



視線を逸らすアバタは聞き上手： 発話前行動としての視線移動が与える印象の調査

杉田修建¹⁾, 福地健太郎¹⁾

1) 明治大学 先端数理科学研究科 (〒164-8525 東京都文中野区中野 4-21-1, cs222023@meiji.ac.jp)

概要：対話エージェントの表情変化による非言語情報の表出による，エージェントが対話相手の話を知的に聞いているような印象を与える手法を調査した．具体的には，プロのインタビュアーによる対話の様子を分析し特徴的な仕草として発話交替時に対話者から視線を逸らす行動が抽出された．この行動を CG キャラクタに実装し対話実験で印象評価を行った．その結果視線移動がアバタを聞き上手であるという印象形成に有効であることがわかった．

キーワード：非言語行動，発話交替，発話前行動

1. 序論

CG キャラクタが人間とリアルタイムに対話する種類のエージェントの実利用が進みつつある．近年では NTT ドコモの「VR 警備システム」のような AI を応用して音声による自動応答が可能なエージェントの研究開発が加速しつつある．こうしたエージェントのアニメーション生成においては，エージェントの発話時だけでなく，相手となる人間側の発話時にそれを聞いていることを示す非言語的情報（バックチャンネル）の表出の重要性が指摘されており，うなずきや相槌などの自動生成手段が検討されている[1][2]．

しかしながら対話エージェントの研究において，発話交替（turn-taking）に関連させたノンバーバル情報の表出についてはまだ議論が深められていない．実際の対話の場面では，ユーザの発話ターンからエージェント発話ターンへの切り替えをある程度誘導することが必要である．人間同士の対話においては，発話交替のために人間がどのような非言語行動や発話を行っているかについて様々な調査が行われている[3]．

我々はここに着目し，人間同士の対話で行われている発話交替にまつわるジェスチャを対話エージェントに取り入れることで，より人間らしい，自然な対話を行えるエージェントが作れるのではないかと考えた．特に，相手の発話内容をよく検討しているかのようなジェスチャを模倣することで，エージェントが知性的な存在であるかのような印象をユーザに与えられるのではないかと考えた．

そこで本研究ではまず，発話交替時前後に現れるジェスチャの内どのようなものが上記目的にかなうか，プロ

のインタビュアーの仕草を観察分析して抽出した．具体的には，模擬インタビューを実施し，インタビュアーの様子を録画して表情や仕草を分析した．その結果，発話交替直後のインタビュアーの発話前に，インタビュアーが視線を横に逸らす仕草が特徴的な行動として抽出された．

次にこの視線逸らし行動をエージェントに実装し，これを行わないものとの比較実験を行った．具体的には被験者とエージェントで対話実験を行い，視線逸らし行動のありなしがエージェントの印象に与える影響をアンケート調査した．なお，今回の実験ではエージェントの発話やノンバーバル情報の表出は実験者が対話中にキーボード操作により行った．

その結果，エージェントが知性的に被験者の発話を理解し応答しているという印象が視線逸らし行動により生起していることが分かった．またそれが，会話がより自然に行われたという印象を被験者に与えることが確認された．これらより，発話交替前後の演出として，発話前行動を利用することでエージェントの態度や個性の演出をより強化できることが示された．

2. 模擬インタビュー実験

2.1 実験内容

まず，人間同士の対話時に表出するノンバーバル情報のうち発話交替時に関連したものに着目し，相手の発言を注意深く聞いているという印象を相手に与えていると考えられるものを抽出することを目的として，模擬インタビュー実験を行った．

今回の実験では，映像作家兼インタビュアーとしてイ

インタビュー映像の制作に従事しているプロに依頼して実験を行った。その理由としては、映像化を前提としたインタビューにおいてはインタビューイの発話中は、声に出してバックチャンネル情報を返すことでそれが音声として記録されることで視聴を妨げるのを避けるために、表情やうなずき等によるフィードバックが多用されることがある。このため、こうしたインタビュアーは非言語的情報の表出について経験が深く、それを録画から分析抽出しやすいのではないかと我々は判断した。なお、インタビュアー本人の談によると、近年ではオンラインでのインタビュー収録の機会が増えており、従来以上に表情や仕草によるフィードバックを強調する必要性が生じているとのことであった。こうした事情もあり、後に述べるように模擬インタビュー実験はオンラインで実施した。

インタビューイは著者の一人である大学教員で、題材はインタビューイが大学で担当している講義の設計方針について、とした。インタビュアーには事前にシラバスや講義資料の一部を渡し、質問内容について検討をしておいてもらった。模擬インタビューではインタビュアーは自由に発言可能で、ノートにメモをとりつつ行った。

模擬インタビューはZoomを用いオンラインで実施した。インタビュアーの表情や仕草の録画はZoomの録画に加え、インタビュアー側のZoom画面の傍らにスマートフォンを設置し、インカメラによって録画した。

インタビュー終了後、録画を分析した。まずインタビュアーの映像を視聴しながら、主にインタビューイの発話中、および発話交替後インタビュアーが発話する際に表出した表情や仕草を抽出し、その時の発話内容とともに記録した。このとき、表情や仕草についての記述では、表情・ジェスチャの種類に加え、頷きの大きさや間隔、顔や視線の向き、メモや資料に手を伸ばすなどの動作について記述するようにした。

その後、それらのフィードバックがどのような情報を返すことを目的になされたかを発話内容から推測し、属性情報として追記した。最後にこれを基におおまかな分類を行った。

2.2 結果

模擬インタビュー結果の分析の結果、インタビュアーに特徴的な非言語的フィードバックとして、インタビュアーの発話直前に、視線を相手（ただし実験では相手が映るディスプレイ）から逸らし、斜め上を向くというジェスチャが抽出された。また、同ジェスチャが表出される際のインタビュアーの発話は、インタビューイの発言を受けて新たな質問をする際に目立った。

同ジェスチャについて我々は、これがインタビュアーはインタビューイの発言を吟味し、次の発話内容をどうするか考えていることの表出である、と捉え、これを実装の対象とすることとした。同ジェスチャを採用した理由としては他にも、視線の移動のみで表現できるため2Dアニメーションとして表現しやすく、被験者からも認識

しやすいという理由にも依る。

なお、似たジェスチャとして斜め下に視線を向けるジェスチャも観測されたが、これは机上のメモを見ている仕草として捉えられ、これを対話エージェントに実装するとするとメモの存在を前提とせねばならず、実装および実験が複雑になる恐れがあったため、これは採用しなかった。

3. エージェントを用いた対話実験

3.1 実験概要

模擬インタビュー実験の結果から、本研究では発話前行動としての視線逸らしジェスチャに着目し、これを対話エージェントに実装して被験者を相手にした対話実験を行うこととした。具体的には被験者から見たエージェントの印象について17項目からなるアンケートにより調査し、左記ジェスチャの有無が与える影響を調査する被験者間実験を行った。

対話実験においては、被験者はディスプレイに表示された対話エージェントとの会話を行った。実験者は被験者からは目視できない場所に着席し、被験者の発話を聞きながらその文脈に応じた発話をエージェントに行わせ、またうなずきや視線逸らし行動を操作した。（図1）

3.2 エージェント設計

実験で用いた対話エージェント（図1右図）の外見は、アニメーション生成ソフト「Live2D Cubism」用のサンプルアバタとして配布されているものを使用した。このアバタに数種類のうなずきおよび視線逸らしアニメーションを作成した。またこのアニメーションを利用し、キーボード操作によりリアルタイムで発話および表情変化ができるソフトウェアをUnityを用いて実装した。

エージェントの音声はプロの声優に依頼して作成した。エージェントの発話中はLive2D Cubism SDKが提供するリップシンク機能を用いて口の開閉動作を行わせた。

図2に、今回実装した視線逸らし行動の概要を示す。左図が通常時の視線方向であり、被験者を直接目視している。実験者により視線逸らし行動が入力されると右図のように向かって右斜め上に視線を向ける。右図の状態でも2秒間経過の後、通常状態へと復帰する。この2秒間という長さは、模擬インタビュー実験においてインタビュアーが視線を逸らしていた時間を参考に定めた。エージェントからの発話は常に、通常状態への復帰後に行われた。



図1: 実験環境。左図は環境の模式図で、右図は被験者から見た実験装置の様子を示している。

3.3 実験手順

以下に各セッションにおける実験手順を説明する。

まず被験者には実験者が口頭で実験概要を伝えた。具体的には、対話エージェントがインタビュアーとして被験者と会話すること、インタビューでは題目がエージェントから与えられ、それに沿って会話して欲しいこと、エージェントは被験者の発話を聞き、それに対する質問を数度行うこと、実験終了後にアンケート調査を行うこと、などを伝えた。ただし、エージェントは実験者が操作すること、またエージェントが視線逸らし行動をとることは被験者には伝えなかった。

実験は2021年の12月に実施した。インタビューの題目は「冬休みにいきたい場所」で、インタビューの開始前に1分間、題目について考える時間を被験者には与えた。

インタビューではまず題目について被験者に自由に話してもらった。被験者の発話中、実験者はそれを聞きながらキー操作でうなずき動作をエージェントにさせた。被験者の発話が終了したら、実験者はそれまでの文脈に適した質問をエージェントに発話させた。質問文は事前に「そこを知らなかったのはなぜですか」「そこで食べたいものはありますか」といった内容のものを12種類用意した。

エージェントが発話するタイミングで、視線逸らし行動あり群においては、発話操作前に視線逸らしジェスチャをエージェントに行かせた。視線逸らし行動なし群ではジェスチャなしですぐに発話操作を行った。

インタビュー時間は3分間として、時間となったら最後にエージェントにその旨を発話させた上で、アンケート調査に移った。なお、すべての実験終了後にインタビュー中の質問回数の平均を確認したところ、エージェントからの質問の平均回数は視線逸らしなし群が6.3回、あり群が5回であった。

実験は、視線逸らしジェスチャなし群・あり群それぞれ10名ずつの、計20名の被験者を対象に実施した。被験者には実験終了後、謝金として600円を支払った。

3.4 アンケート内容

アンケートは「対話エージェントに対する評価」「対話エージェントの質問に対する評価」「対話内容に対する評価」の3種類の評価を調査するもので全17項目からなり、それぞれ5段階のリッカート尺度による評価をもらった。アンケートの質問項目は文献[4]を参考にしたほか、



図2：作成した対話エージェント。
左：通常状態，右：視線逸らし状態

エージェントの知性的印象に関する項目などを独自に追加して構成した。質問項目は図3を参照されたい。なお、「対話エージェントの質問に対する評価」「対話内容に対する評価」に属する質問項目については、それぞれなぜそう感じたかについて記入する記述欄を設けた。

4. 結果

アンケート調査の結果を図3に示す。それぞれの質問項目について、リッカート尺度の各段階の占める割合を縦軸に示し、視線逸らしジェスチャなし群・あり群の順に並べてある。

各質問項目について、対応のないt検定により検証したところ、「聞き上手だと感じたか」「会話内容を理解して発言していると感じたか」の二つの項目において、有意水準5%で帰無仮説が棄却され(それぞれ $p=0.0379$, 0.0308)、有意差が確認された。いずれの項目でも、視線逸らし行動ありのエージェントがポジティブな評価を受けていることが示された。

また、有意差は確認されなかったが差が大きく注意を要する項目としては、「実在の人間に対して話していると感じたか ($p=0.134$)」「質問内容をアバタが考えて発言していると感じたか ($p=0.075$)」「質問は定型文のように感じたか ($p=0.138$)」で、いずれの項目においてもやはり視線逸らし行動ありのエージェントの方が、より人間らしい対話を行えているという印象を与えている可能性を示唆している。

記述回答を確認したところ、「会話内容を理解して発言していると感じたか」「質問内容をアバタが考えて発言していると感じたか」の2項目において、エージェントの非言語動作に言及したものが多く見られた。「会話内容を理解して発言していると感じたか」の項目では、視線逸らし行動あり群ではそう感じた理由として、うなずき動作を挙げた者が2名、質問内容が会話に沿っていることを挙げた者が5名おり、一方なし群ではアバタの質問が被験者の回答に沿っておらず、質問ごとに繋がりが無いことに違和感を覚えた者が3名いた。「質問内容をアバタが考えて発言していると感じたか」の項目では、視線逸らし行動あり群ではその理由として、視線逸らし行動によってそれを感じたと回答した者が5名、うなずき動作によって感じたと回答した者が2名いた。なし群ではエージェントの発話に思考が感じられないと答えた者が7名いた。

被験者の発話量やその内容について、両群の違いがないかを検討したが、被験者間での個人差が大きく、両群間に明確な差異を見出すことはできなかった。

5. 議論

実験の結果より、視線逸らし行動のある対話エージェントに対して被験者は、それが被験者の発話を聞いて理解しているという印象を強化する効果を持つことが明らかとなった。

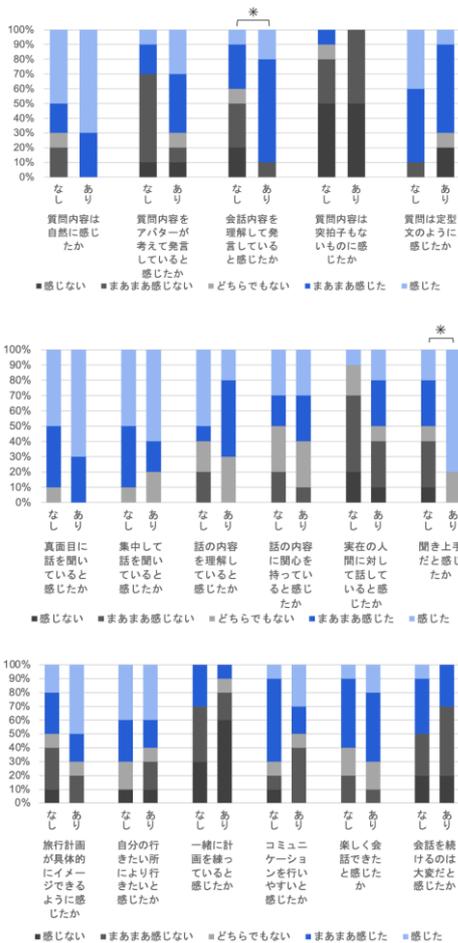


図3: アンケート結果

「聞き上手だと感じたか」という質問項目は複合的な印象を問うものであり、この項目で有意差が確認されたのは著者らにとってもやや意外な結果であった。今回の実験では質問文は事前に用意したものの中から選択するという手法をとったため、被験者の発話内容に応じて柔軟に対話を進めることはできず、総合的な項目で高評価を得ることは難しいというのが著者らの想定であった。加えて、視線逸らし行動は被験者の発話が完了してから表出されたので、被験者の話を聞いているときのエージェントの態度、という意味での「聞き上手」という印象形成を視線逸らし行動が直接喚起したとは考えにくい。「会話の内容を理解して発言していると感じたか」がポジティブに評価されたこと、また有意差は確認されなかったものの他に視線逸らし行動がポジティブに評価された項目と考え合わせると、相手の話を理解し、その文脈に沿って発話しているという印象を視線逸らし行動によって与え、それによって総合的に「聞き上手」の印象形成に成功していると考えられるであろう。これはまた、手本としたインタビュアーの行動を分析抽出して対話エージェントに実装する、という本研究の手順の有効性を裏付けてもいる。

一方で、本手法はあくまでも「聞き上手」という印象形

成を目的とした対話、すなわち相手の話を聞いている割合の多い場合でのみ有効であるのかもしれない。事実、他の項目では両群であまり印象が変化していないものも多い。対話テーマを変えてさらなる調査をする必要がある。

「質問内容をアバターが考えて発言していると感じたか」の項目については、両群間での回答傾向の差は大きかったものの有意差は確認できなかった。この項目は発話前行動が直接影響を与えそうなものであり、有意差が確認されることを著者らは想定していた。今回の実験ではエージェントの発話は声優が吹き込んだものであり、音声合成による発話特有のぎこちなさはないため、かえってその発話が事前に準備されたものであるかのように感じさせた可能性がある。より高品質な音声合成が今後一般化し、リアルタイムに生成された発話の不自然さが無くなればこれは解消される可能性がある。

6. 結論

対話エージェントを知性的に見せるための工夫として、まず人間のインタビュアーの行動を分析し、発話前行動としての視線逸らしを抽出し、対話エージェントに実装した。対話実験の結果から、同行動によってエージェントが「聞き上手」であるという印象を強化することがわかった。また発話前行動がエージェントとの対話全体の印象に対して影響を与えることが示された。

本研究の結果は、エージェントの表情演出の設計として、対話相手となるユーザの方を見続けたり、文脈に応じてうなずいたりするだけではなく、時によ見をしたりするような、様々な表情演出がより印象を強化しうることを示している。

謝辞 本研究は株式会社 Live2D の教育機関向け支援プログラムによる支援を受けたものである。ここに記して感謝の意を表す次第である。

参考文献

- [1] 渡辺富夫, 大久保雅史: コミュニケーションにおける引き込み現象の生理的側面からの分析評価, 情報処理学会論文誌, Vol.39, No. 5, pp. 1225-1231, 1998.
- [2] 渡辺富夫, 大久保雅史: 身体的コミュニケーション解析のためのバーチャルコミュニケーションシステム, 情報処理学会論文誌, Vol. 40, No.2, pp. 670-676, 1999.
- [3] 堀口順子: コミュニケーションにおける聞き手の言語行動, 日本語教育, Vol. 64, pp. 13-26, 1988.
- [4] 山口貴史ほか: 傾聴対話システムのための言語情報と韻律情報に基づく多様な形態の相槌の生成, 人工知能学会論文誌, Vol. 31, No. 4, pp. C-G31_1-10, 2016.