

複数人対話における役割に応じた視線の振る舞いの解析とロボットへの実装

Analysis of Role-Based Gaze Behaviors and their Implementation in a Robot in Multiparty Conversation

新谷太健^{1,2*} 石井カルロス寿憲^{2,3} 石黒浩^{1,3}
Taiken Shintani^{1,2} Carlos T. Ishi^{2,3} Hiroshi Ishiguro^{1,3}

¹ 大阪大学

¹ Osaka University

² 理化学研究所

² RIKEN

³ 国際電気通信基礎技術研究所

³ ATR

Abstract: In a multi-person face-to-face dialogue, people naturally gaze according to their roles. The goal of this research is to develop an agent that can control eye movement according to its role in a face-to-face dialogue with multiple users. In this study, we analyze the gaze behaviors in three-party dialogue data accounting for dialogue roles, implement gaze models on a robot based on the analysis results, and conduct evaluation experiments. We show that natural behaviors are achieved by our proposed gaze control system, which accounts for dialogue roles and eyeball movement control.

1 はじめに

近年、遠隔地に存在する他者とコミュニケーションを行うとき、多人数が参加するテレビ会議のようなものだけでなく、アバターを介する対話も増え始めている。モニタ上に表示された擬人化エージェントのようなものから物理的媒体を持つロボットを使用したものまで多種多様なアバターが開発され、それらを遠隔対話のインターフェースとしたアプローチによる研究がなされ、その有用性が明らかにされている。たとえば、小川らはロボットである Telenoid を用いて携帯電話を用いた対話よりも対話エージェント Telenoid を介した対話の方が有益であることを示している [1]。また、遠隔対話のインターフェースとしてのアバターだけでなく、自律型エージェントのアプローチによる研究もなされている。

対話において、言語情報は会話を進める上で相手に発話意図や意味・内容・情報を中心となる要素である。同様に、非言語情報も言語情報を補佐する機能だけでなく、言語・発話だけでは伝えられない意図・欲求を

表出したり、対話を円滑に進めたりと対話において重要となる要素である。非言語情報の重要な要素の1つとして、視線がある。視線の動きは対話を円滑に進める腕最も重要な要因の一つであり、視線に関する研究は様々になされている。Kendon らや Argyle らは視線には「会話の流れを制御する機能、対話相手にフィードバックを与える機能、感情情報を伝える機能、共存感覚を与える機能、対話相手に正もしくは負の評価を与える機能」があると論じている [2][3]。

対話においても、遠隔操作エージェント・自律型エージェント問わず視線は非常に重要な役割を持つ。ロボットやバーチャルエージェントの視線が対話に影響を与えることは様々な研究者によって示されている [4]。しかし、エージェントの視線が対話において重要な機能を果たすことが明らかにされているものの、視線行動の人工的生成手法に関してはまだ十分に研究されていない。特に、眼球検出機器の推定精度がまだ十分であると言えず、実際の対話時における眼球的動きに関する研究・眼球的動作生成手法に関する論文は少ない。

二者対話に関する研究は盛んに進められている。一方、現実の対話において大部分を占める複数人対話に関する研究は少ない。複数人対話では、二者対話以上

*連絡先：大阪大学大学院基礎工学研究科
〒560-0043 大阪府豊中市待兼山町1丁目3
E-mail:shintani.taiken@irl.sys.es.osaka-u.ac.jp

に複雑な情報(今誰が発話権を所持しており、発話しているのか、話題は何かなど)を認識する必要があり、また、振る舞い方もより困難なものとなる。しかし、三者以上の複数人で行われる対話の場にエージェントが居合わせ、適切な振る舞いをする動作生成システムの存在は必要である。そのため、本研究では複数人対話、その中でも三者対話を対象として扱う。

また、タスク指向の対話のように何か目標が存在しそこに向かって進んでいくような対話に関する研究は多いが、井戸端会議のような目的がなく明確に議論の進め方のルールが存在しないにもかかわらず会話の弾むような対話に関する研究も少ない。エージェントが社会進出していくうえで、エージェントは会議のような意味のある対話から何気ない対話まで対応していく必要がある。そのため、本研究では、主に雑談(フリートーク)に着目し、扱っていく。

Sacks らによってターンテイキングと視線に相関があることは既に示されている [5]。また、複数人対話において非常に重要となる要素として今の自分の役割を認識することがあげられる。今自分自身が話者であるのか、話者から主に話しかけられている聞き手であるのか、会話に少ししか関与しないサブリスナーであるのかという役割と視線の振る舞いに相関があると考えられる。そこで、本研究では、実際の対話データに基づいた視線の動き・眼球の動きをターンテイキングに注意しながら、また、役割ごとに分類して解析を行い、視線パターンの行動分析を行った。その後、社会的対話ロボットである CommU に提案である視線生成手法を実装し、その有用性を示す。

2 従来研究

複数人対話における視線の解析・実装に関しては、さまざまな方向から研究されている。

例えば、J.Lee らは 4 人の被験者に対して 7 分の西部劇を演じさせ、それぞれの役割 (Speaker, Addressee, side participant, bystander)、4 者のキャラ間の関係性 (好き、嫌い、どちらでもない)、実際の演じ方から視線の特徴を抽出しようとした [6]。

また、中野らは 1 人のロボットと 3 人の人による 4 者対話を行わせ、対話における支配度と対話者の役割に視線が相関するのではないかと考え、解析・実装を行っている [7]。

前述したように、ターンテイキングと視線の動きに相関があることは明らかにされているが、例えば、石井らは 4 者対話を解析し、IPU 終了時における視線パターンからターンテイキングの推定を行おうとした [8][9]。

本研究では、対話内容に目的がなく、また、エージェントが司会進行するような対話でない、対話者全員が

対話参加者となり、何気ない会話をエージェントも一人の対話者に過ぎない雑談の場において自然な振る舞いをするエージェントの解析・実装を目的としている。次に、本研究で扱う三者対話に関する研究について述べる。三者対話に関しては主に 2 種類の状況が考えられている。一つは三者で対話を行うものであり、もう一つは二者対話に陪席者を用意した対話である。前者に関する視線の研究として、武川らは発話交替時のスピーカーではない聞き手二人に着目し、聞き手同士の視線の配分・スピーカーへの視線量と次の話者に関する関係を解析した [10]。榎本らは三者対話における参与役割の交代にかかわる非言語行動の分析を行っている [11]。三者対話の会話分析を行った結果、現話者は次話者を見ていることが多く、次話者・非話者は話し手を見ていることが多く、現話者の役割から、会話参与者のうち視線を合わせることがその参与者を次話者に選択する手段とみなせるが、次話者の立場からは現話者を見ることにより発話が増えたと解釈が可能であると主張している。また、後者の研究としては、有本らは陪席ロボットと対話ロボットの二つのロボットを用意し、この両者のロボットがアイコンタクトを行うことで対話ロボットと会話をする被験者からのロボットへの印象・社会的存在感が上がることを示した [12]。酒井らは 2 つのロボット議論を行わせ、その議論を見て参加した気になっている陪席者である被験者から対話に参加したように見える陪席者の視線モデルを作成した [13]。Mutlu らは両者の場合において、話者の視線配分がどのようになっているのかを解析し実際にロボットに実装している [14]。

本研究では、主に三者対話を扱うことを考えている。そのため、陪席者の役割となる人は存在はせず、すべての参与者に発話権は存在していると考え、陪席者とは異なるサブリスナーという考え方を採用している。役割ごとに視線モデルは異なる。しかし、対話における陪席者でなくサブリスナーとなる人の視線、特に発話中のサブリスナーの視線の解析・実装を行った研究は不十分であったと考えられるため、本研究では収録されたデータセットからサブリスナーの役割となる人の視線配分についての解析を行い実装した。また、収録したデータセットから実際の対話における視線の動きに関して顔よりも目玉の動きの方が支配的であったが、実際に三者対話における黒目の位置の配分について解析・実装を行っている研究がなかったため、瞳の位置についても収録されたデータセットから解析・実装を行った。

3 データ解析 (ラベル付け)

3.1 データセット

データセット本研究で使用する三者対話のデータセットについて述べる。1 グループ 3 人 6 グループ、男女 9 人の実験参加者が対話を行った。対話の内容はフリートークであり、各々自由に 20 分から 30 分間談笑しあった。各被験者は三角形を描くように椅子に座って話を行い、各被験者の前にカメラが設置され真正面から顔・体の動きが撮影されている。また、各被験者ごとにヘッドセットマイクが装着されており、各参加者の音声データ・議論全体の動画が収集されている。また、録画・音声データから対話の書きおこしを作成した。

本研究では、視線方向・眼球の方向・発話権の有無・ターンテイキング・役割が視線の動きと非常に強い相関があると考えた。そのため、アノテーターの方に前述した対話データのラベル付けを行ってもらい、このラベルを下に解析を行った。また、ラベル付けのルールの説明を行う。

3.2 データの分析

このセクションでは、実際にどのようなラベルを作成・付与し解析を行ったかについて述べていく。

3.2.1 視線方向

被験者は図??に示すように三角形を描くように斜めに座り対話を行っている。被験者がどの被験者を見ているのかラベル付けを行う。ラベルは被験者番号 2 つとそれ以外の 3 つである。この時、顔の角度だけでなく眼球の方向も含めて視線とし、この 2 つの要素から総合的に判断された方向を視線の方向としてアノテーターの主観により決定している。実際にラベル付けをしている様子を図 1 に示す。例えば、A という ID を持つ人は $B \rightarrow C \rightarrow env \rightarrow B \rightarrow env$ と視線が遷移している。ここで、env というラベルは B も C も見ていない状態、つまり視線を逸らした状態を意味する。

3.2.2 眼球の方向

人を見る動作は、ラベルと顔・瞳の位置にそこまで大きなずれは生じない。しかし、視線を逸らすという動きにおいては、視線・特に瞳の位置は非常に重要な要素となる。だが、瞳の位置検出を自動でやることは精度が低く困難であり、眼球の方向を自動で検出する精度は現状低い。そのため、眼球の方向も上下左右斜め真ん中の 9 方向でラベル付けを行う。特に、被験者

を見ていない動き、つまり視線逸らしを行うとき顔の角度だけでなく眼球の方向は非常に重要な要素となる。実際のラベルの様子を図 1 に示す。視線を逸らすラベルの時、付加情報として、瞳はどこを向いていたのかをラベル付けしている。

3.2.3 発話権の有無・ターンテイキング

図 2 に実際の対話の様子を示す。

まず、ターンを取った時、“Turn Taking” というラベル付けを行っている。これにより、実際にターンが交替したタイミングを知ることができる。“Turn Talking” ラベルはアノテーターの方がターンが交替した・話者が発話権を取ったと判断した発話の開始時に付与している。そのため、オーバーラップや相槌・フィラーのような発音に影響されず、また、これにより IPU (Inter-Pausal Unit) を用いなくても正しいターンテイキングのタイミングを知ることができる。

また、ターンテイキング開始時がわかっているため、ターンが切り替わった瞬間の前後 1 秒ずつ計 2 秒間 (図 2 の赤丸で示す部分) における視線ラベルの解析を行う。

3.2.4 役割

本研究では、対話における役割が重要であると考えている。本研究では以下のように役割を定義した。

- **Speaker (Sp)** 発話権を持っている人/発話権を持っており、話している人・発話をしている人
- **MainListener (ML)** 上記のスピーカーが主に話しかけている対象
- **SubListener (SL)** Sp でも ML でもない聞き手の人

本研究ではスピーカー・メインリスナー・サブリスナーをターンテイキング時に着目し以下の手法により分別して解析を行った。図 2 において、このターンテイキングで発話権を取得した人 (図 2 では C に当たる人) を話者 (Sp)、発話権を譲渡したまたは取られた人 (図 2 では A に当たる人) をメインリスナー (ML)、このターンの交替に関与しなかった人 (図 2 では B に当たる人) をサブリスナー (SL) とする。

- **Speaker (Sp)** ターンテイキング時 (発話交替) に発話権を取った人
- **MainListener (ML)** ひとつ前にターン (発話権) を持っていて、スピーカーに発話権を譲った人
- **SubListener (SL)** ターンのやり取りに関与しなかった人

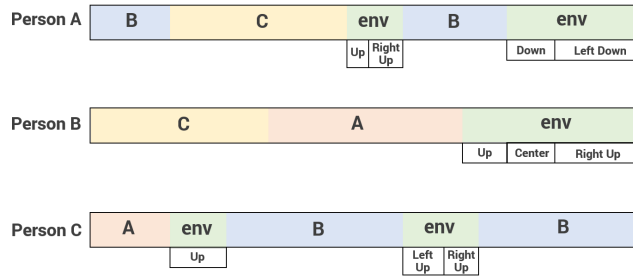


図 1: Gaze Label

3.3 解析結果

本研究ではターンテイキングに注意して、ターンテイキング時とそれ以外に分けて視線パターンの考察を行った。

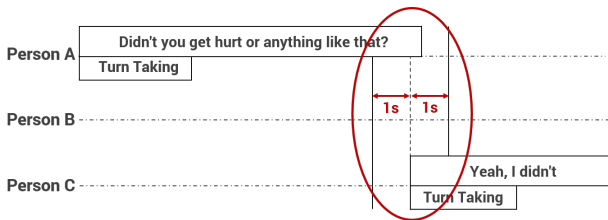


図 2: ターンテイキング

3.3.1 ターンテイキング時の解析

ターン交替時の各役割ごとの視線の動きの割合の推移について図 3 に示す。横軸は時間 (単位は秒) であり、時間 $t=0.0$ の時に話者がターンを取り、話し始める。また縦軸は時間 t においてどこを見ているのかの統計を取りその割合 (0.0-1.0) を表している。例えば図 3 の (a) の図の横軸 $t=0.0s$ の縦軸の値を見ると ML, SL, env それぞれ 0.37, 0.19, 0.44 であるが、これは話し始めたタイミングでの話者の視線の先の総計のうち 0.37 がメインリスナーを、0.19 がサブリスナーを、そして残る 0.44 視線を逸らしていたことを表している。0.1 秒ごとに算出しており、図 3 では、話者がターンを取り話し始めたタイミングの前後 1 秒ずつ合計 2 秒の区間における視線のやり場の割合が時間経過とともに遷移する様子を表している。

スピーカー

図 3 の一番左の図 (a) にターン交替時における話者の視線の割合の推移を示す。話者はターンを取る 1 秒前はメインリスナー (ML), つまり、前の話者を見ている割合が高い。しかし、ターンを取り、発話を開始する地点 ($t=0.0s$) に向かうにつれて、視線を外す割合が高くなっている。そして、話し始めて 0.1 秒経過したあた

りで視線を逸らす割合はピークを迎えた後、視線はメインリスナー (ML) を向く割合と視線を逸らす割合がほぼ同等に高くなりターン交替における視線の遷移を終える。この結果より、人は最初前の話者を見て、話を始める際に視線を逸らし、以降は視線を逸らしたまま、あるいは前の話者であるメインリスナーを見る傾向にあると考えられる。

メインリスナー

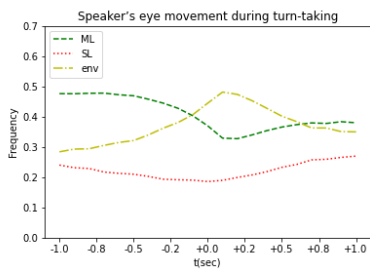
図 3 の真ん中の図 (b) に、0.0 秒に話者が話し始めるがその直前までターンを持っていたメインリスナーの視線の割合の遷移を示す。このターンテイキング時に、メインリスナーがターンを渡した、取られた、もしくは自然に交替した等、状況は様々に考えられるが全体として話し始める前から次話者がわかっている或いは決めており、1 秒前からターンを取られた後 1 秒後までのターンテイキングの区間において話者の方を見続ける傾向にあると考えられる。

サブリスナー

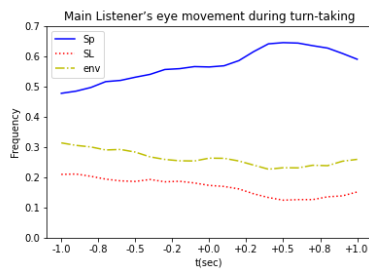
最後に、ターンを渡しターンを取るターン交替のやり取りに関与せず、その様子を見ていたサブリスナーの視線の割合の遷移を図 3 の右の図 (c) に示す。横軸 $t=-1.0s$ の時、即ち、話者が発言する 1 秒前までは前話者であるメインリスナー (ML), 次話者であるスピーカー (Sp) を見る割合が同様に高く、そこから $t=1.0$ つまり話者が発話を開始し 1 秒経過した地点に向かうにつれて話者を見る割合が高くなっている。そのため、サブリスナーは 1 秒前まではメインリスナー (前話者) の方もしくは次話者を推測しスピーカー (Sp) の方に視線を向け、そこから次の話者へと視線を動かす傾向にあると考えられる。

3.3.2 ターンテイキング外の解析

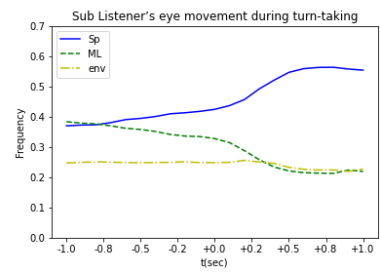
では、ターンテイキング時以外での視線の動きの解析を行う。前後 2 秒ずつ合計 4 秒はターンテイキングからの影響を受けているものと考えられるため、この 4 秒を省いて、また、短すぎる発話を省いて解析を行った。



(a) 話者の視線の動きの遷移の割合

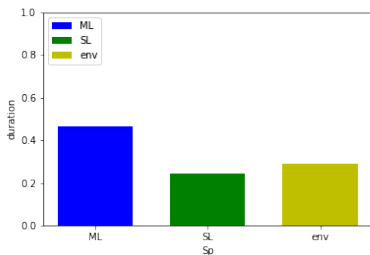


(b) メインリスナーの視線の動きの遷移の割合

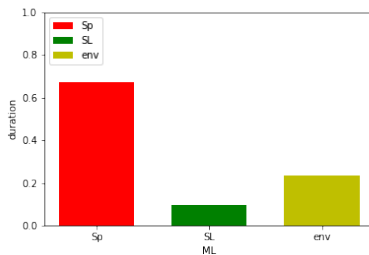


(c) サブリスナーの視線の動きの遷移の割合

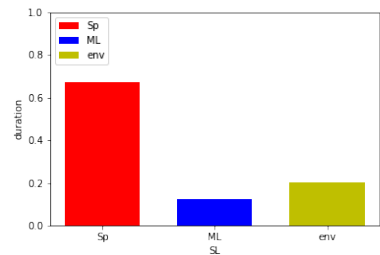
図 3: ターン交替時の各役割の視線の動き



(a) 話者の発話中の視線の時間の割合



(b) メインリスナーの発話中の視線の時間の割合



(c) サブリスナーの発話中の視線の時間の割合

図 4: 発話中における各役割の視線を向ける時間の割合

3.4 発話中の視線の動きの解析

次にターンテイキングでない区間、つまり発話の冒頭部分、末端部分を除いた話者が発話をしている区間における視線の解析を行う。図4に各役割ごとの発話中における視線の時間の割合を示す。横軸は視線の方向を現しており、縦軸は時間の割合を示す。ここで、ターンテイキングを除くため、4秒以上の発話を抽出し、前後1秒ずつ合計2秒間を取り除いている。また、冒頭部分、末端部分のラベルに関して、発話区間外と一続きでかつ、区間内では視線方向の持続時間が1.0秒以内、区間外で視線の持続時間が1.0秒以上の時、ターンテイキング要素と強く結びついていると考え取り除いている。

表1に図4で示した発話中の視線の割合の具体的な数字を示している。

図5に発話中における視線を逸らした回数の散布図を示す。横軸は発話中の時間(先ほど選別したものを使用)を示しており、縦軸はヒトから視線を逸らした回数を示している。また、その時の近似直線も表している。近似直線の回帰係数 r だが図(a),(b),(c)それぞれ0.83,0.82,0.67である。図5及び表1より、役割に依らず視線を逸らす感覚が変わらないことがわかる。

話者

まずは発話中の話者の視線の解析を行う。図4の一番左の図(a)、表1より、話者は発話中に前の話者であ

るメインリスナー (ML) を見る割合が少し高いが、全体としてバランスよく視線の配分をしていることがわかる。

メインリスナー

次に、発話中のメインリスナーの視線の解析を行う。図4の真ん中の図(b)、および表1より、メインリスナーは7割近く話者の方向を向いている。

サブリスナー

発話中のサブリスナーの視線についてもメインリスナー同様、図4の一番右の図(c)、および表1より、7割近く話者の方向を向いている。

3.4.1 視線逸らしの解析

では、次に視線逸らしにおける解析を試みる。ヒトを見るとき視線の動きは非常に単純であり、違和感のないように顔を向けばよい。しかし、視線を逸らすときは単純な動きだけでは不十分である。実際に収録されたデータを見ると眼球だけ移動させて視線をそらしている動きがみられた。そのため、視線を逸らすときの動きの解析として、今眼球がどこを向いているのかに特に着目した。また、視線を逸らすときにどれだけの時間逸らしているのかも分布にして解析を行った。

視線逸らし時における瞳の配分

次に、視線を逸らす動きをした時の継続時間のヒス

表 1: 発話中の視線の割合と逸らす時間

Role	Speaker	Main Listener	Sub Listener	Look away	Interval(look away)(s)
Speaker	0.0	0.47	0.24	0.29	5.03
Main Listener	0.67	0.0	0.10	0.23	6.05
Sub Listener	0.67	0.13	0.0	0.2	6.90

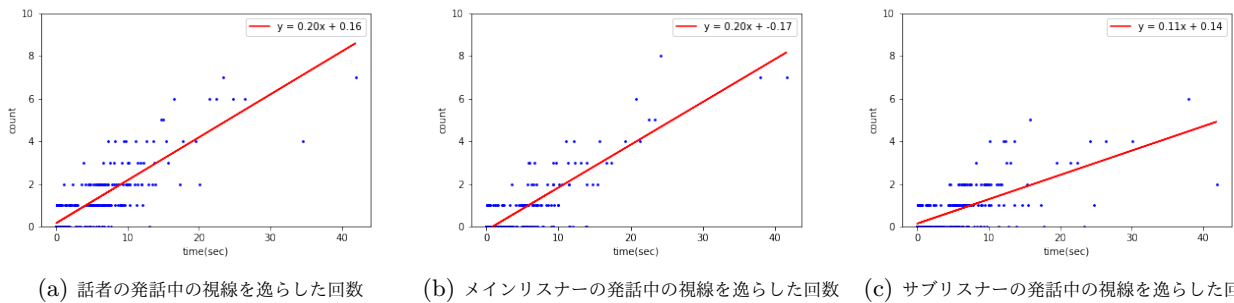


図 5: 発話中における各役割の視線を逸らした回数

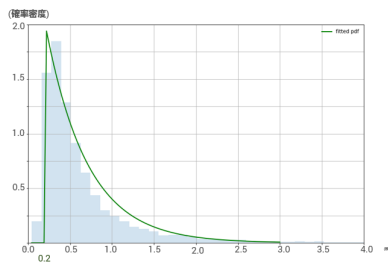


図 6: 視線を逸らすときの時間のヒストグラム

トグラムを図 6 に示す. 指数分布

$$p(x; \mu, \lambda) = \lambda \mathbb{1}_{x \geq \mu} \exp(-\lambda(x - \mu)) \quad (1)$$

で近似される. ここで, $\mu = 0.2, \lambda = 0.63$ である. なお, 0.2s 以下のサンプルは実際のエージェントへの実装を考えたとき, 非現実的であると考え, 0.2s 以上のものに限った. 平均値は 0.83s, 中央値は 0.55s である.

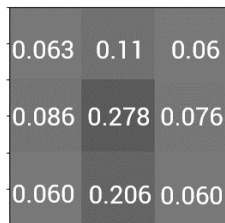


図 7: 視線を逸らすときの瞳の位置の割合

視線逸らし時における瞳を向ける時間

次に, 視線を逸らす動作をしたときにおける黒目の位置についての解析結果を図 7 に示す. 黒目が上下左右斜め中央の 9 か所のうち, それぞれの時間の合計がわかっているため, そこから相対度数の計算を行った. 傾向としては, 中央・ましたに瞳を動かして視線を逸らす傾向強く, 残りの 7 か所は同割の割合でまばらに見ていることがわかる.

4 実装

解析した内容を視線制御システムとして実装を行う. 瞳の動きは Pejsa らの提案した手法を参考に行う [15].

4.1 ターンテイキング時の実装

次に, ターン交替時におけるロボットへの実装方法について記述する. 図 3 に示すように, 各役割ごとにターン交替時における視線の動きには傾向がある. そのため, 自分の今の役割に応じたターンテイキングの視線の動きを生成し実装する必要がある.

本研究では, ターンテイキングの 2 秒の区間を -1.0s - 0.3s, -0.3s 0.3s, 0.3s 1.0s の 3 つの区間に分割し, それぞれの区間における特徴的な所の割合から, その区間ごとにどこを見るべきかを抽出, 算出し, ターンテイキングの 2 秒間で 2 回の遷移, 合計 3 か所を見る実装を行う.

スピーカー

役割がスピーカーの時の視線の割合の推移が図 3 の (a) に示されている. この図より, -1.0s -0.3s の区間

において、-1.0sの時に特徴があると考え、この時のML,SL,envを見る割合がそれぞれ、0.48,0.24,0.28であった。この割合を確率とし、一様分布からどこを見るかを決める。次に、-0.3s 0.3sの区間において、0.1sの時に特徴がある。この時のML,SL,envを見る割合はそれぞれ、0.33,0.19,0.48である。最後に、0.3s 1.0sの区間における特徴は1.0sに見られる。この時のML,SL,envを見る割合はそれぞれ、0.38,0.26,0.36である。

メインリスナー

次に、役割がメインリスナーの時のターンテイキング時の視線の動きの実装について考える。この時の視線の割合の推移は図3の(b)に示している。この図より、-1.0s-0.3sの区間において、-1.0sの時に特徴がみられ、この時のSp,SL,envを見る割合はそれぞれ、0.48,0.21,0.31であった。次に、-0.3s 0.3sの区間において、0.0sの時に特徴がみられ、この時のSp,SL,envを見る割合はそれぞれ、0.56,0.17,0.27であり、また、0.3s 1.0sの区間においては0.5sの時に特徴がみられ、その時のSP,SL,envの割合はそれぞれ0.65,0.12,0.23であった。

サブリスナー

最後に、役割がサブリスナーの時のターンテイキング時の視線の動きの実装について考える。この時の視線の割合の変化の推移は図3の(c)に示している。-1.0s-0.3sの区間において、-1.0sの時に特徴がみられ、この時のSp,ML,envを見る割合はそれぞれ、0.37,0.38,0.25であった。また、-0.3s 0.3sの区間において、0.2sの時に特徴がみられ、この時のSp,ML,envを見る割合はそれぞれ0.46,0.28,0.26であり、0.3s 1.0sにおける特徴は0.8sの時に見られ、Sp,ML,envを見る割合は0.57,0.21,0.22であった。

ターンテイキング時の実装

以上の3つの役割における特徴を用いて各区間における確率より、一様分布から抽出した値から視線を向けるべき方向を算出し、視線の動きを作成する。

4.2 ターンテイキング時以外の実装

図4に示すような時間の割合を維持しつつ、また、各場所を見る時間も実際のデータに近くなるように発話中の視線の動きの実装を行う。今、誰を(或いは視線を逸らす)どれだけの時間継続して見ていたのかがわかる視線ラベルデータが大量にある。エージェントの役割ごとにラベルの数がわかり、そこから割合が決まる。次に割合を確率と考えどの役割の人を見るのかを決める。見る対象が決まったら、図6のように視線を向ける継続時間のヒストグラムがわかっているため、そこから見る時間を抽出する。

4.3 視線逸らし時の動き

実際に視線を逸らす動作をロボットのCommUに実装した様子を図8に示す。左右における、上中下それぞれの場所に瞳を動かした様子である。画像8から目玉の位置が印象に大きな影響を与えそうなのがわかる。

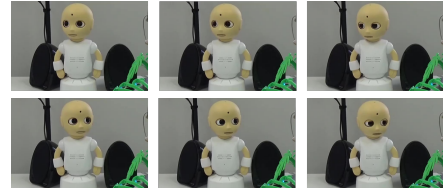


図 8: 実験の様子

5 実験

本研究では、役割ごとの視線パターン・視線を逸らすときの瞳の動きが対話における重要な要素であると考えている。そのため、以下の4つの条件で比較することにした。

同割合-頭部モデル

ベースラインとして、話者の時、2人の対話者を同じ割合で自然配分するモデルを使用した。2人の対話者とそれ以外のところを見る割合であるが、先行研究[14]より、人・人・それ以外を見る割合をおおよそ3:3:4になるように設定した[14]。また、視線を向ける場所であるがこれも同論文より、人の顔ないし2人の対話者の間に角度4度を分散とした2次元ガウス分布を適用して得られた場所を見るようにした。聞き手の時は違和感のないように話者の方向を向くようにした。

同割合-頭部-眼球モデル

眼球の動きが大切であると考えているため、前述の同割合モデルで動かす部位を頭だけでなく眼球も追加し共に動かすモデルとした。

提案-頭部-眼球モデル

4節に示した役割に基づいた視線制御システム・視線逸らしシステムの実装を行った提案モデルである。

提案-頭部モデル

前述の提案モデルのうち、眼球の動作をなくし、頭部動作のみを行う。

5.1 印象評価

ヒトが自然にこなしている役割の理解・視線逸らしを実装することで一番期待したい効果としてロボットが自然らしいかどうかが大切であるため、「全体的にロボットの振る舞いは自然に感じましたか」という質問項目を設けた。

男女 30 人 (平均 32.1 歳, 分散 10.3) に対して 4 つの動画を視聴していただきそれぞれに対して印象評価をしてもらった。

本実験では, 実際に収録した 3 人で行われた対話の中の 1 人の会話音声ロボットに搭載したビデオを視聴し, その時の視線の動き方について評価してもらった。実際には図 8 で示すような動画を見ていただき, 2 名の対話者は, ロボットの斜め左側と斜め右側にいて, 左側の人の声は左耳に, 右側の話者の声は右耳に, ロボットの声は両耳に聞こえるようにした。そして 4 つの手法についてそれぞれ個別に印象評価をして頂いた。

実験手順は次のとおりである。まずは, 順序効果を減らすために 4 つの動画をの提示順序をランダム化した。次に, 各 4 つの手法の動画について個別に 7 点スケール (1 : とても不自然, 4 : どちらともいえない, 7 : とても自然) で印象評価していただいた。このセットを 3 つの対話区間 (各対話区間の長さは 1 分程度) に対し合計 12 個の動画について印象評価していただいた。

5.2 評価結果

3 つの対話区間のうち, 2 つでは条件間に有意差がみられなかったため, 有意差がみられた 1 対話区間の結果について報告する。

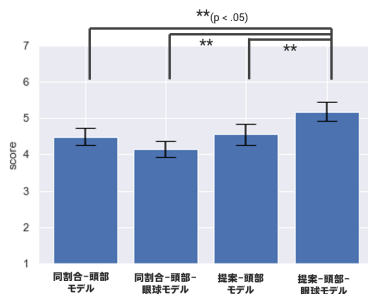


図 9: 「全体的にロボットの振る舞いは自然に感じましたか」の結果

「全体的にロボットの振る舞いは自然に感じましたか」への評価結果を図 9 に示す。各モデルにおいて結果の平均値, 標準誤差, ライアン法に基づいた多重比較の結果を算出した。ライアン法の結果, 同割合-頭部と提案-頭部-眼球モデル間に p 値 $0.020 (\leq .05)$, 同割合-頭部-眼球と提案-頭部-眼球モデル間に p 値 $0.001 (\leq .05)$, 提案-頭部と提案-頭部-眼球モデル間に p 値 $0.008 (\leq .05)$ とそれぞれ有意差が算出された。この結果から, 視線逸らしモデル・役割に応じた視線制御モデル及び頭部・眼球を共に動かすモデルがロボットの振る舞いが最も自然であることが示された。しかし, その他の対話区間では, 有意差はみられなかったことから, どのよう

な条件において, 提案法が効果的であるのかを今後詳しく調べる必要がある。

6 考察

評価実験の結果から, 視線逸らしと役割に応じた視線配分を実装し, 頭部動作だけでなく眼球も動かしたモデルがロボットの振る舞いを自然に見せることに有効であることが示された。しかしながら, 本実験では視線逸らしと役割に応じた視線配分の 2 要因のうちどちらが強く影響を与えたのか, あるいは, 両者がそろって初めて自然な振る舞いとなるのかを明らかにすることはできなかった。そのため, より細分化して再実験を行う必要がある。また, ロボットの振る舞いは対話状況や対話者間の関係・対話者の心的状態に大きく影響を受けることが考えられる。対話の役割ごとのモデルの効果は今後詳細に調べる必要があり, 今後の予定とする。

7 まとめ

本研究では, 対話役割に応じた視線パターンと視線を逸らすモデルがロボットの自然な振る舞いに影響を与えることを示した。現在, 人間酷似型ロボットであるアンドロイド ERICA にも実装を試みており, より人に近いロボットでも自然に振る舞えるようなモデルへと改良している。また, 現在のモデルは乱数に依存しすぎているため, 対話内容などから考えて明らかに不適切な動きをしないような条件付きモデルを作成することを今後の課題とする。

謝辞

本研究は, JSPS 科研費 JP20H05576 の助成を受けたものである。データ分析にご協力いただいた村瀬妙子氏, 中西京子氏, 中山祐佳氏に感謝する。

参考文献

- [1] K. Ogawa, S. Nishio, K. Koda, K. Taura, T. Minato, C. T. Ishii, and H. Ishiguro. Telenoid: Telepresence android for communication. In *ACM SIGGRAPH 2011 Emerging Technologies*, SIGGRAPH '11, New York, NY, USA, 2011. Association for Computing Machinery.

- [2] A. Kendon. Some functions of gaze-direction in social interaction. *Acta psychologica*, Vol. 26, pp. 22–63, 1967.
- [3] M. Argyle and M. Cook. Gaze and mutual gaze. 1976.
- [4] M. Shimada, Y. Yoshikawa, M. Asada, N. Saiwaki, and H. Ishiguro. Effects of observing eye contact between a robot and another person. *International Journal of Social Robotics*, Vol. 3, pp. 143–154, 2011.
- [5] H. Sacks, E. A. Schegloff, and G. Jefferson. A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction*, pp. 7–55. Elsevier, 1978.
- [6] J. Lee and S. Marsella. Modeling side participants and bystanders: The importance of being a laugh track. In *International Workshop on Intelligent Virtual Agents*, pp. 240–247. Springer, 2011.
- [7] Y. I. Nakano, T. Yoshino, M. Yatsushiro, and Y. Takase. Generating robot gaze on the basis of participation roles and dominance estimation in multiparty interaction. *ACM Trans. Interact. Intell. Syst.*, Vol. 5, No. 4, December 2015.
- [8] R. Ishii, K. Otsuka, S. Kumano, R. Higashinaka, and J. Tomita. Analyzing gaze behavior and dialogue act during turn-taking for estimating empathy skill level. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, New York, NY, USA, 2018. Association for Computing Machinery.
- [9] R. Ishii, K. Otsuka, S. Kumano, M. Matsuda, and J. Yamato. Predicting next speaker and timing from gaze transition patterns in multi-party meetings. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, New York, NY, USA, 2013. Association for Computing Machinery.
- [10] 徳永弘子, 湯浅将英, 武川直樹. 3人会話における発話交替時の視線行動分析: 聞き手の立場から見た発話・非発話の戦略. 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. 106, No. 268, pp. 23–28, sep 2006.
- [11] 榎本美香, 伝康晴. 3人会話における参与役割の交替に関わる非言語行動の分析 (テーマ:一般). 言語・音声理解と対話処理研究会, Vol. 38, pp. 25–30, jul 2003.
- [12] T. Arimoto, Y. Yoshikawa, and H. Ishiguro. Multiple-robot conversational patterns for concealing incoherent responses. *International Journal of Social Robotics*, Vol. 10, No. 5, pp. 583–593, 2018.
- [13] K. Sakai, F. Dalla Libera, Y. Yoshikawa, and H. Ishiguro. Generation of bystander robot actions based on analysis of relative probability of human actions. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 21, No. 4, pp. 686–696, 2017.
- [14] B. Mutlu, T. Kanda, J. Forlizzi, J. Hodgins, and H. Ishiguro. Conversational gaze mechanisms for humanlike robots. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, Vol. 1, No. 2, pp. 1–33, 2012.
- [15] T. Pejisa, S. Andrist, M. Gleicher, and B. Mutlu. Gaze and attention management for embodied conversational agents. *ACM Trans. Interact. Intell. Syst.*, Vol. 5, No. 1, March 2015.