注視行動の認識において有用であることを示している．

一方，より厳密なユーザの視線情報をアイトラッカによって測定し，ユーザの状態を理解する試みが，HCI の研究分野で行われている．Qvarfordt ら[14)] は，人同士の協同作業における視線パターンと継続時間長の知識を利用し，街の観光案内のインタラクティブシステム “iTourist” を開発している．ユーザ評価の結果，ユーザの興味を視線パターン，注視継続長といった視線情報を用いることで推定可能であることが示されている．Iqbal ら[15),16)] は，インタラクティブな対話型タスクにおいて，心的作業負荷と瞳孔反応の関連性について実験的に検証を行った．その結果，処理時間が長く，処理が困難である心的作業負荷の大きいタスクにおいて，瞳孔反応が顕著に大きくなると報告している．Eichnerl ら[17)] は，アイトラッカを用いてユーザが興味を持つ対象物を検出する方法を提案し，この機構をインタラクティブシステムに統合している．このように，アイトラッカにより取得される視線情報からユーザの興味対象や内的状態を推定できる可能性が示されていることから，我々の目指す会話参加態度の推定においてもこれらの視線情報が大きく寄与すると考えられる．

以上の関連研究は，対面会話において視線や頭部姿勢・動作が重要なコミュニケーションシグナルとなっていること，アイトラッカやヘッドトラッカから得られた計測データはこれらを推定するうえで有用であることを示唆するものである．そこで本研究では，視線および頭部の複数の情報を用いて，ユーザの会話参加態度の推定技術を構築することを目的とする．

3. マルチモーダルコーパスの構築

3.1 人対会話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，会話エージェントの言語・非言語行動を制御可能な Wizard-of-Oz システムを開発し，被験者と会話エージェントとの 1 対 1 の会話を収録した[4]．被験者はシステムの “ユーザ” として，携帯電話の販売員を務める会話エージェントと対話を行い，6 台の携帯電話についての説明を受けた（**図 1** 参照）．各携帯電話の説明は 3〜5 分間であり，会話エージェントの説明発話総数は 109 発話，会話所要時間は平均約 16 分間であった．会話中にユーザは自由に会話エージェントに質問することが可能であり，携帯電話の機能や価格についての質問，「はい（YES）」または「いいえ（NO）」で回答可能な質問，さらには今の説明を中断し，説明対象を次の携帯電話に変えること（話題転換の依頼）が許された．

3.2 取集データ

上記 3.1 節の手続きに従い 10 会話分の会話データを収録し，以下のユーザおよび会話エージェントの言語・非言語行動のデータを収集した．

図 1　対話実験における会話エージェント
Fig. 1　Conversational agent used in the WOZ experiment conversation.

- 言語データ：ユーザの発話データは録音した音声データから書き起こしを行った（総数 61 発話）．会話エージェントの発話データは，Wizard-of-Oz システムのログを用いて生成した（総数 951 発話）．
- 非言語データ：ユーザの眼球運動（注視位置，瞳孔径）は，非装着型の視線計測装置 Tobii X-50 で計測され，会話エージェントと携帯電話が映し出されたスクリーン（図 1）上の注視位置が 50 Hz で計測・収集された．また，計測データに含まれるノイズを除去するために，半径 20 pixel の円内に 50 ms 以上視線が停留した区間のみをデータとして採用した．一方，ユーザの頭部動作は，非装着型の頭部姿勢計測装置[18] により計測され，頭部の 3 次元位置と回転角度が 30 Hz で計測された．ユーザの視線情報については 4 章で，頭部動作については 5 章で詳しく述べる．会話エージェントの注視行動は，Wizard-of-Oz システムのログから各時間における会話エージェントアニメーションの種類を取得し，正面顔のアニメーションが実行されていた時間を，“ユーザへの注視”，説明対象の携帯電話に顔を向けているアニメーションが実行されていた時間を “携帯電話への注視” と見なし，50 Hz でデータ化した．その他，ユーザのバストショットおよび会話エージェントの映像のビデオデータを収録した．
- 会話参加態度の指標：実験後，対話実験には参加していない別の 10 人の被験者が，実験で収録されたユーザと会話エージェントのビデオを視聴し，ユーザの会話参加態度についてラベリングを行った．ラベリングにはアノテーションツール anvil[19] を用い，被験者はユーザが会話に積極的に参加していないと感じられる箇所にラベリングを行った．このとき，ビデオは自由に再生，コマ送り，一時停止が可能であった．これら計 10 人の被験者のラベリングデータを合算した．この 0〜10 の 11 段階の合算値を “会話参加態度の非積極度” と定義し，本研究では，会話参加態度の指標として扱った．この指標において，たとえば 0 であれば，非常に積極的な会話参加態度であり，10 であれば極めて非積極的な会話参加態度であることを示す．なお，この会話参加態度の非積極度は 30 fps のデータである．

これらすべての情報を，アノテーションツール anvil を用いて統合し，言語・非言語情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．

4. 視線情報の分析

本研究では，これまでに行った視線遷移パターンの分析[4)–6)] に加えて，相互注視，注視時間，視線の移動距離，瞳孔径の 4 つの視線情報と，頭部の位置，回転角度の 2 つの頭部

情報を主要な分析項目とし，会話参加態度との相関関係を明らかにする．

4.1　視線遷移パターン 3-gram

4.1.1　3-gram の定義

収集された注視行動データは，最短視線停留時間を 50 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，我々は，注視対象の遷移パターンを注視 3-gram として定義し，会話参加態度との関連性を明らかにしてきた[4)–6)]．次に，3-gram の作成手順を説明する．アイトラッカは，ユーザが瞬きをした際に，視線方向の計測ができない．瞬きは最大で 200 ms 程度の継続長があることから，200 ms 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，注視対象が変化したとき，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．このとき，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．

次に，3-gram の構成素となる注視行動を以下の 3 つに分類しラベル付けした．

- **T**：会話エージェントが説明している対象を見る．
- **AH**：会話エージェントを見る．
- **F1〜F3（F1 ≠ F2 ≠ F3）**：会話エージェントの説明対象およびエージェント以外の対象（説明対象以外の携帯電話や広告）を見る．なお，F1，F2，F3 はそれぞれ違う対象物を示す．たとえば，会話エージェントの説明対象が携帯電話 A であったとき，携帯電話 B–携帯電話 E–携帯電話 B の 3-gram は，F1-F2-F1 と分類される．

なお，3 章の実験では，会話エージェントは多くの時間，説明対象となる携帯電話を注視しながら説明していたため，ユーザの注視行動が T の場合には，エージェントとユーザが説明対象を共同注視している状態となった．

図 2 に 3-gram 作成の具体例を示す．会話エージェントの説明対象が携帯電話 A であったとき，ユーザの注視対象が “携帯電話 A，(99 ms)，エージェント，(66 ms)，エージェント，(132 ms)，携帯電話 E，(66 ms)，広告”，つまり，携帯電話 A と 1 つ目のエージェントの間に 99 ms の間隔，1 つ目のエージェントと 2 つ目のエージェントの間には 66 ms，2 つ目のエージェントと携帯電話 E との間には 132 ms，携帯電話 E と広告の間には 66 ms の

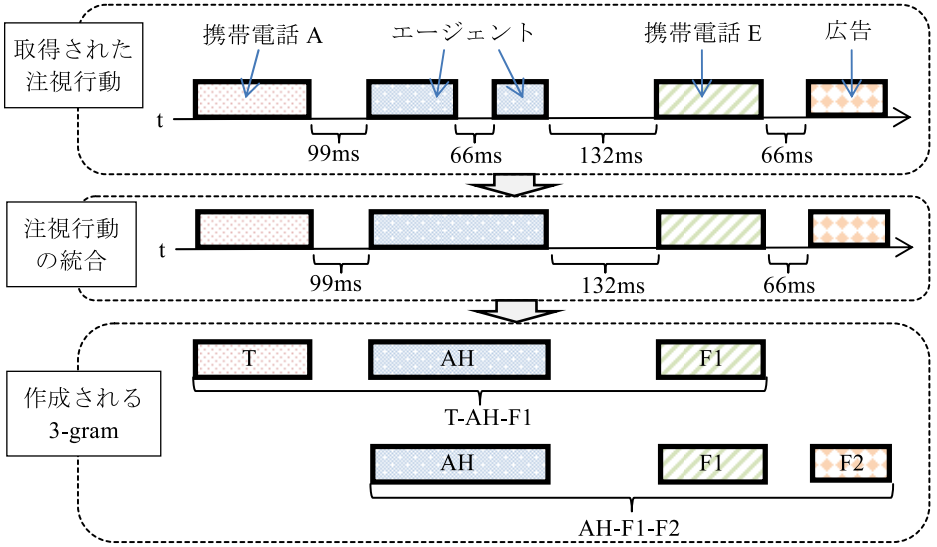

図 2　3-gram の作成
Fig. 2　Construction of 3-gram.

間隔がある場合を考える．まず，連続する 2 つのエージェントへの注視行動の間隔は 99 ms であるから，2 つのエージェントへの注視行動は統合される．統合後の視線行動である，“携帯電話 A，(99 ms)，エージェント，(132 ms)，携帯電話 E，(66 ms)，広告” からは，携帯電話 A–エージェント–携帯電話 E とエージェント–携帯電話 E–広告の 2 つの 3-gram が生成され，ラベル付けを行った結果，それぞれ T-AH-F1 と AH-F1-F2 の 3-gram が生成される．

4.1.2　3-gram と会話参加態度の相関

10 人のユーザ役被験者の視線データから 3-gram を作成し，会話参加態度の非積極度との相関を分析した．分析には，3-gram の 3 つ目の構成要素の開始から終了までの時間における会話参加態度の非積極度を算出し，これを 3-gram の非積極度の指標とした．たとえば，図 2 に示すような T-AH-F1 の 3-gram の 3 つ目の構成素としての F1，つまり T，AH に引き続き観測される F1 は，収集データ中で合計 62.07 秒（1,862 フレーム）観測された．このとき，各フレーム（30 fps）における会話参加態度の非積極度の平均値は 3.89 であり，これを T-AH-F1 の非積極度の値とした．

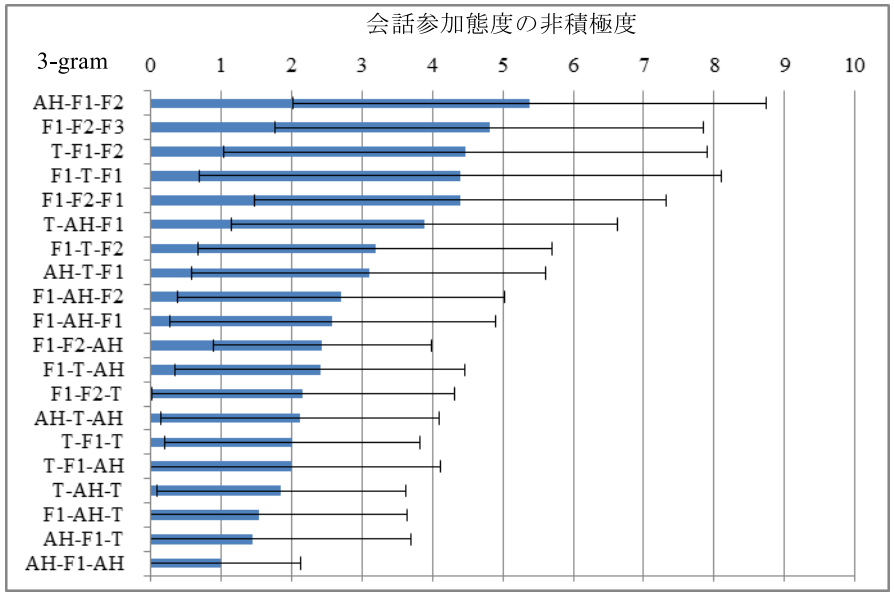

図 3　注視対象 3-gram と会話参加態度の相関関係
Fig. 3　Eye-gaze 3-grams and disengagement scores.

以上の方法で算出した会話参加態度の非積極度の平均値を，全 3-gram において算出した結果を**図 3** に示す．図 3 の横軸は非積極度の平均値，縦軸は 3-gram の種類を示す．この図から，3-gram の種類によって会話参加への非積極度が大きく異なっていることが分かる．たとえば，AH-F1-F2 は最も非積極度が高く，値が 5 を超えているのに対して，AH-F1-AH は 0.99 である．この結果から，会話参加への非積極度が高い 3-gram は，積極的な会話参加態度から逸脱した注視行動を示し，逆に値が低い 3-gram は，積極的な会話参加態度を表すシグナルとして機能していると考えられる．3-gram によって会話参加への非積極度に違いが見られたことから，3-gram は会話参加態度を推定するパラメータとして有用であると考えられる．

4.2　相 互 注 視

対面会話において，相互注視は話者に対する聞き手からのフィードバックとして機能をする．具体的には，相互注視は聞き手が話者にフィードバックを与えるとともに，話者は聞き手からのフィードバックを受け取る機会となる．これにより，話者は発話を続けるか，聞き手に発話権を譲る等の会話戦略を決定する．ここでは，このような相互注視に注目し，会話参加態度との関連性を分析する．

本分析では，先の 4.1 節で生成した 3-gram における注視対象の分類において，ユーザが会話エージェントを注視している時間（ラベル AH）における会話エージェントの注視行動に着目し，ユーザとエージェントが相互注視を行っていたか否かにより，ラベル AH を以下のように細分類した．

ユーザが会話エージェントを注視しているときに，

- **M**：会話エージェントもユーザを注視（相互注視）している．
- **A**：会話エージェントはユーザ以外（説明対象）を注視している．

10 人のユーザの視線データから M ラベルを含む新しい 3-gram を作成し，4.1.2 項と同様に，会話参加態度の指標との相関を集計した結果を**図 4** に示す．M ラベルを含む 3-gram は，図 4 中で赤い斜線の棒で示されている．M ラベルを含む 3-gram はその他の 3-gram に比べて，会話への非積極度が 10 または 0 に二極化する傾向にあることが確認できる．たとえば，AH-F1-AH を構成要素として持つ 3-gram の平均値は 0.99 であった（図 3 参照）．次に相互注視を考慮した場合，AH-F1-AH を構成要素として持つ 3-gram は A-F1-A，M-F1-M，M-F1-A，A-F1-M の 4 つの 3-gram に分類される．このとき，会話参加態度はそれぞれ 2.87，0.26，0.27，1.70 であった（図 4 参照）．また，会話参加態度の非積極度が極端な値（4 以上もしくは 1 以下）となる 3-gram が，M ラベルを考慮しない場合は全体の 30.0%であったのに対し，M ラベルを考慮した場合は，45.2%と増加した．このことから，ユーザの注視行動だけでなく会話エージェントの注視行動を考慮し，相互注視に着目することで，3-gram と会話参加態度との相関関係がより明確化されたと考えられる．これを検証するために，M ラベルの考慮の有無によって，3-gram の非積極度の分布に統計的な差があるかを検定した．4.1.2 項での 3-gram の非積極度の平均値の分散と M ラベルを含む 3-gram を導入した場合の分散が異なっているか否かを F 検定により調べた結果，有意な差は見られなかった（$F(19, 41) = 1.55$，n.s.）．この原因として，今回の会話収集実験では，会話エージェントがユーザに注視を行う頻度が少なく，3-gram が全部で 42 種類であるのに対し，M ラベルを含む 3-gram の種類が 21 種類と十分に多くなかったことが考えられる．以上より，3-gram に M ラベルを考慮することで会話参加態度の推定に寄与するか否かは明確には結論づけられなかった．

4.3　注 視 時 間

我々はこれまで注視時間と会話参加態度の関連性を分析し，会話エージェントの説明対象

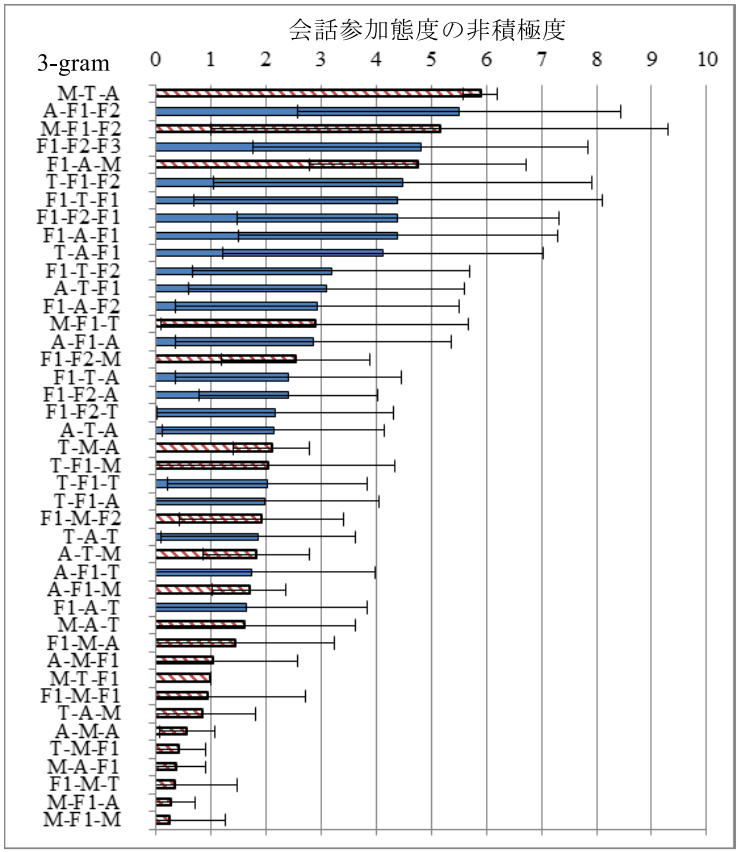

図 4　相互注視を加味した 3-gram と会話参加態度の相関関係
Fig. 4　Disengagement scoresfor 3-grams with M label.

となる携帯電話や会話エージェントを注視することは，積極的な会話参加態度と考えられることを明らかにした．また逆に，説明対象以外の対象物を長い時間注視することは非積極的な会話参加態度と考えられる[4]．しかしながら，これまでの分析では，一定時間（会話エージェントの 1 発話）に対する注視時間の割合を考慮していたが，注視時間と会話参加態度の相関は強固なものではなかった．

そこで本節では，4.2 節までに分類した 3-gram において，注視時間長を考慮した分析を行う．具体的には，3-gram の構成要素となる 3 つの注視行動の合計継続時間長によって 3-gram を細分類し，会話参加態度との相関を調査した．まず，3-gram の細分類のために，各 3-gram において継続時間長に関する閾値を設定し，長，中，短の 3 つに分類した．閾値は，各 3-gram の時間長の平均 μ と標準偏差 σ を用いて，以下のように定めた．

- 長：合計継続時間長 $>= \mu + \sigma/2$
- 中：$\mu + \sigma/2 >$ 合計継続時間長 $>= \mu - \sigma/2$
- 短：合計継続時間長 $< \mu - \sigma/2$

すべての 3-gram をこの方法で細分類し，継続時間長グループごと（長，中，短）の 3-gram の会話参加態度の平均値を算出した結果を**図 5** に示す．会話参加態度の値の算出方法は，4.1.2 項での分析と同様に，3-gram の第 3 構成素において得られた非積極度の値の平均値を用いた．たとえば，M-F1-T を構成要素に持つ 3-gram は，注視時間の μ が 7.85(s)，$\sigma/2$ が 2.03(s) であった．この結果を基に分類をした結果，時間長が長の 3-gram（時間長 $>= 9.88$）における会話参加値度の平均値は 4.93，時間長が中の 3-gram（$9.88 >$ 時間長 $>= 5.82$）では平均値は 3.47，短の 3-gram（時間長 < 5.82）における会話参加値度の平均値は 0.86 であった．時間長を考慮する前の会話参加態度の平均値が 2.89 であったのに比べて，時間長を考慮することで会話参加態度の値の差が広がっていることが分かる．そこで，時間長の考慮の有無によって，3-gram の非積極度の値の分布に統計的に差があるかを検証した．4.2 節での各 3-gram の非積極度の平均値の分散と，時間長によって長，中，短に分類された各 3-gram グループの分散との差異に関して F 検定を行った．その結果，時間長が長および中の 3-gram のグループにおける会話参加態度の分散と 4.2 節の時間長を考慮しない分析における会話参加態度の分散に有意差および，有意傾向が見られた（長：$F(19, 41) = 2.66$, $p < .01$，中：$F(19, 40) = 2.49$，$.05 < p < .10$，短：$F(19, 40) = 1.50$，n.s.）．この結果から，注視継続時間長を考慮することで会話参加態度との相関がより明確になると考えられる．また特に，時間長が長と中でのみ有意差および有意傾向が見られたことから，3 つの連続した注視行動の時間長がある程度長いとき，遷移パターン 3-gram によって，会話参加態度が積極的であったか否かをより明確に判断できる可能性があるといえる．よって，3-gram に加えて，注視時間を考慮することでより高精度に会話参加態度を推定できる可能性が示唆された．

4.4　視線移動距離

ユーザが積極的な会話参加態度であったとき，会話エージェントが説明する携帯電話を注意深く観察するため，単位時間あたりの視線移動距離は短くなると考えられる．一方，非積

(a)　時間長：長　　(b) 時間長：中　　(c) 時間長：短

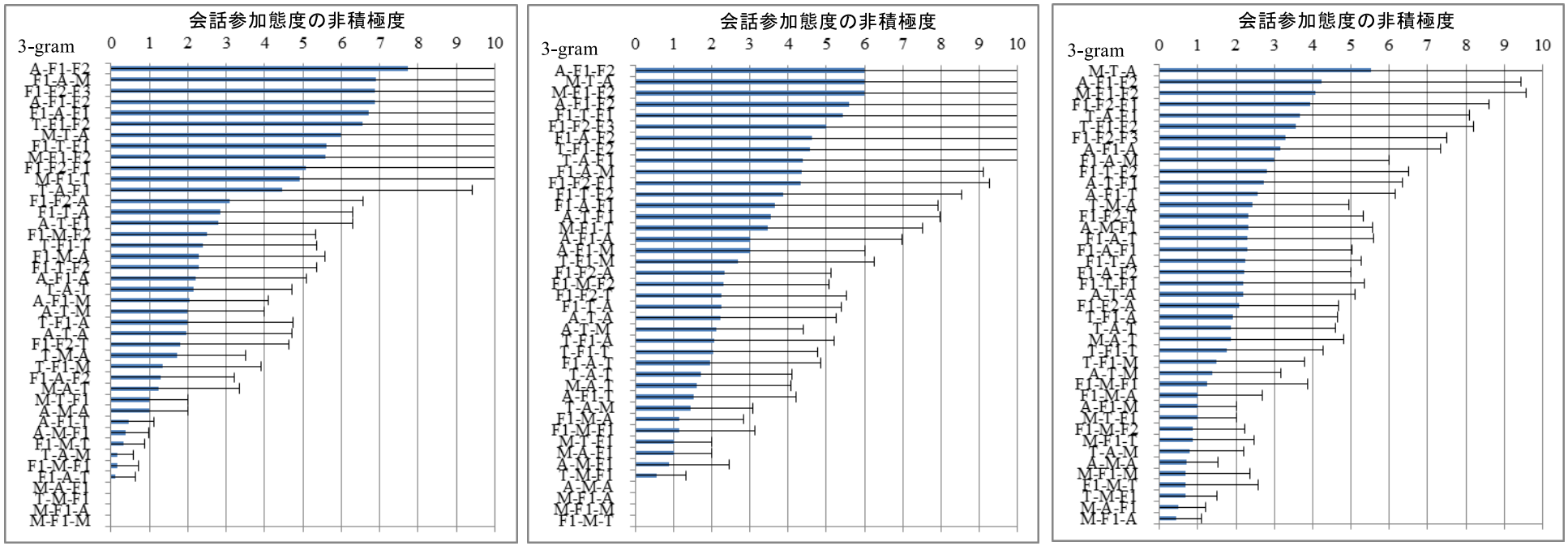

図 5　注視時間を考慮した 3-gram と会話参加態度の相関関係
Fig. 5　Disengagement scores with respect to 3-gram duration.

極的な会話参加態度である場合には，説明対象以外の携帯電話や広告等を頻繁に注視するため，視線移動距離は積極的な会話参加態度であったときに比べて，大きくなると考えられる．以上のような仮説を立て，本研究では単位時間あたりの視線移動距離と会話参加態度の関連性について分析を行った．

過去 400 ms の移動距離を算出し，距離の変化に応じて会話参加態度の平均値がどのように変化するかを調べた．具体的には，移動距離を 1 ピクセル刻みで集計し，会話参加態度の非積極度の平均値を算出した．その結果を**図 6** に示す．グラフの横軸は視線移動距離を示し，縦軸は会話参加態度の非積極度の平均値を示す．グラフに示されるように，ユーザが積極的な会話参加態度であるほど，視線移動距離は小さくなっており，両者の相関係数は 0.76 であった．以上より，視線移動距離は会話参加態度の推定に有効的なパラメータとなることが示唆された．

4.5　瞳　孔　径

人間は関心を持ったものを見たときや，興奮状態にあるときに瞳孔径が大きくなることが知られている[20]．そのため，ユーザの会話参加態度と瞳孔径にも関連性があると考えられる．たとえば，積極的な会話参加態度であったときは説明対象を注意深く観察するため瞳孔径が大きくなり，逆に非積極な会話参加態度であれば対象物を注意深く注視することはせずに瞳孔径が小さくなると考えられる．これらの仮説を立て，本研究では瞳孔径と会話参加態度の非積極度との関連性を分析する．

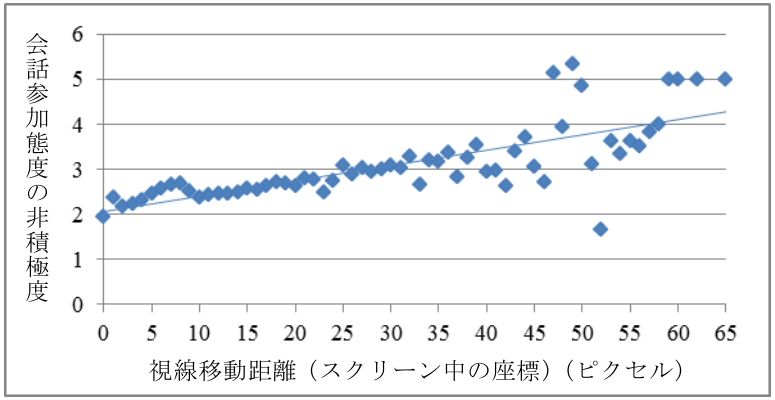

図 6　視線移動距離と会話参加態度の関係
Fig. 6　Relationship between eye movement distance and disengagement score.

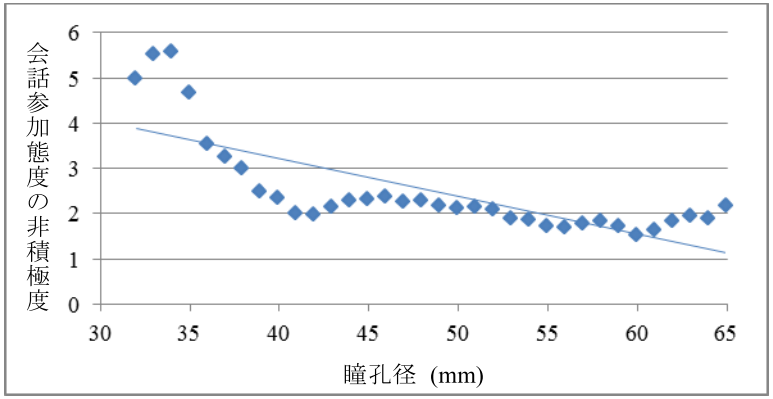

図 7　瞳孔径と会話参加態度の関係
Fig. 7　Relationship between pupil size and disengagement score.

左眼の瞳孔径の大きさに応じて，会話参加態度の平均値がどのように変化するのかを調べた．具体的には瞳孔径を 1 mm 刻みで集計し，会話参加態度の非積極度の平均値を算出し，その結果を**図 7** に示す．グラフの横軸は瞳孔径を示し，縦軸は，会話参加態度を示す．グラフに示されるように，会話参加態度が非積極的であるほど瞳孔径が小さくなり，瞳孔径が大きいときには非積極度が小さくなる（つまり，積極的な態度である）傾向が見られる．また，相関係数を算出した結果 −0.56 であった．以上より，瞳孔径は会話参加態度の推定に有効なパラメータであることが示唆された．

5. 頭部動作の分析

視線の変化は直接的にはアイトラッカにより計測されるべきものであるが，視線の移行が頭部動作をともなう場合もあり，ヘッドトラッカにより計測される頭部姿勢データはアイトラッカの視線データをある程度近似していると考えられる．もしそうであれば，視線計測が良好でない場合にも頭部姿勢データが視線データを補うことにより，より頑健性の高い会話参加態度の推定が可能になる．また，会話関心低下による体全体の姿勢が崩れることにより，頭部姿勢にも変化が現れる可能性も考えられる．

そこで，ヘッドトラッカから得られた頭部の位置（x，y，z）と回転角度（roll，yaw，pitch）の 7 人分のデータを用いて，頭部姿勢と会話参加態度との関連性を分析した．まず，トラッキングデータのノイズを除去するために，30 fps で計測されたデータにおいて，1 秒間（30 フレーム）のウィンドウを 1 フレームずつずらすことにより移動平均を求め，データのスムージングを行った．**図 8** (a-1)，(b-1) は非積極度 0〜10 において x，y，z，roll，yaw，pitch のそれぞれの平均値を求め，プロットした結果である．これらの相関係数を求めたところ，−0.17〜0.16 程度であり，頭部の位置と回転角度のデータは直接的には会話参加態度に関係しないことが分かった．

次に，頭部の位置と回転角度の変化を波としてとらえ，振幅と周波数を算出した．振幅のグラフを図 8 (a-2)，(b-2) に示す．グラフに示されるように，参加態度が非積極的であるほど頭部位置や回転角度の振幅が大きくなっていることが分かる．これらの相関係数は 0.45〜0.59 であり，中程度の相関があるといえる．一方，周波数と会話参加態度の関連をプロットしたところ，図 8 (a-3)，(b-3) に示されるように，明確な相関関係は見られず，相関係数も −0.26〜0.15 と無相関と見なせる範囲にとどまった．

以上の結果から，ヘッドトラッカから得られる頭部姿勢データ，特に頭部の位置や回転角度の変化量を示す振幅の情報は，会話参加態度の推定に有用であることが示唆された．

6. 会話参加態度の推定

6.1 機械学習による推定モデル

前述 4，5 章では，注視対象遷移パターンに加えて，相互注視，注視時間長，注視位置の移動距離の視線情報と，頭部位置と角度の振幅が会話参加態度の非積極性と相関があること

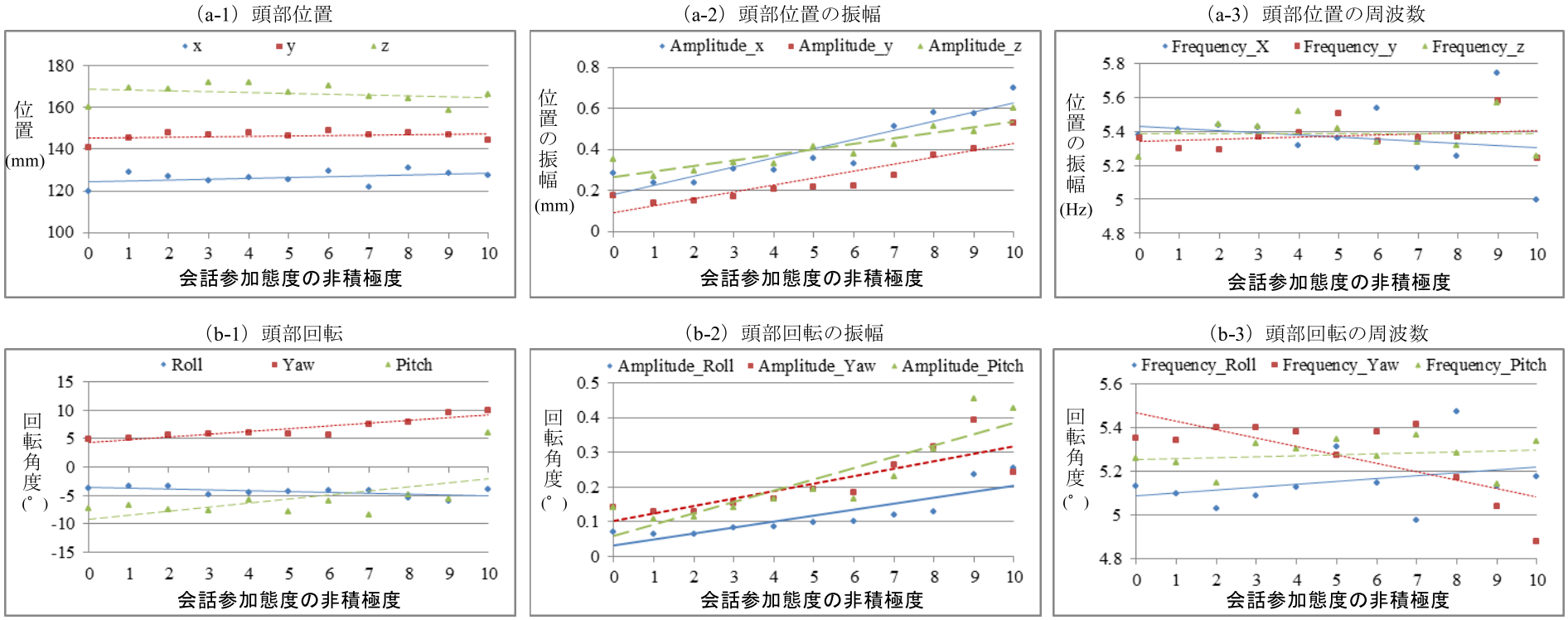

図 8　頭部姿勢と会話参加態度の相関関係
Fig. 8　Relationship between head pose and disengagement score.

を示した．本章では，これらのパラメータが会話参加態度の推定に有用であるかを検証するため，各パラメータを用いて機械学習（SVM: Support Vector Machine）[21] を行い，会話参加態度の推定を行った．使用した機械学習のライブラリおよび各種設定は以下のとおりである．

- ライブラリ：LIBSVM（R のパッケージ e1071 を使用）
- カーネル：RBF（Radial basis function）
- C パラメータ：C = 1.0（デフォルト値）

本機械学習においては，会話参加態度の非積極度を積極/非積極の 2 値に変換し，会話参加態度が非積極的か否かを分類する問題として扱った．具体的には，全会話データの非積極度の平均値である 2.42 を閾値とし，クラスを以下のように定義した．

- 積極的な会話参加態度：会話参加値度の非積極度が 0〜2 のとき
- 非積極的な会話参加態度：会話参加値度の非積極度が 3〜10 のとき

また，特徴量となるパラメータは，以下のとおりである．

- 注視対象遷移パターン：M ラベルを含まない 3-gram の種類（20 種類）
- M ラベルを含む注視対象遷移パターン：M ラベルを含む 3-gram の種類（42 種類）
- 注視時間長：3-gram の各構成要素の時間長，3 つの構成要素の総時間長，総時間長に

対する各構成要素の時間長の比率の 7 つの特徴量[⋆1]

- 注視位置の移動距離：過去 400 msec の注視位置の移動距離
- 瞳孔径：両眼の瞳孔径の平均値
- 頭部動作の振幅：x，y，z 軸に対する頭部位置の振幅，roll，pitch，yaw の回転角の振幅

これらのパラメータを用いて，推定モデルを以下のように設定し，SVM アルゴリズムによる評価を行った．

- 3-gram：注視対象遷移パターンを使用
- 3-gram+M：M ラベルを含む注視対象遷移パターンを使用
- 3-gram+M+Dr：M ラベルを含む注視対象遷移パターンと注視時間長を使用
- 3-gram+M+Ds：M ラベルを含む注視対象遷移パターンと注視位置の移動距離を使用
- 3-gram+M+PS：M ラベルを含む注視対象遷移パターンと瞳孔径を使用
- 3-gram+M+Dr+Ds+PS：M ラベルを含む注視対象遷移パターン，注視時間，注視位置の移動距離，瞳孔径を使用
- Head：頭部動作の振幅を使用
- All：すべてのパラメータを使用

機械学習に用いるデータは，会話実験で取得された全ユーザのデータ中の，すべてのパラメータが取得されている時刻のデータから，サンプリングした 60,000 データ（会話 2,000 秒間相当）である．これらのデータの半数を学習データ，残りの半数をテストデータとした．

6.2 評価結果

機械学習によるモデルの評価結果を**表 1** に示す．各モデルにおいて，積極的な会話参加態度および非積極的な会話参加態度の各クラスについて，適合率，再現率，F 値を算出した．全体的に見ると，積極的な会話参加態度および非積極的な会話参加態度に対して，All モデルは F 値で 0.930，0.753 であり最も性能が高かった．この結果は，使用したすべてのパラメータが会話参加態度の推定に有効的に働いていることを示唆する．

3-gram と 3-gram+M モデルを比べると，非積極的な会話参加態度で F 値 0.130，0.155 と大きな差は見られなかった．これは，今回の会話収集実験で会話エージェントがユーザに注視を行う頻度が少なく，M ラベルを含む 3-gram のデータが全体の 4.5%と少なかったことが理由として考えられる．次に 3-gram+M+Dr モデルは，3-gram+M モデルに比べて，非積極的な会話参加態度の F 値が 0.155 から 0.366 と大幅に向上が見られた．この結果から，注視時間を考慮することが会話参加態度推定の精度向上に大きく寄与することが示された．3-gram+M モデルと 3-gram+M+Ds モデルを比較すると，非積極的な会話参加態度で F 値が 0.155 から 0.217 と向上した．また，3-gram+M モデルと 3-gram+M+PS モデルを比較すると，非積極的な会話参加態度で F 値が 0.155 から 0.238 と向上した．この結果から，注視位置の移動量と瞳孔径の情報も会話参加態度の推定に有用であることが確認された．以上のような，視線パラメータをすべて加味した 3-gram+M+Dr+Ds+PS モデルは，その他の視線パラメータを組み合わせたモデルに比べて，最も性能が高く，非積極な会話参加態度の F 値が 0.631 であった．我々が分析を行った視線情報をすべて考慮することで，より高精度な推定モデルを構築できることが確認された．

次に，Head モデルは，3-gram+M+Dr+Ds+PS モデルには及ばないものの，非積極な会話参加態度の推定で F 値が 0.419 とその他のモデルに比べて高い性能を示したことから，頭部の位置・姿勢の振幅の情報が会話参加態度の推定に有効であることが示された．

表 1　機械学習によるモデルの評価結果
Table 1　Results of evaluation.

Model \ Result	Engagement			Disengagement		
	Precision	Recall	F-measure	Precision	Recall	F-measure
3-gram	0.704	0.964	0.814	0.475	0.075	0.130
3-gram+M	0.750	0.991	0.854	0.597	0.089	0.155
3-gram+M+Dr	0.787	0.979	0.872	0.796	0.237	0.366
3-gram+M+Ds	0.764	0.982	0.859	0.712	0.128	0.217
3-gram+M+PS	0.866	0.975	0.858	0.667	0.145	0.238
3-gram+M+Dr +DS+PS	0.849	0.968	0.904	0.845	0.504	0.631
Head	0.874	0.996	0.931	0.931	0.270	0.419
All	0.887	0.979	0.930	0.913	0.641	0.753

7. 議　　論

評価実験の結果から，視線と頭部動作のすべてのパラメータは会話参加態度の推定に有効

⋆1 4.3 節の分析では，注視時間を 3 つに分類した．しかしながら，機械学習では，特徴量に応じてより最適な分類を行える可能性がある．そのため，注視時間長をそのままデータとして扱うこととした．また，より多くの特徴量を扱うことで性能を上げられる可能性があるため，4.3 節の分析で扱わなかった注視時間に関連する情報についても特徴量に含めた．

であることが示された．しかしながら，これらのパラメータはユーザの心的状況やシステムの環境によって影響を受けることが考えられる．たとえば，瞳孔径はユーザの感情や環境の明るさに大きく影響を受ける．また，そもそもの瞳孔径の大きさの個人差は小さくはない．また，注視位置の移動距離においても，対話相手や注視対象物との位置関係によって大きく変化する．そのため，瞳孔径や注視移動距離といったパラメータを用いて，汎用的な推定モデルを構築するためには，これらの要因に適応的な推定モデルを構築する必要がある．これに対して，注視対象 3-gram や注視時間は上記のような外的な影響を受けにくいと考えられ，視線のみを用いて会話参加態度を推定することを考えた場合，3-gram+Dr モデルは 3-gram+Ds および 3-gram+PS に比べて頑健な推定モデルであると考えられる．また，頭部動作においても，対話相手や注視対象物との位置関係によって頭部運動が大きく変化する可能性があるが，本研究では微小な頭部位置および姿勢変化の振幅に着目したため，影響は受けにくいものと考えられる．

評価実験で最も評価の高かったモデルは，視線と頭部動作をパラメータとしてあわせ持つモデル（All）であったが，アイトラッカによる視線計測成功率は全会話時間の 52.8%であったのに対し，ヘッドトラッカによる頭部データは 100%計測できていた．一般的に頭部データは，視線データに比べてよりロバストに計測可能であると考えられるため，本研究で提案した会話参加態度推定機構を実装することにより，アイトラッカによる眼球計測が不能になった場合でも，頭部動作を用いることでよりロバストに会話参加態度が推定可能になると考えられる．

8. ま と め

本研究では，人と会話エージェントの対話における頑健なユーザの会話参加態度推定方式の確立に向けて，注視パターン，注視時間，注視位置の移動距離，瞳孔径に関する複数の視線情報，および眼球検出によらないものとしてヘッドトラッキングによる頭部姿勢の情報と会話参加態度との関連性を検証し，会話参加態度と相関があることを示した．さらに，パラメータの様々な組合せによる会話参加態度推定モデルを比較した結果，視線と頭部姿勢の両方の情報を用いた推定モデルが最も性能が高く，複数の視線情報および頭部姿勢のパラメータを用いることで，頑健な推定モデルを構築可能であることを示した．我々は，すでに注視対象遷移パターン（3-gram）を用いて，リアルタイムに会話参加態度を推定可能な機構を組み込んだ対話システムの実装を完了しており，本稿で提案した新しい推定方式を組み込むことにより，さらに対話システムの高度化を図る予定である．

謝辞　機械学習による推定モデルの評価に協力いただいた NTT 数原良彦氏に感謝する．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」（課題番号：21013042）によるものである．ここに記して感謝する．

参 考 文 献

1) Sidner, C.L. et al.: Explorations in Engagement for Humans and Robots, *Artificial Intelligence*, Vol.166, No.1-2, pp.140–164 (2005).
2) Argyle, M. and Cook, M.: *Gaze and Mutual Gaze*, Cambridge, Cambridge University Press (1976).
3) Kendon, A.: Some Functions of Gaze Direction in Social Interaction, *Acta Psychologica*, Vol.26, pp.22–63 (1967).
4) Ishii, R. and Nakano, Y.I.: Estimating User's Conversational Engagement Based on Gaze Behaviors, *8th International Conference on Intelligent Virtual Agents* (*IVA'08*), pp.200–207, Springer (2008).
5) 石井　亮，中野有紀子：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，情報処理学会論文誌，Vol.49, No.12, pp.3835–3846 (2008).
6) Nakano, Y.I. and Ishii, R.: Estimating User's Engagement from Eye-gaze Behaviors in Human-Agent Conversations, *2010 International Conference on Intelligent User Interfaces* (*IUI2010*), pp.139–148 (2010).
7) Clark, H.H.: *Using Language*, Cambridge, Cambridge University Press (1996).
8) 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).
9) 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).
10) Anderson, A.H. et al.: The Effects of Face-to-face Communication on the Intelligibility of Speech, *Perception and Psychophysics*, Vol.59, pp.580–592 (1997).
11) Whittaker, S.: Theories and Methods in Mediated Communication, *The Handbook of Discourse Processes*, Graesser, A., Gernsbacher, M. and Goldman, S. (Eds.), pp.243–286, Erlbaum, NJ (2003).
12) Nakano, Y.I. et al.: Towards a Model of Face-to-Face Grounding, *The 41st Annual Meeting of the Association for Computational Linguistics* (*ACL03*), pp.553–561 (2003).
13) Miyauchi, D., Sakurai, A., Nakamura, A. and Kuno, Y.: Bidirectional eye contact for human-robot communication, *IEICE Trans. Information and Systems*, Vol.E88-D, No.11, pp.2509–2516 (2005).

14) Qvarfordt, P. and Zhai, S.: Conversing with the user based on eye-gaze patterns, *Proc. SIGCHI Conference on Human Factors in Computing System, CHI '05*, pp.221–230, ACM Press (2005).
15) Iqbal, S.T., Adamczyk, P.D., Zheng, X.S. and Bailey, B.P.: Towards an Index of Opportunity: Understanding Changes in Mental Workload during Task Execution, *Proc. CHI'05*, Portland, OR, ACM, pp.311–320 (2005).
16) Iqbal, S.T., Zheng, X.S. and Bailey, B.P.: Task-Evoked Pupillary Response to Mental Workload in Human-Computer Interaction, *Proc. CHI'04*, Vienna, ACM, pp.1477–1480 (2004).
17) Eichner, T. et al.: Attentive Presentation Agents, *The 7th International Conference on Intelligent Virtual Agents* (*IVA*), pp.283–295 (2007).
18) 岡　兼司，菅野裕介，佐藤洋一：頭部変形モデルの自動構築をともなう実時間頭部姿勢推定，情報処理学会論文誌：コンピュータビジョンとイメージメディア，Vol.47, No.SIG10 (CVIM 15), pp.185–194 (2006).
19) Kipp, M.: Anvil – A Generic Annotation Tool for Multimodal Dialogue, *The 7th European Conference on Speech Communication and Technology*, pp.1367–1370 (2001).
20) Hess, E.H.: Attitude and Pupil Size, *Scientific American*, Vol.212, pp.46–54 (1965).
21) Vapnik, V.: *The Nature of Statistical Learning Theory*, Springer-Verlag (1995).

(平成 23 年 4 月 12 日受付)
(平成 23 年 9 月 12 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．現在，NTT サイバースペース研究所勤務．2010 年より，京都大学大学院情報学研究科博士後期課程在学．2011 年より，成蹊大学客員研究員．人同士の映像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会話促進技術の研究に従事．映像情報メディア学会会員．

大古　亮太

2010 年成蹊大学理工学部情報科学科卒業．現在，同大学院理工学研究科博士前期課程在学中．視線を用いた会話エージェントの研究に従事．

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）．ACL，ACM，人工知能学会各会員．

西田　豊明（フェロー）

1977 年京都大学工学部卒業．1979 同大学院修士課程修了．1993 年奈良先端科学技術大学院大学教授，1999 年東京大学大学院工学系研究科教授，2001 年東京大学大学院情報理工学系研究科教授を経て，2004 年 4 月京都大学大学院情報学研究科教授，現在に至る．会話情報学，原初知識モデル，社会知のデザインの研究に従事．日本学術会議連携会員（2006 年〜），人工知能学会会長（2010 年〜），国立情報学研究所運営会議委員（2008 年〜）．日本学術振興会学術システム研究センター主任研究員．