



Mirroring Agent : 自己の感情状態を反映するエージェントの設計とその対話がもたらす自己認識への影響

禰寝崇之¹⁾, 高下修聡¹⁾, 松平幸大²⁾, 齊藤寛人¹⁾, 櫻田国治¹⁾, 稲見昌彦¹⁾

Takayuki NESHIME, Shuto TAKASHITA, Koudai MATSUDAIRA, Hiroto SAITO, Kuniharu SAKURADA,
and Masahiko INAMI

1) 東京大学 先端科学技術研究センター (〒 153-8904 東京都目黒区駒場 4-6-1)

2) 東海大学 情報通信学研究科 (〒 108-8619 東京都港区高輪 2-3-23)

概要: 会話ができる AI エージェントは現代に浸透しつつあるが, Mirroring Agent では表情検出でユーザーの感情を読み取り, その結果がエージェントの表情, 声質, そして返答内容に現れる. さらに, スライダーコントローラーにより, 会話のスタイル等を自分の手で調整することも可能である. これによりエージェントを通じて自分自身の感情表現を客観的に振り返ることができ, 対話を通じた自己認識の深化に繋がる可能性がある. 本稿では, 提案システムの設計について報告するとともに, このアプローチがコミュニケーションや心理に与える影響について考察する.

キーワード: 対話エージェント, 感情鏡映, コミュニケーション, 自己認識

1. はじめに

自分自身を客観的に認識すること, 特に自己の感情状態を客観視することは, 円滑な社会的コミュニケーションや自己改善, メンタルヘルスにおいて重要だと指摘されている. しかし日常生活において, 自分が他者からどう見えているか, 特に「今この瞬間」の自分の表情や声のトーンをリアルタイムに知る機会ほとんどない. 対面場面では自分の顔を見ることはできず, 多くの場合フィードバックは他者の反応や事後の録画確認に頼るため, 感情表現の自己認識が難しいという課題がある. このギャップを埋める新たなアプローチとして, 本研究ではユーザ自身の感情状態その場で鏡映する対話エージェントに着目する.

従来, 多くの対話エージェントは事前に設計されたペルソナを持ち, ユーザとの対話において一貫したキャラクターや情報提供を行うものが中心であった. 例えば, ユーザの情動に反応する対話システムとして, Semaine プロジェクトの Sensitive Artificial Listener (感情傾聴エージェント) や Rapport Agent では, ユーザの音声や表情を解析し対話エージェントが感情的な応答をリアルタイムに生成する試みがなされている [1, 2]. また, 対話エージェント以外にもコミュニケーション支援ツールとして, 社会的スキルトレーニング (Social Skills Training, SST) や就職面接練習システムの研究が多数報告されている [3]. 例えば TARDIS プロジェクトではバーチャルエージェントを用いた就職面接のコーチングシステムを開発し [4], MACH システムではカメラとマイクでユーザの表情・発話を解析してエージェントがリアルタイムに質問や相槌を行い, 対話後に詳細なフィードバックを提供することで面接スキルの向上を図っ

ている [5]. 実際, MACH を用いた 1 週間の訓練実験では, 使用しなかった対照群に比べて面接時のパフォーマンス評価が有意に向上し, 参加者は「自分の行動について多くを学べた」と報告している. さらに近年では, 自閉スペクトラム症 (ASD) 者やニートの対人技能訓練へのエージェント活用や, 学生を対象とした面接練習等が行われており, エージェントとの練習が従来の訓練手法を上回る効果を示した例も報告されている [6]. これら既存研究の多くは, 特定のシナリオ (例えば就職面接や社会スキル訓練) に基づいており, ユーザーの行動に対して「正解」または「間違い」を示すような形式でフィードバックを提供する.

一方で, 特定のシナリオに縛られない汎用的な自己認識支援を行う場合には, ユーザー自身が「正解」を理解することよりも, 自分の行動や感情表現に気づくことの方が重要になる. この場合, ユーザーは他者からのフィードバックを通じて, 自分の内面に気づき, 自己改善のきっかけを得ることができる. 例えば, 「今この瞬間」の自分の感情表現に焦点を当て, ユーザー自身がそれを客観的に鏡映することができるシステムがあれば, 自己認識が深まる可能性がある.

以上を踏まえ, 本研究ではユーザ自身の動的な感情状態をリアルタイムにアバターエージェントの表情・声質・対話内容に反映させる「Mirroring Agent」を提案する. これによりエージェントとの対話を通じて, 自分の感情表出をまるで鏡で見るように客観視でき, 自己認識の変容を促すことを目指す. 本稿ではそのコンセプトとシステム設計について報告し, このアプローチが自己認識を深める新たなコミュニケーション手法となり得ることを示唆する.

2. 提案システム

2.1 コンセプト

本システムの核心は「ミラーリング」の概念にある。ユーザーの感情をリアルタイムに取り取り、まるで鏡を見るようにエージェントを介してその感情をユーザー本人に返す対話体験を提供する。ユーザーにとってエージェントは他者でありながら、自分自身の感情状態を映し出す存在となる。このような自己の感情の鏡映体験が何故自己認識の深化に繋がるのか、以下にその背景を述べる。

2.1.1 心理学的背景（バイオフィードバックと感情鏡映）

人は自身の内的状態を外部からのフィードバックによって学習・調整できることが知られている。例えばバイオフィードバックでは、生理指標のリアルタイム表示により自身のストレスや興奮状態を認識・制御しやすくなる [7]。同様に、感情表現のフィードバックも自己認識に重要である。発達心理学では、乳児が養育者から示される表情フィードバック（例えば自分が泣いた時に母親がオーバーな悲しい顔を見せる等）を通じて、自分の感情状態に気づきラベリングできるようになるとする感情の鏡映理論が存在する [7]。本研究の Mirroring Agent は、こうした人間の情動鏡映のメカニズムを対話エージェントで再現し、利用者が自分の感情状態を言語・非言語両面からフィードバックされることで認識を深めることを目指す。

2.1.2 自他の境界と VR の特性

エージェントが自分の感情を表出する様子を見ることは、言わば他者の姿を借りて自分自身を見ることでもある。このとき、特に VR など没入型の環境では、自身と他者の境界が通常より曖昧に感じられることが指摘されている。実際、Mirror Ritual と呼ばれるインタラクティブアートでは、ユーザーの感情状態を詩的なテキストに変換して鏡に映し出すことで、内面の感情と外在的表出との調和や不一致にユーザー自身が気づきを得る試みがなされている [8]。このように、システムを通して自分の感情表出を客観視する体験そのものがユーザーに内省を促し、自己と他者の視点を行き来する新たな自己認識をもたらすと期待される。

以上のコンセプトに基づき、本研究では「自分の感情がエージェントを通じて返ってくる」という独自の対話体験を設計する。ユーザーはあたかも自分自身と対話しているかのような感覚を得る一方で、エージェントという客観的な鏡を通して今の自分を見つめ直すことになる。これにより、一人では得にくいリアルタイムの自己フィードバックを得て、自己の感情表現や状態に対する新たな気づきや理解が生まれることを目指す。

2.2 システム設計

提案システム「Mirroring Agent」の実装画面とアーキテクチャ全体をそれぞれ図 1、2 に示す。図 2 では、ユーザからの入力に対し、感情推定を行い、エージェントの表情・対話・声に反映する一連のフローを矢印付きで示している。以下、各コンポーネントの役割を具体的に説明する。



図 1: Mirroring Agent の実装画面。

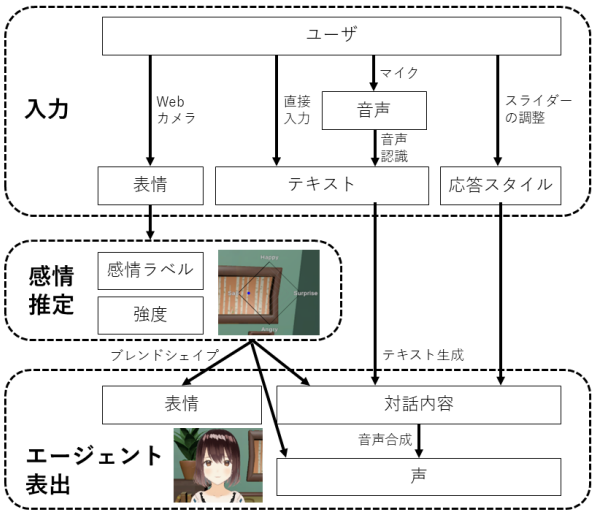


図 2: Mirroring Agent のアーキテクチャ全体。

2.2.1 入力

ユーザの表情及び音声情報を収集する。表情は Web カメラから取得して感情推定に用い、音声情報はマイクから取得し、こちらは音声認識を用いてテキスト情報へと変換する。音声認識には Whisper を使用しているが、音声入力に代わり直接テキスト入力を行うことも可能である。

また、本システムにはユーザがエージェントの応答スタイル（例えば語調が丁寧かフランクか等）を調整できるようにコントロール用のスライダーを用意している。これは、本システムが様々なユースケースに対応できるようにするための仕組みであり、例えば日常のパートナーとして使用する場合はフランクで距離感の近い返答を返すようにし、就職活動の面接練習として活用する場合は丁寧で落ち着いた応答を行うように調整が可能である。音声（テキスト）情報を入力する前にこちらのスライダーを調整しておけば、最終的なエージェントの発話内容がこのスライダーの値に応じて調整される。このスライダーは発話内容のみに対して作用するものではあるが、適切に用意を行えば様々な調整が可能であり、ユーザは自分が相談したい相手像（優しい聞き役、率直にフィードバックするコーチ等）に近いものをスライダー調整を通じてカスタマイズすることができる。

このようなユーザ主導の介入要素を設けることで、より個人に合った自己フィードバック体験を提供することを可能とする。

2.2.2 感情推定

入力で取得した表情に基づき、ユーザーの現在の感情状態をリアルタイムに推定する。今回は Unity Sentis に感情推定の ONNX モデルを組み合わせたものを使用する。このモデルは多様な感情を識別でき、入力で得られた表情に対し各感情それぞれの予測確率を推定するが、エージェントへの反映と認識の簡易化のため、「喜び」「怒り」「驚き」「悲しみ」「平常」の5種類の感情に絞り、予測確率が一番高いものを分類結果として採用する。また、その予測確率を感情の強度を表すパラメータとして使用する。

2.2.3 エージェント表出

ユーザより得られた感情は3つの形でエージェントにミラーリングされる。

表情のミラーリング：ブレンドシェイプ機能を用い、エージェントに感情推定で得られた感情を反映させる。ただし、「平常」は基準の表情として用い、「喜び」「怒り」「驚き」「悲しみ」の4つに対してブレンドシェイプによる表情変化を適用する。感情の強度とブレンドシェイプの値を対応させることにより、例えばユーザが「少し驚いている」ときはエージェントも少し眉を上げる程度であるのに対し、「大きく驚いている」ときはエージェントも大きく目を見開くようにスケーリングを行う。

対話内容のミラーリング：エージェントが発する言語の応答内容の生成には AzureOpenAI の gpt-4.1-mini モデルを使用する。単に発話内容を生成させるだけでなく、そこに感情を反映させる。例えば「喜び」の感情が得られたときは発話内容もポジティブで楽観的な内容にし、一方で「悲しみ」の感情が得られたときはネガティブで悲観的な返答をするようにプロンプトで調整を行う。また、入力時に調整した応答スタイルのスライダーの内容も同じく gpt-4.1-mini モデルを利用してこの時点で同時に反映される。

声のミラーリング：音声合成には StyleBertVITS2 を使用する。このモデルは日本語特化で学習を行うことにより自然な日本語の発声ができるとともに、複数スタイルでの学習を行えば感情の出し分けが可能となる。得られた感情に合わせて、ピッチと話速、そして StyleBertVITS2 のスタイルを指定することで、エージェントの感情の籠もった発話を実現する。

以上のように、ユーザの感情推定結果を多面的にエージェントの表現にマッピングすることで、ユーザはエージェントとの対話を通じて自分の感情と向き合うことになる。

2.3 実装

改めて図1が本システムの実装画面であり、開発には Unity を用い、3D アバターには VRoid より提供されているアバターサンプルを使用した。画面中央にミラーリングを行うエージェントを配置し、画面左に会話ログを表示、画面右に感情値や応答スタイルのスライダー、テキスト入力

の UI 等を配置している。

表情のミラーリングは常にリアルタイムで行うが、発話内容と声のミラーリングはユーザの発話もしくはテキスト入力完了したタイミングの感情を取得し、その感情を反映させて処理を実行する。したがって、感情推定と表情のミラーリングに関してはリアルタイム処理が重要になってくるが、入力時の音声認識、発話内容と声のミラーリングに関しては会話体験を損なわない範囲で遅延を許容する。そこで、通信の遅延が発生しないように感情推定はローカルで処理を行い、残りの機械学習推論は外部サーバで処理を行う。言語処理は上記に記したように AzureOpenAI の API を利用し、音声認識と音声合成は専用の API サーバを Python で構築した。処理が重い機械学習推論を外部サーバで行うことにより、スペックが高くないオンボード GPU の PC でも自然な会話を体験することができ、本システムは一般的なノート PC (CPU: Intel Core i7-1360P, GPU: Intel Iris Xe) で動作することを確認している。

カメラやマイクもノート PC 内蔵のもので動作することを確認しているが、場面に応じて適切に外付けの機器を使用することを想定している。例えば、オンライン上での会話を想定する場合はそのままノート PC 上での使用で問題ないが、対面での面接練習に用いるような場合であれば、エージェントが実際の人間と大きくなるように大型モニターを用意し、それに合わせてユーザを真正面に捉えるようにカメラやマイクを設置する。

3. 想定される応用と今後の展望

本システムはまだ実証段階であるが、その研究の価値と将来性について、現時点で想定している応用例および今後計画している評価・発展内容を述べる。

3.1 ユースケース

本システムは、自己の感情表現を客観視しながら対話練習や自己調整を行うためのツールとして、様々な場面での応用が期待できる。具体的な利用シナリオとして、以下のようなケースを想定している。

3.1.1 面接練習

面接中の自分の表情や話し方の癖を客観的に把握する目的で活用できる。緊張すると顔が強ばったりする人は少なくないと思われるが、本システム上ではそれがエージェントの緊張した態度として即座に返ってくるため、自分がどれほど緊張して見えるかをその場で認識できる。従来の自動面接コーチシステム MACH においても、エージェントとの模擬面接後に参加者の面接スキルが向上し、自分の行動を振り返るのに役立ったとの報告があるが、本システムではその場のフィードバックがより直感的な「鏡」として得られるため、短時間での改善サイクルを促進できる可能性がある。

3.1.2 感情理解

自己・他者問わず、感情を認識したり表現することが難しい状態をアレキシサイミア (失感情症) と呼ぶ。このよう

な人は、自分の感情表現が他者にどう伝わっているかを読み取ったり想像することが難しく、それが原因で対人トラブルを招く可能性がある。本システムは安全な仮想対話環境で、自分の感情表現とそれに対する他者の反応を学ぶ場を提供できる。例えば、本人は自覚していなくても表情が不機嫌そうな顔をしている場合、エージェントもそれに合わせて不機嫌そうな応答を行う。これをきっかけに「自分は不機嫌なのか」と気づき、感情コントロールや表情の練習に繋げることができる。特にエージェント相手であれば失敗を恐れず繰り返し利用できる利点もあり、本システムを対人スキルトレーニングの場としても活用できる。

3.1.3 日常的な感情セルフモニタリング

ユーザが日々の中で自分の感情変化を把握し、セルフコントロールするツールとしての利用も考えられる。例えば、ストレスフルな仕事終わりに本システムと対話することで、自分がどれほど疲れた顔・声になっているかを可視化し、必要に応じてリラクセス法を試すきっかけになる可能性もある。あるいは、家庭内で感情的になってしまった後に本システムと話してみると、エージェントが怒りの表情で反応することで自身の怒り具合に気づき、クールダウンできるといったセルフケア的な使い方も考えられる。要するに、本システムは「感情の鏡」として日常の感情状態を振り返る習慣づくに貢献し得る。

以上のように、コミュニケーショントレーニングからメンタルヘルスまで、幅広い領域で Mirroring Agent の応用が期待できる。エージェントの感情表現は実際の人よりも単純でわかりやすいが、そのエージェントを通じて自己を見つめることにより、自己の感情に向き合いやすくなる効果があると考えており、それによって得られる気づきや学びが各領域の課題（コミュニケーション改善、自己理解、セルフケア等）に寄与すると思われる。

3.2 今後の評価計画と展望

Mirroring Agent の利用がユーザの自己認識に与える影響を定量・定性の両面から検証する予定である。具体的には、本システムを面接練習等の特定のシナリオで使用するにより、ユーザが自分の感情表現に対する気づきや自己理解が深まるのか、また対話エージェントとしての有用性（練習効果や満足度）がどの程度得られるのかを明らかにする。「自己認識を深める新たなコミュニケーション手法」としての本システムの有効性を示すことが目標となる。実験参加者へのアンケート調査やインタビューにより主観的指標を測定する。また、対照群としてミラーリング機能をオフにしたエージェントを用意して、表情（感情推定）の変化を観測することにより、定量的にどの程度感情を意識していたかを測定する。

本システムを利用することにより参加者の自己認識指標が向上し、特に自分の感情表現に対する気づきが高まると考えている。もし、実験でポジティブな結果が得られれば、次の段階として長期的な使用による自己認識の変化や行動変容の調査を検討する。将来的には、生体情報の活用（心

拍や皮膚電位による内的情動のフィードバック）との組み合わせや、より自然な対話理解・応答生成（大規模言語モデルの活用など）によって、ユーザの感情に深く寄り添いながらも客観視を促すインテリジェントな対話エージェントへと発展させる予定である。

4. むすび

本稿では、ユーザ自身の感情状態をリアルタイムに反映する Mirroring Agent のコンセプトとシステム設計について説明を行った。従来の対話エージェントやコミュニケーション支援ツールにはないアプローチとして、ユーザの「今この瞬間」の感情を対話の中で鏡映し、自己の感情表出を客観視させる仕組みを提案した。今後は被験者実験を通じて、本システムの有効性を検証していくとともに、個人の自己理解を支援する新たなコミュニケーションツールとして本システムを発展させていく。

謝辞 本研究は、JST ムーンショット型研究開発事業（JP-MJMS2292）及び、公益財団法人セコム科学技術振興財団の支援を受けたものである。

参考文献

- [1] Douglas-Cowie, E., Cowie, R., Cox, C., Amir, N., Heylen, D.: The Sensitive Artificial Listener: an induction technique for generating emotionally coloured conversation, Proc. ACHI 2009, pp. 1–7, 2009.
- [2] Gratch, J., Wang, N., Gerten, J., Fast, E., Duffy, R.: Creating Rapport with Virtual Agents, Proc. IVA 2007, pp. 125–138, 2007.
- [3] Tanaka, H., Takahashi, T., Yoshimura, T.: Automated Social Skills Trainer, Proc. IUI 2015, pp. 17–27, 2015.
- [4] Anderson, K., et al.: The TARDIS Framework: Intelligent Virtual Agents for Social Coaching in Job Interviews, Proc. ACM ACE 2013, pp. 476–491, 2013.
- [5] Hoque, M., Courgeon, M., Kaplan, F., Mutlu, B., Picard, R. W.: My Automated Conversation Coach (MACH), Proc. UbiComp '13, pp. 697–706, 2013.
- [6] Damian, Ionut; Baur, Tobias; Lugin, Birgit; Gebhard, Patrick; Mehlmann, Gregor; André, Elisabeth: Games Are Better Than Books: In-situ Comparison of an Interactive Job Interview Game with Conventional Training, Proc. AIED 2015, pp. 84–94, 2015.
- [7] Gergely, G., Watson, J.: The social biofeedback theory of parental affect-mirroring: The development of emotional self-awareness and self-control in infancy, Int. J. Psycho-Anal., vol. 77, pp. 1181–1212, 1996.
- [8] Rajcic, N., McCormack, J.: Mirror Ritual: An Affective Interface for Emotional Self-Reflection, Proc. CHI 2020, pp. 1–13, 2020.