ヒューマン・エージェントインタラクションにおける会話関心度推定のためのユーザ視線パターンの分析

石井 亮† 中野 有紀子†

†東京農工大学大学院 工学府 〒184-8588 東京都小金井市中町 2-24-16
E-mail: †nrhc_ryo@hotmail.co.jp; nakano@cc.tuat.ac.jp

あらまし 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，話題を変えるなど，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，ユーザの注視行動に着目して，エージェントとの会話におけるユーザの興味，関心，飽きといった状態と注視行動との間に関連性があることを明らかにする．まず，ユーザの視線・頭部行動の計測データ，会話への関心度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ために行った Wizard-of-Oz 法による実験について報告する．次に，ユーザの注視行動について分析し，対面会話における理想的な注視行動からの逸脱が大きいほど会話への関心度が低下傾向にあることを示す．この結果より，視線データから会話への関心度の推定を行うことの妥当性が確認された．

キーワード 注視行動，会話関心度，会話エージェント

Gaze Pattern Analysis for Estimating the Degree of Conversational Engagement in Human-Agent Interaction

Ryo ISHII† and Yukiko NAKANO‡

†Graduate School of Tokyo University of Agriculture and Technology

2-24-16 Naka-cho, Koganei-si, Tokyo, 184-8588 Japan
E-mail: †nrhc_ryo@hotmail.co.jp; nakano@cc.tuat.ac.jp

Abstract In face-to-face conversations, the speaker is 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 with the user in such an adaptive way, this study focuses on the user's gaze behaviors, and reveals that the user's degree of engagement to user-agent conversations is related to the user's gaze behaviors. First, we report a Wizard-of-Oz experiment that collects user's gaze and head movement, and user's subjective reports and observer's judgment about whether the user is interested in the conversation. Then, our analysis of user's gaze behaviors reveals that the more the user's gaze behaviors are deviated from the ideal gaze patterns, the lower the user's interest in the conversation. This result supports the validity of the idea of estimating the user's degree of conversational engagement by sensing the user's gaze behaviors.

Keyword gaze behaviors, conversational engagement, conversational agent

1. はじめに

会話を円滑に遂行するために，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．例えば，視線行動やうなづ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過程は会話を成立させる上で，基本的，かつ不可欠なプロセスであり，engagement の問題として，人間同士の対面コミュニケーションの研究において，また人対コミュニケーションロボットの研究においても議論されてきた[1]．[1]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，ヘッドトラッカーにより認識されたユーザの頭部運動の情報を用い

て判断している．

本研究では，ユーザの会話参加態度に応じて，適応的に振舞う会話エージェントを目指し，会話参加態度を推定する情報として，さらに詳細なデータであるユーザの注視行動に着目する．まず，Wizard-of-Oz システムによる実験から得られた視線データの分析を行い，会話参加態度を推定するための指標を提案するとともに，これにより会話に積極的に参加している状態とそうでない場合とを識別可能であることを示す．さらに，会話参加態度に関するユーザ自身の主観的判断や観察者による判断と視線行動の指標との間に十分な相関関係があることを確認し，提案指標の妥当性を示す．

2. ユーザ対エージェント会話の収録実験

2.1. 目的

ユーザが会話エージェントシステムとの会話に関心が無くなっていることを分析・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，ユーザと会話エージェントとの対話収録実験をおこなった．大型のスクリーンにはアニメーションキャラクターが投影され，ユーザには，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．スクリーン投影画面を図１に示す．

図１：スクリーン投影画面

2.2. システム概要

ユーザは，図２のように，120 インチリア型スクリーンとパーティションで仕切られた個室スペースに入り，およそ 1.5m 離れたスクリーンに，3000ml DLP プロジェクターから投影された販売員エージェントと対面するような位置に立つ．ユーザが対話に集中できるよう環境を設定した．

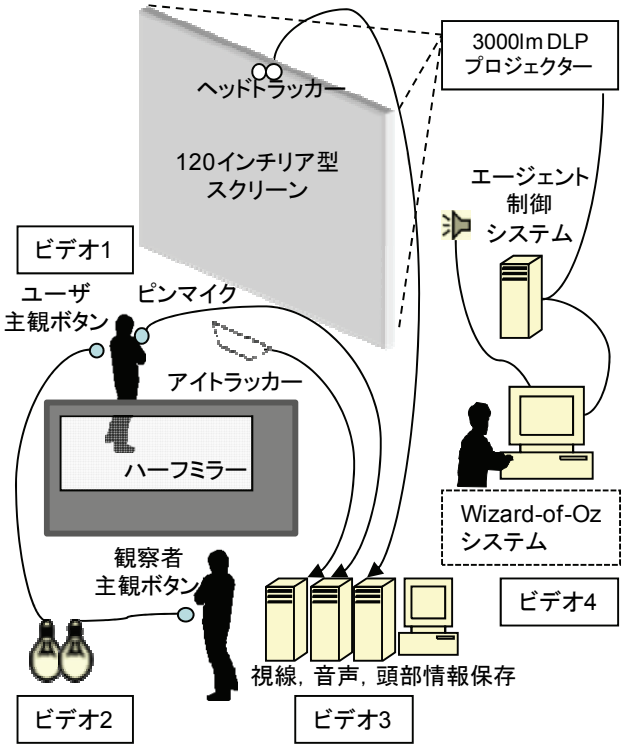

図２：対話収録実験のシステム構成

（1）エージェント生成部

会話エージェントの言語・非言語情報は，自動生成システム CAST[2]を用いて生成した．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，エージェントのアニメーション実行用スクリプトを生成する．会話エージェント描画には Haptek アニメーションシステムを用い，Haptek の動作スケジューリングの操作と音声再生をおこなう VB アプリケーションを Wizard-of-Oz システムとして開発した．また，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．

（2）ユーザ情報の取得

対話中のユーザの様子を収録するためにハンディカムビデオカメラで上半身部分を録画した．また，会話への関心が無くなったことを知らせるボタンをユーザと観察者に持たせ，ボタンが押されるとランプが点灯する装置を作成し，会話中のランプの点灯／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）を通じて，記録した（DigiOnSound5）．

ユーザの視線計測には，非接触型視線計測装置（Tobii）を使用し，スクリーンに投射されたシステム

画面上のユーザの注視位置を 50fps で取得する．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 ㎝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．

また，ユーザの頭部方向の取得には，頭部姿勢推定システム[3]を利用した．このシステムは，パーティクルフィルタにおける仮説の拡散を適応的に制御することにより，ユーザが空間中のある点を注視している場合の推定精度を高く維持すると同時に，ユーザが突発的に動作する場合にも追随性を保つことを可能にしている．図４に頭部姿勢推定システム動作例を示す．

これらの視線，頭部方向データの時間を同期させるため，画面出力されたシステム時間をビデオカメラに収めた．

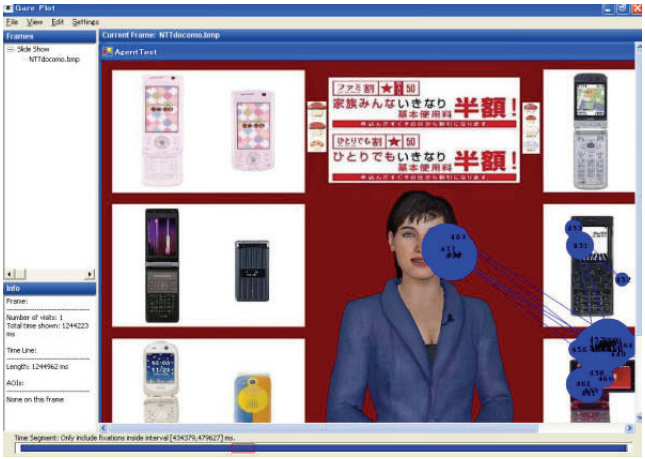

図３：ユーザの注視位置のプロット

図４：頭部姿勢推定システム動作画面

2.3. 実験手続き

システムのユーザとなる被験者（以下，ユーザと呼ぶ）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(以下，観察者と呼ぶ）として，各セッション 1 名，計男性 7 人が参加した（複数回参加者を含む）．

ユーザには，研修中の新人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，どうしても話題を変えたいときのみ，話題を変えるようエージェントに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，会話を続けざるを得ない状況を作った．

また，ユーザには，店頭の 6 つ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ことを課題として課した．対話終了後に，質問紙に回答させ，正解すれば 1000 円の報酬があると伝え，被験者の実験への動機付けを行った．

ユーザの会話関心度低下状態の内観を取得するために，対話中にエージェントの会話内容に関心が無くなったり，話題を変えたいと思った時に，押式のユーザ主観ボタンを押し続けるように指示した．一方，観察者には，対話内容とユーザの挙動をよく集中して観察し，ユーザが会話への関心を失っていると思ったら，ボタンを押し続けるよう指示した．

ユーザへの提示刺激は 6 種類の携帯それぞれについて，説明を全て聞いた場合，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．具体的に，ジェスチャは，各携帯の説明開始時（発話「○○にございますのが，××でおなじみの（機種名）です.」と同時）に一様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した．エージェントの視線行動として，各携帯の説明開始時に携帯を注視し，各携帯の説明中の 10 発話ごとに，発話中に約 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．その他，携帯の説明をしていないときは，エージェントは全てユーザを見るようにした．図５にエージェントが携帯電話を説明している様子を示す．

図５：携帯電話に向きながら説明中のエージェント

3. 収集データ

収録した会話データは，１会話で平均約 16 分から

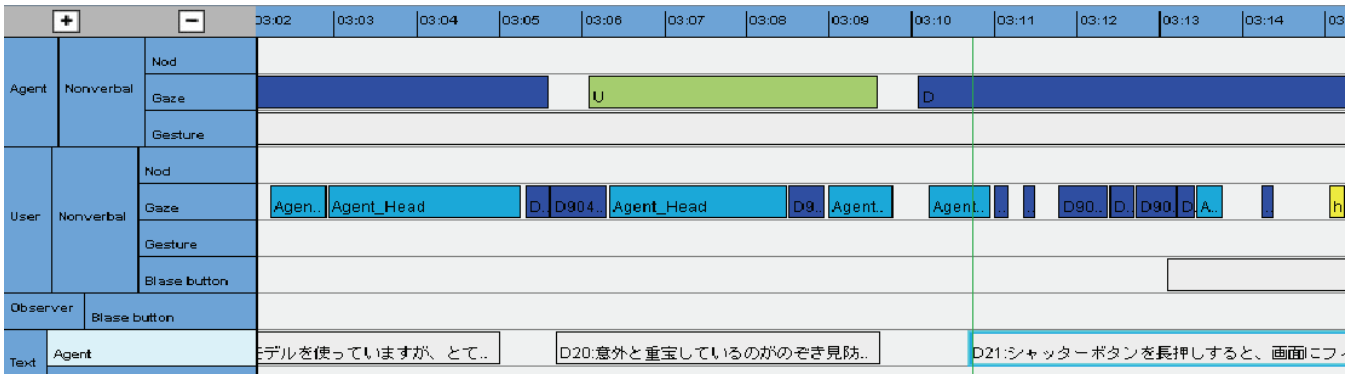

図 6 ：作成した Anvil ファイル

なる．本分析では，アイトラッカーで取得されたユーザの視線データの中から，半径 20pixel 以内の円内に 20ms 以上視線が停留した区間のデータを使用した．また，ユーザならびにエージェントの発話音声はテキストに転記した．全データの発話総数はエージェントが 951 回，ユーザが 61 回である．エージェントのジェスチャならびに視線行動は，発話情報をもとに CAST で生成したスクリプトから算出した．テープに収めた会話関心度主観ランプはアノテーションツール (anvil) を用いて，30frames/秒の精度でデータ化した．最終的に，エージェントとユーザそれぞれの発話テキスト，エージェントのジェスチャ，視線，ユーザの視線，会話関心度主観ランプ，観察者ランプの全ての情報を 1 つの anvil ファイルに統合し，これらの言語・非言語情報の共起関係を視覚的にとらえられるコーパスデータを作成した．図 6 にその一例を示す．

4. 分析 1：ユーザの注視行動と会話関心度

本節では，ユーザの視線パターンに関する基礎的な分析を行い，ユーザの注視行動から会話への関心度を推定することの妥当性を検証する．

人がエージェントに対して人と会話をするのと同じような非言語行動を無意識のうちに行っているのであれば，人の視線行動に関する研究から得られた知見をエージェントとの会話への関心度推定の基礎とすることが可能である．そこで，次の 2 つの知見に注目した．

(1) 話し手が聞き手に視線を向けるのは，聞き手からのフィードバックを促すシグナルであり，その際，聞き手は話し手への視線やうなづき等のフィードバックを返す[4].

(2) 一方，話し手が会話において共有されている対象について説明している際には，多くの時間は両者がその対象物を注視する共同注視が行われている[5].

4.1. 注視行動の全体的特徴

まず，各ユーザの注視行動の全体的な特徴を調べるために，各発話における発話時間長に対するエージェント注視時間 (A-ratio)，説明対象注視時間 (T-ratio)，および説明対象外の物への注視時間の割合 (NT-ratio) を求めた．これらの値についてのユーザ毎，つまり会話毎の平均値を図 7 に示す．

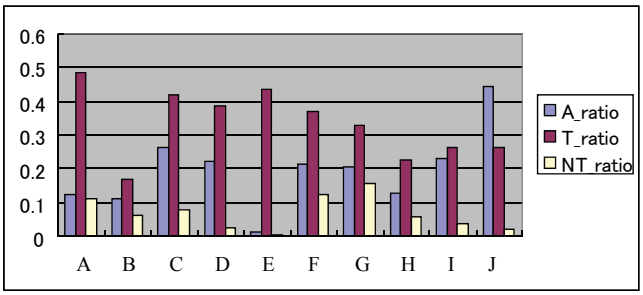

図 7 ：ユーザ毎の注視行動の特徴

本実験では，大部分の時間エージェントは説明対象に向かって発話を行っているため，当然のことながら，10 人中 9 人のユーザにおいて，程度の差はあるが説明対象を注視する時間の方がエージェントを注視する時間よりも多くなっている．しかし，J のユーザのみでそれが逆転しており，(2)の知見に反した注視行動をとっていたと考えられる．一方，E のユーザはほとんどエージェントを見ておらず，エージェントからのフィードバックの促しにも応答してなかったと考えられ，(1)の知見に反した行動であったと考えられる．

4.2. 会話関心度と注視行動逸脱度の関連性

そこで，注視行動のルールを大きく逸脱していなかった残りのユーザについてより詳細な分析を行った．上記(1)，(2)の知見を本実験状況における，ユーザ対エージェントの関係に置き換えて考えると，ユーザがエ

ージェントとの会話に関心を持っている場合，ユーザによる望ましい注視行動は，次のようになる．

(a) エージェントがユーザを注視している場合
　ユーザの主要な注視対象：エージェント
　ユーザの副次的な注視対象：説明対象
(b) エージェントが説明対象を注視している場合
　ユーザの主要な注視対象：説明対象
　ユーザの副次的な注視対象：エージェント

また，どちらの場合においても説明対象以外の物への注視時間 (NT) が増えることは会話への関心が下がっていることを示す情報であると考えられる．

以上に基づき，望ましい注視行動からの逸脱度を示す指標として次のような評価式を定義した．

逸脱度(D)＝(副次的対象注視時間(ST)+対象外注視時間(NT)) / 主要対象注視時間(MT)

MT に対して ST や NT が大きくなるほど逸脱度は大きくなる．但し，各注視時間は，実測値ではなく．発話時間長に対する割合を用いている．

表１：会話関心度主観ボタン押下行動の予測

利用情報	適合率	再現率	F-measure
逸脱度	0.62	0.44	0.51
逸脱度 and エージェント注視の有無	0.62	0.76	0.68

次に，逸脱度がユーザ自身や観察者による会話への関心度に対する主観的な判断とどの程度関連性があるかを調べる．但し，主観ボタンのデータの信頼性を確保するため，ユーザか観察者のどちらかが全くボタンを押さなかったデータは分析対象から除外し，計 7 名分のデータについて分析を行った．

逸脱度が 0.8 以上の場合を関心度低下状態とみなし，主観ボタンが押されていると予測する．その評価結果を表 1 の上段に示す．特に再現率が 0.44 と低く，実際には主観ボタンが押されているにも関わらず，予測できない場合が多い．

4.1 の分析で述べたように，共同注視が主要な状態である場合でも，エージェントを全く見ないことは，一種の逸脱であり，時々話し手の様子を伺うことは重要な視線行動であると考えられる．そこで，エージェントを全く見ていない発話についても，関心度低下状態と判断するという条件を加えた．その結果を表 1 の下段に示す．この条件を加えた結果，適合率 0.62，再現率 0.76, F-measure0.68 という結果が得られた．これらは，注視行動を会話の関心度の推定に利用することの妥当性を示すには十分な結果であり，今後，さらにモデルを詳細化することにより，さらに精度を向上させることができると期待できる．

5. 分析 2：相互注視によるユーザ行動への影響

5.1. 分析結果

ここまで，人対エージェントの対話における視線行動の特徴と，ユーザ主観ボタンあるいは観察者主観ボタンによる会話関心度についての分析を進めてきた．これにより，会話関心度の低下と注視行動の逸脱との間に関連性があることが明らかになったが，ボタンの押下は全くユーザの主観に任されていたため，ユーザによっては，とらえ方が大きく異なっている可能性がある．そこで，会話参加態度に関するユーザ自身による主観的な判断の等質性の検証として，ユーザ主観ボタンを押すタイミングについて分析を行った．

ここでは特に，相互注視行動に着目し，相互注視とユーザ主観ボタン押下との相関関係について検証した．ユーザ主観ボタンが押下されるのは，エージェントが携帯の説明をおこなっている時のみであった為，分析範囲は，エージェントが各携帯の説明をおこなう最初の発話の開始から,最後の発話の終了時点までとした．この範囲において，相互注視中と直後 5 秒間の区間においてユーザ主観ボタンが押された頻度（以下，相互注視中頻度と呼ぶ）と，それ以外の区間で主観ボタンがおされた頻度（以下，相互注視外頻度と呼ぶ）を算出した．尚，ここでは 1 分間での頻度に換算して指標化した．これを各被験者ごとに算出した結果を表 2 に示す．一度もユーザ主観ボタンを押さなかった被験者２人のデータは除外した．

表２：相互注視有無によるユーザ主観ボタン押下頻度（回/分）

被験者	相互注視中頻度	相互注視外頻度
A	1.54	0.23
B	10.28	1.08
C	0	0.98
D	2.87	0.50
E	1.83	1.19
F	4.76	0.44
G	3.73	0.52
I	3.21	0.92

これらの結果に対し，両側検定での対応あり t 検定

をおこなったところ，有意差が認められた．($t(7)=2.46$, $p<.05$)．よって，会話参加態度に関するユーザ自身の主観的判断を行うタイミングは，エージェントとの相互注視中に行われることが多いといえる．

5.2. 考察

相互注視時にユーザの主観ボタンが押下された理由を考察する．Argyle&Cook [6] は，聞き手からの視線は会話への関心や意欲を示すと報告している．これに立脚すると，相互注視は，聞き手が話し手にフィードバックを返すよい機会であると同時に，話し手が聞き手のフィードバックを受け取り，それに適応するストラテジを決定，あるいは実行するタイミングであると考えられる．

従って，本実験において，聞き手であるユーザが，自身の関心度低下や，話題の転化ならびに話者交替の要求を相互注視時に会話エージェントに示すのは，人間同士の対面会話時と同じコミュニケーション方法をエージェントに対しても採用していたことを裏付ける結果である．また，ユーザがこのタイミングで主観ボタンを押す頻度が多いことに関しては，ユーザからのフィードバックを無視してエージェントが説明を継続するため，非言語的シグナルを無視されたユーザがボタンを押下してさらに明確に意思を伝えようとした可能性が考えられる．

6. 議論

本稿では，ユーザ対エージェントの会話における，視線情報に着目し，ユーザの会話への関心度と注視行動との関連性について調べた．その結果，人間同士の対面場面での会話と同様に，エージェントがユーザに視線を向けている場合は，ユーザもエージェントを注視し，両者間での相互注視が成立していることが多く，一方，エージェントが説明対象を見ながら発話を行っている際には，ユーザも説明対象に注意を向ける共同注視の状態になることが多いことが観察された．さらに，これから逸脱する行動が多くなると，会話の関心度が低下することも確認された．

本稿では，会話の関心度をユーザや観察者による主観ボタンにより計測し，視線情報との関連性を調べたが，主観ボタンを押す行動自体を予測することが目的ではない．ユーザの会話参加態度において，何らかの異変がおこっていることを視線データから検出し，それを会話の制御に利用することが最終的な目標である．そのために，今後は，視線行動のパターンやタイミングをより詳細に分析し，ユーザの会話関心度の推定モデルをより精緻化していく予定である．

最後に，本研究の分析では，エージェントとの会話において，ユーザは，人間同士の対面会話の際と類似した行動をとること，つまり人間に対するのと同じようにエージェントに対する行動が決定されていることが確認された．特に，相互注視と主観ボタン押下との間に強い共起関係が確認されたことは，相互注視がユーザの心的状態に大きく影響していることを示唆する結果である．これらの結果は，人間のコミュニケーションモデルを会話エージェントのコミュニケーション機構の基礎とすることの妥当性を示すと同時に，ヒューマンエージェントインタラクションの研究において，このようなユーザによる無意識の行動を解明することの重要性を示唆するものである．

文　献

[1] Sidner, C.L., et al. Where to Look: A Study of Human-Robot Engagement. In Proceedings of *ACM International Conference on Intelligent User Interfaces (IUI)*, pp. 78-84, 2004.

[2] Nakano, Y.I., et al. Converting Text into Agent Animations: Assigning Gestures to Text. In Proceedings of *Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2004), Companion Volume*, pp. 153-156, 2004.

[3] 岡兼司, et al., 適応的拡散制御を伴うパーティクルフィルタを用いた頭部姿勢推定システム. *電子情報通信学会論文誌 D-II*. **J88-D-II**(8): pp. 1601-1613, 2005.

[4] Duncan, S., On the structure of speaker-auditor interaction during speaking turns. *Language in Society*. **3**: pp. 161-180, 1974.

[5] Argyle, M. and J. Graham, The Central Europe Experiment - looking at persons and looking at things. *Journal of Environmental Psychology and Nonverbal Behaviour*. **1**: pp. 6-16, 1977.

[6] Argyle, M. and M. Cook, Gaze and Mutual Gaze. Cambridge: Cambridge University Press, 1976.