関係性に応じた我慢と限界を混合表情により表す不快感情抑制エージェントの印象評価

藤野惇士 [1*]　柳朋輝 [1]　米澤朋子 [1]

[1] 関西大学
[1] Kansai University

Abstract: 本研究では，不快感情を抑制しつつ適切に表現する人間らしいエージェントの実現を目指す．まず我慢を伝える表情として，FACS の Action Unit を用いた部分選定を行い，ユーザの接近に応じて蓄積した不快感情を「我慢」として表出するエージェントを提案する．我慢とその限界後に表出される表現の印象や，関係性による変化を検証した．その結果，複数の感情表現が異なる顔部位に並行して表現される混在表情が我慢を表すことが確認され，我慢をすることにより好感を持たれ，今後も交流をしたいと評価された．しかし，我慢後の表現の印象には，関係性による差は見られなかった．

1 はじめに

近年，店舗での接客業務や施設の案内，教育や介護の現場など，多岐にわたる領域で人型アンドロイドや対話エージェントといった対話のある自律システムの社会実装が進められている．しかし，普及に伴って，ロボットやエージェントに対して，むやみやたらに触る，不適切な言動や態度を取るといったハラスメントが問題として提起された [1]．これらの行為は，ロボットの故障に繋がる可能性だけではなく，人間相手のハラスメント行為に対する心理的なハードルを下げたり，目撃した周囲の人間に不快感や心理的な悪影響を与える恐れがある．社会へのさらなる普及に支障をきたす可能性があるため，ロボット自身が拒絶の意思を示す必要がある．しかし,「嫌」の意思表示のみでは，円滑なコミュニケーションを阻害し，ユーザとの良好な関係性を構築しづらくなる．このような状況は共生社会の実現に悪影響を及ぼしかねず，ユーザとの関係性や状況を踏まえた拒絶の度合いを含む意思表示能力が必要になる．

そこで本研究では，日本人特有のコミュニケーション様式である「本音と建前」[2] の概念を取り入れた内部状態モデルを実装し，そのモデルを搭載するエージェントを作成し，その効果について検証を行う．内部状態モデルは，ストレス値，我慢値，感情表出値，感情制御値の 4 つのパラメータで構成されている．ストレス値は，ユーザの行動によってエージェントが受ける即時的な不快の変化による時間的な蓄積を考慮して計算される．我慢値は，エージェントが不快感情の抑制を開始するための閾値である．蓄積されたストレス値がこの値を上回った場合，エージェントは即座に怒るのではなく，社会的な建前としての「我慢状態」へ移行する．感情表出値は，エージェントが我慢状態を維持できる限界点を示す閾値である．ストレス値が蓄積し，この値を上回った場合，エージェントは本音である「怒り状態」へと移行する．感情制御値は，時間の経過とともにストレスを自然減衰させるためのパラメータであり，この値は，エージェントが受けたストレスを内部で処理し，時間の経過とともに平常心を取り戻す能力を表現している．本モデルにより，エージェントは不快な刺激に対して即座に反応するのではなく,「我慢(建前)」を行い，限界を超えた時点で「本音」を表出するという，状況に応じたふるまいを行おうとするプロセスを再現する．そして，実装したエージェントによる不快感情の抑制が，ユーザに人間らしさや社会的知能を感じさせ，関係維持に寄与するか，またその効果がユーザとの関係性によって変化するかという仮説を検証した．仮説が支持されれば，ハラスメント等の不快な状況下においても，単なる拒絶ではなく，エージェント自身の尊厳を守りつつユーザとの関係を破綻させないエージェントの有効性が確認できる．また，将来的に人間とエージェントが共生する社会において，エージェントが「本音と建前」を使い分けることで，長期的な信頼関係を築くことが可能になることが期待される．

*連絡先： 関西大学
〒 569-1095 大阪府高槻市霊仙寺町 2-1-1
E-mail: [k655714,yone]@kansai-u.ac.jp

2 関連研究

2.1 関係性に応じた感情表現

対人コミュニケーションにおいて，人間は相手の感情や行動を単に受動的に受け取るのではなく，社会的な文脈や関係性に基づいた「期待」や「規範」に照らし合わせて解釈している．Holt は，怒りの表出が対人関係および個人内に及ぼす影響について論じ，怒りを適切に制御することが社会的な適応において重要であることを示している [3]．特に日本の文化においては，怒りの直接的な表出は，避けられる傾向にある．木野は，日本人が怒りを表出する際,「表情・口調」や「遠回し」といった抑制的な行動を多用することを明らかにした [4]．また，対人場面において相手の非を冷静に指摘するように制御された表出は，特に公的な場面において良いとされる傾向があり [5]，人間は文脈に応じた適切な反応を期待しながら対話を行っている．こうした感情制御は，相手との関係性によって変化する．吉田は，親密で対等な二者関係における怒りの制御方略を調査し，関係性が高いほど，関係維持のために建設的な感情表出が促進される．一方で，関係が浅い場合や自身の表出に否定的な態度を持つ場合，抑制的な方法が選択されることを示した [6]．一方エージェントとのやりとりにおいても，既知かそうではないかにより人間の行動選択に変化が認められる [7]．人間側が関係性に応じた行動変化を示すように，ロボットにも類似した機構があることが相互理解につながると考えられる．これらのことから，エージェントにおいても，相手との関係性や状況に応じて「怒り」と「抑制（我慢）」を使い分けることが，円滑な関係性を維持するために不可欠な機能であると言える．

2.2 エージェントの表情

人間の顔表情と感情に関わる多くの先行研究がある [8, 9, 10]．Ekman の Facial Coding System は，顔の動きを Action Unit という単位にし，表情を客観的に分析するためのものとなっている [11]．このような人間の表情に倣い，エージェントの感情表現が人間らしい複雑さを持つことは，インタラクションのリアリティや体験の質を増幅させることに繋がる．感情は必ずしも一つではなく，心の社会 [12] の概念のように，人間の心には単一の状態があるわけではない．このような複雑で並行した様々な感情が混在することを混合感情と呼ぶ [13]．加納らは，感性会話ロボット Ifbot を用い，生理的な現象と筋肉による表情変化を区別して制御することで，相反する感情が同居する「混合感情」の表現を実現した [14]．また,「本音と建前」のような二重構造を持つ感情表現に関しては，目良らが「ツンデレ」を題材とした表情生成を行っている．彼らは顔部位の変化の随意性・不随意性に着目し，頬の紅潮（不随意）で本心の喜びを，口元の動き（随意）で演技的な怒りを表現することで，本心を隠蔽しようとする状態をユーザに認知させた [15]．曹らもまた，喜びと怒りのモーフィング画像を用いた認知実験を行い，日本人が表情から他者の「内心の感情」と「表出された感情」を区別して認知している可能性を示唆している [16]．さらに，Richoz らが指摘しているように，表情の知覚には静的な特徴だけでなく動的な特性が影響する [17]．大道らは Action Unit の組み合わせを用いた快感情の抑圧表現において，微細な動きが抑制の印象に寄与することを報告しており [18]，我慢という動的な葛藤を表現する上でこれらの知見は重要だと考える．本研究では，エージェントの不快感情に対する反応として，単に怒るのではなく「我慢」を行い，その際の複雑な内部状態を，異なる部位に混在するそれぞれの表情を示すことにより表現することとした．

2.3 人間らしさと社会性

エージェントが社会的パートナーとして受容されるためには，単なる知能だけでなく，社会的なふるまいや「人間らしさ」を感じさせる要素が不可欠である．Breazeal は，ヒューマノイドロボットが人間と自然に交流するためには，感情モデルに基づく表出行動が重要であるとし，社会的相互作用における感情の調整機能を実装した [19]．日本的な文脈における社会性として，吉田らは「本音と建前」に着目し，エージェントの隠された欲求（本音）を生理反応（心拍変動）で，抑制された振る舞い（建前）を視線や距離で表現するモデルを提案した．心拍の変化と抑制的な行動の組み合わせが，ユーザに「欲求を我慢している状態」を強く知覚させることを示している [2]．また，目良らは，ユーザとの対話から生起する感情に基づいてエージェントの気分状態を遷移させることで，人間らしい一貫性のある反応を実現している [20]．エージェントが単に反応を返すだけでなく，社会的な「我慢」のふるまいを見せることは，関係性の破綻を防ぐための有効なインタラクション戦略となり得ると考えられる．

2.4 パーソナルスペースの定義と心理的影響

対人関係における社会的規範の一つとして，物理的距離などの非言語シグナルが挙げられる．Hall は，人間の周囲に形成される不可視の領域を「パーソナルスペース」と定義し，この領域への侵入が防衛反応を引き起こすと論じた [21]．八重澤らは，他者の接近に対する生理的・認知的反応を多次元的に解析する実験を

行った．その結果，他者が接近するにつれて，不安や緊張といった主観的な不快度が増加するとともに，瞬目（まばたき）などの生理反応が変化することを確認している [22]．この現象は対エージェントにおいても同様であり，Bönsch らは VR 空間内のエージェントの感情（怒りなど）がユーザのパーソナルスペースに影響を与え，ネガティブな感情を持つエージェントに対してはユーザが距離を取る傾向があることを示した [23]．これらのことから，他者の接近が心理的な負荷（ストレス）を伴うことを示しており，本研究においてユーザの接近行動をストレス蓄積を誘発するものと定義し，それに基づく反応モデルを構築することとした．

3　不快感情抑制表現エージェント

3.1　システム概要

本システムは，VR 内のユーザ位置とエージェント位置の間の距離に応じて，パーソナルスペースの侵害により発生するエージェントの内部のストレス値を計算し，関係性に応じてそのストレスを抑制（我慢）したり，表出させたりするエージェントの表情を出力する．すなわち，エージェントに対するネガティブ刺激としてパーソナルスペースへの侵入を用いる．開発環境には Windows 11 搭載 PC を用い，Python 3.12.7 および Unity 2022.3.48f1 にて実装を行った．エージェントモデルの作成には VRoid Studio を使用した．入力には Web カメラを用い，カメラ映像から MediaPipe を用いたユーザの顔領域の検出を行い，その大きさに基づいて VR 内のユーザとエージェントの距離をリアルタイムに推定している．推定された距離を仮想空間内のカメラ位置に反映させることで，VR 内のユーザの視点移動を制御するとともに，距離に応じたエージェントのストレス値に基づき，表情出力を制御する．具体的には，ストレス値 $S(t)$ が閾値 α 以上の場合は「怒り表情」を出力し，閾値 $\alpha > S(t) \geq \beta$ の場合は「我慢の表情」，閾値 β 未満の場合は「平常表情」を出力する．顔位置推定には Python 上で OpenCV および MediaPipe を利用し，UDP 通信により Unity のスクリプトへ値を送信する．Unity 側では，受信した距離データに基づくストレス値の計算，表情の遷移制御，および仮想カメラの移動制御を行っている．これらのシステムの処理フローを図 1 に示す．以下に，各部の処理の詳細を述べる．

3.2　距離計測部

距離計測部では，Python 環境下で Google が開発した機械学習フレームワークである MediaPipe を用い，

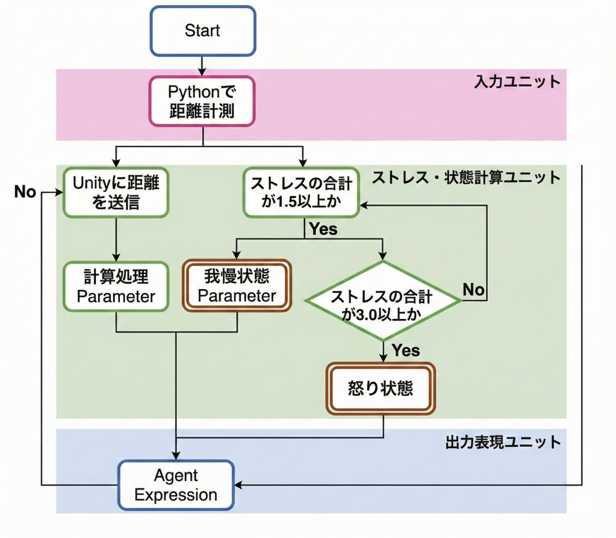


図 1: システムの処理フロー

Web カメラから顔位置を検出する．距離推定の手法には，検出された顔の位置情報とカメラの結像関係を利用する．具体的には以下の式 (3) に示すように，検出された顔領域の幅を w，一般的な成人の顔の平均幅を W（本システムでは 16.0 cm），事前にキャリブレーションにより求めたカメラの焦点距離を f（本システムでは 600）とし，これらを用いてエージェントからユーザまでの物理的距離 d (cm) を算出する．計測された距離データは，UDP 通信を用いて Unity へフレームごとに送信される．本システムでは，エージェントのパーソナルスペースへの侵入を不快の源泉とし，距離 $d(t)$ に応じたストレスを計算する式を以下の節で定義する．

3.3　エージェントのストレス値計算部

ストレス値計算部では，UDP 通信から受信したユーザとの距離データ $d(t)$ に基づき，エージェントの内部状態であるストレス値を計算する．まず，時刻 t における瞬間的なストレス値 $s(t)$ を計算する．エージェントのパーソナルスペースへの侵入により，ストレスを受けることとし，距離 $d(t)$ に応じたストレスを以下に定義する．

$$s(t) = e^{-d(t)/225} \tag{1}$$

ここで，225 は距離に対する感度を調整する減衰係数である．ユーザがエージェントに接近するほど，指数関数的に高いストレスが発生する．次に，蓄積されるストレスの総量 $S(t)$ を以下に定義する．

$$S(t) = S(t-1) \times 0.7 + s(t) \tag{2}$$

ここで，係数 0.7 は時間の経過とともにストレスが自然に減衰することを表す．ユーザが接近を続けるとストレスが蓄積され，離れると徐々に落ち着くことを再現している．最終的に，計算された蓄積ストレス値 $S(t)$ が閾値 α を超え，怒り表情を表出した直後はストレス値の計算を一時的に停止し，表情が頻繁に切り替わることを防いでいる．

3.4 エージェントの外観表現部

エージェントの外観モデルの作成には，VRoid Studio を使用した．特定のキャラクター性による印象へのバイアスを排除するため，装飾の少ない標準的な衣服を着用したモデルを採用した．エージェントの表情生成には，Facial Action Coding System（FACS）に基づく Action Unit（AU）を用いた．我慢表情については，快状態の目と不快状態の口，または不快状態の目と快状態の口，というように，相反する感情パラメータを混在させた表情を実装した．この表情の出力制御は，前節で算出されたストレス総量 $S(t)$ に基づき，以下の条件分岐によって行われる．$S(t) > \alpha$（閾値以上）で怒りを表出し，$\alpha > S(t) > \beta$（2 種の閾値の間）では我慢を表出する．また，$S(t) < \beta$（閾値以下）では平常の表情を出力する．ここで，怒り状態への遷移直後はストレス値の更新処理そのものが一定時間停止する仕様となっている．そのため，$S(t)$ の値や距離の変化に関わらず，怒りの表情が継続して出力され，閾値付近での頻繁な切り替えを防いでいる．また，表情が瞬時に切り替わることによる不自然さを低減するため，表情遷移には線形補間処理を適用した．先行研究 [24] を参考に，表情遷移時間は 1.0 秒とし，連続的に滑らかに変化するように設定した．

4 実験

4.1 実験概要

本実験は，2 つの実験から構成される．実験 1 では，不快感情の我慢状態を伝える表情の選定のための評価を行った．実験 2 では，実験 1 から得られた我慢を表す表情を利用し，我慢の有無によりエージェントの印象評価にどのように影響するのか（実験 2-1），また我慢後に表出される表情が関係性の設定により印象がどのように変化するか（実験 2-2）を検証した．

4.2 実験 1

4.2.1 目的

不快感を示す我慢の表現に適切な表情および混在表情の影響を明らかにすることを目的とする．なお，ここでの混在表情とは，VRoid Studio の表情パラメータにおける怒り（Angry）および喜び（Joy）の値を顔の上部（眉）と下部（口）のそれぞれに異なる設定をする，あるいは，部位に両方の感情による表情の特徴点を中間的な位置に移動させた表情の設定を意味する．

4.2.2 仮説

検証にあたり，次のように仮説を設定した．仮説 1）感情抑制時のエージェントの表情が感情の混在である場合，ユーザはエージェントが本心の感情を隠そうとしていると認識し（i），人間らしさ（ii）の評価を高める．仮説 2）エージェントの表情が感情の混在，特に顔上部の眉が快感情で顔下部の口が不快感情である場合、エージェントが感情を抑制していることを認識できる．

4.3 実験刺激

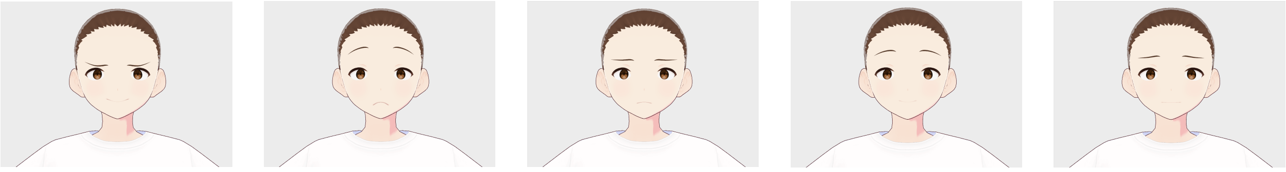
提案システムで利用した VR 環境のエージェントの顔画像を条件に基づき撮影し実験刺激に利用した．

4.3.1 実験手順

実験には，インフォームドコンセントに基づく同意書を読み参加に同意した，20-28 歳の男女 24 名（男性 15 名，女性 9 名，平均年齢 22.20，標準偏差 1.802）がオンラインで参加した．手順は，1）実験設定に関する説明，2）各条件に基づく表情をしたエージェントの顔画像の観察，3）各評価項目への回答の 3 段階である．まず，提示エージェントをユーザが不快にした状況を想定してもらうことを説明した．次に，エージェントの平常時の表情の顔画像を提示した後，各条件に基づいた感情抑制表情の顔画像を提示した．また，各条件の提示順はラテン方格法に基づいてカウンタバランスを考慮して決定した．

4.3.2 実験条件

顔の眉と口の感情値の組み合わせ要因（A1：眉不快・口快，A2：眉快・口不快，A3：眉不快・口不快，A4：眉快・口快，A5：眉混合・口混合）の計 1 要因 5 条件の被験者内実験計画とした．刺激として提示した顔の表情を図 2 に示す．A1 と A2 は部分混在，A5 は位置混合である．



(a) A1　(b) A2　(c) A3　(d) A4　(e) A5

図 2: 実験に使用した刺激画像（A1：眉不快・口快，A2：眉快・口不快，A3：眉不快・口不快，A4：眉快・口快，A5：眉混合・口混合）

4.3.3　評価項目，評価方法

感情表出（Q1），感情抑制（Q2，Q3，Q6），表情の自然さ（Q4），人間らしさ（Q5）に関する項目を設けた．各質問項目に対し，Visual Analogue Scale（VAS）法に従い，0–99（0: 全くあてはまらない，99: 非常に当てはまる）で，Google フォームを用いて回答を求めた．

- Q1 このエージェントは不快感を感じている
- Q2 このエージェントは本当の感情を隠そうとしている
- Q3 このエージェントは本当の感情が表情に出ないようにしている
- Q4 このエージェントの表情は自然な表情である
- Q5 この表情をするエージェントは人間らしいと感じる
- Q6 この表情は我慢を表現している

4.3.4　結果

実験により得られた各評価結果に対して繰り返しのある分散分析を有意水準 $\alpha = 0.05$ で実施した．図 3 に各評価項目の平均値およびその標準偏差を，表 1 に分散分析の結果を示す．

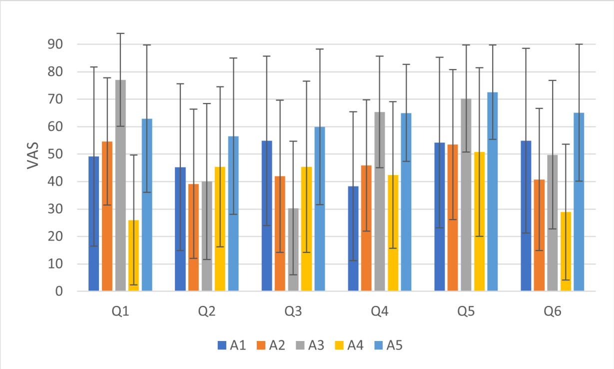


図 3: 実験 1 各評価項目の平均値および標準偏差

まず，感情表出（Q1）について，要因 A で有意差が確認され，$A_3 > \{A_1, A_2, A_5\} > A_4$ が示された．このことから，A3 条件が最も不快感を表出していることが示された．次いで，A1，A2，A5 となった．感情抑制（Q2，Q3，Q6）について，Q2 では内的な意図を評価させたが要因 A では有意差が確認されなかった．外的な意図を評価させた Q3 では要因 A で有意差が確認され，$A_5 > \{A_2, A_3\}$，$A_1 > A_3$ が示された．我慢をしているかの評価の Q6 では，要因 A で有意差が確認され，$A_5 > \{A_2, A_3, A_4\}$，$\{A_1, A_3\} > A_4$ が示された．このことから，感情の混在および混合表現の条件（A1，A3，A5）が不快感では重要となり，感情抑制では高い評価を得ており，感情が混在や混合して表現された表情は本心の感情を隠そうとしていると認識される可能性が示された．表情の自然さ（Q4）については，要因 A で $A_3, A_5 > \{A_1, A_2, A_4\}$ が示された．A3，A5 が自然な表情である可能性が示された．人間らしさ（Q5）については，要因 A で有意差が確認され，$A_3, A_5 > \{A_1, A_2, A_4\}$ が示された．このことから，人間らしさの評価は，A5 の感情が混合した表情が人間らしい評価を高める可能性が示された．しかし，A3 の感情が不快感情に統一されている条件でも人間らしい評価を高める可能性が示されている．これらの結果から，A5 が我慢している表情であるとする可能性が高く，仮説 1 が支持され，仮説 2 は一部が支持された．

4.4　実験 2

4.4.1　目的

接近を伴う VR 内のインタラクションにおいて我慢の表情表出効果を検証した．実験 2 は 2-1 と 2-2 の 2 つで構成される．実験 2-1 は不快感情の抑制表現の有無によるエージェントに対する評価を検証する．実験 2-2 は関係性により我慢後に表出される表情はどのように表出されるのが好まれるか検証した．

表 1: 実験 1 における分散分析および多重比較の結果

評価項目	F 値	p 値	多重比較
Q1	20.719	$< .001^{**}$	$A_3 > \{A_1, A_2, A_5\} > A_4$
Q2	1.572	.188	n.s.
Q3	4.227	$.003^{**}$	$A_5 > \{A_2, A_3\}, A_1 > A_3$
Q4	7.216	$< .001^{**}$	$A_3, A_5 > \{A_1, A_2, A_4\}$
Q5	4.084	$.008^{**}$	$A_3, A_5 > \{A_1, A_2, A_4\}$ $A_5 > \{A_2, A_3, A_4\}$
Q6	7.113	$< .001^{**}$	$\{A_1, A_3\} > A_4$

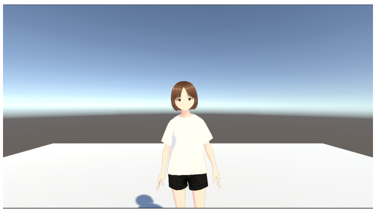
$^{**}p < .01, n.s. : notsignificant$

4.4.2 仮説

検証にあたり，次の仮説を設定した．仮説 1）エージェントが関係性に応じて感情を抑制・表出することで，ユーザとのインタラクションが促進され i），人間らしさが向上し ii），好意的な評価をされる iii）．仮説 2）我慢の表現ができるエージェントは人間らしいと評価される．仮説 3）関係性に応じて我慢を行うエージェントは，親しい場合は我慢をしない方が好意的に評価され i），親しくない場合は我慢をする方が好意的に評価される ii）．

4.5 実験刺激

提案システムの VR 環境内でエージェントに接近する動画を，実験条件に基づき撮影し，刺激として利用した． 実験動画の生成において，我慢の持続時間は 10 秒に設定している．なお，我慢状態の表情は実験 1 での条件 A5 を採用した．条件 A5 は，眉と口ともに快感情と不快感情が混在したものを表情に適用している．以下に，動画内での接近前と接近後の画像を図 4 に示す．



(a) 接近前

(b) 接近後

図 4: 動画内での接近前と接近後の様子

4.5.1 実験手順

実験には，20-28 歳の男女 24 名（男性 17 名，女性 7 名，平均年齢 22.38，標準偏差 1.703）が参加した．実験 2 はオンラインで実施し，Google フォームを用いた．手順は，1）実験設定に関する説明，2）各条件に基づくエージェントの動画の視聴，3）各評価項目への回答の 3 段階であるまず，提示エージェントをユーザが不快にした状況を想定してもらうことを説明した．実験 2-1，2-2 ともに実験参加者はエージェントに主観視点で接近している動画を視聴し，実験 2-1 では条件に基づき我慢の有無を観察してもらう．実験 2-2 では我慢後の表現を条件ごとに評価してもらう．

4.5.2 実験条件

実験 2-1 では，要因 B：表情における我慢の有無（B1：我慢なし，B2：10 秒間我慢あり）を被験者内要因，要因 D：エージェントとの関係性設定（D1：親しい関係，D2：他人の関係）を被験者間要因とする 2 要因 2 水準（2 × 2）の被験者間－被験者内混合計画とした．実験 2-2 では，要因 C：我慢後の表現（C1：不快を表す，C2：不快を表さない，C3：感情を偽る）を被験者内要因，要因 D：エージェントとの関係性設定（D1：親しい関係，D2：他人の関係）を被験者間要因とする 2 要因（2 水準× 3 水準）の 6 条件について，2-1 と同様に被験者間－被験者内混合計画とした．また，各条件の提示順はラテン方格法に基づいてカウンタバランスを考慮して決定した．

4.5.3 評価項目，評価方法

感情表出（Q1），感情抑制（Q2，Q3），ふるまいの自然さ（Q4），人間らしさ（Q5），好意度（Q7）に関する項目を設けた．ただし実験 2-1 は対話意欲（Q6），実験 2-2 は表現の適切さ（Q6-2）を設定した．各質問項目に対し，Visual Analogue Scale（VAS）法に従い，

0–99（0: 全くあてはまらない，99: 非常に当てはまる）で回答を求めた．

Q1 このエージェントは不快感を感じている

Q2 このエージェントは本当の感情を隠そうとしている

Q3 このエージェントは本当の感情が出ないようにしている

Q4 このエージェントは自然なふるまいをしている

Q5 このエージェントは人間らしいと感じる

Q6-1 このエージェントとこれからも交流をしたい

Q6-2 このエージェントの表現は今の関係性にふさわしい表現をしている

Q7 このエージェントに好感を覚える

4.5.4 結果 実験 2-1

各評価項目に対して，繰り返しのある分散分析を有意水準 $\alpha = 0.05$ で行った．図 5 に各評価項目の平均値およびその標準偏差を，表 2 に分散分析の結果を示す．

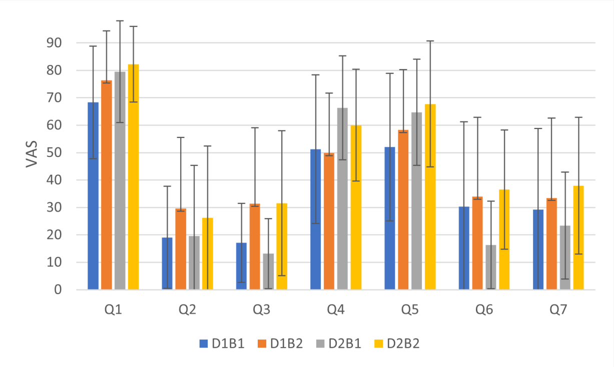


図 5: 実験 2-1 各評価項目の平均値および標準偏差

まず，感情抑制（Q2，Q3）について，まず Q2 では有意差が認められなかったが Q3 では要因 B の主効果が確認され，$B_2 > B_1$ が示された．このことから，エージェントが感情を抑制していることを認識させることが可能であることが示された．次に対話意欲（Q6-1）について，要因 B の主効果が確認され，$B_2 > B_1$ が示された．また，要因 B と要因 D の交互作用（要因 B × D）において有意傾向が確認され，D_2 条件（他人関係）において $B_2 > B_1$ が示唆された。このことから，感情の抑制表現を行うエージェントとこれからも交流を続けたいと感じさせることや，関係性が他人であるときに感情の抑制表現をすると，関係を継続したいと考えられる可能性が示唆された．好意度（Q7）について，要因 B の主効果が確認され，$B_2 > B_1$ が示された．一方，不快感（Q1）やふるまいの自然さ（Q4）と人間らしさ（Q5）について，有意差は認められなかった．このことから，感情の抑制を行うことが好意度を向上させる可能性が示された．以上の結果から，仮説 1 の一部が支持されたが，仮説 2 は支持されなかった．

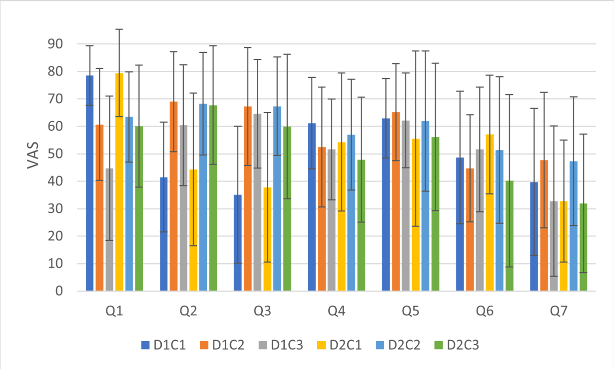


図 6: 実験 2-2 各評価項目の平均値および標準偏差

4.5.5 結果 実験 2-2

各評価項目に対して，繰り返しのある分散分析を有意水準 $\alpha = 0.05$ で行った．図 6 に各評価項目の平均値およびその標準偏差を，表 3 に分散分析の結果を示す．

まず，不快感（Q1）について，要因 C の主効果が確認され，$C_1 > \{C_2, C_3\}$ が示された．このことから，C1 がエージェントの不快感情をユーザに伝えられることが示された．次に感情抑制（Q2，Q3）について，Q2 と Q3 で要因 C の主効果が確認され，$\{C_2, C_3\} > C_1$ が示された．ふるまいの自然さ（Q4），人間らしさ（Q5），表現の適切さ（Q6-2），好意度（Q7）では有意差が確認されなかった．仮説 2 と仮説 3 は支持されなかったが，C2 と C3 が感情を抑制している表現であったと認識することが可能なことが示唆された．

5 考察

本研究では，エージェントが不快感情を適切に抑制・表出することで，ユーザとのインタラクションを促進させ，好意度を向上させるかを検証した．まず，ユーザが「我慢している（感情を抑制している）」と知覚する表情要件を選定し（実験 1），その結果に基づいた表情を実装したエージェントが，関係性に応じて抑制行動をとった際に，ユーザにどのような印象を与えるか

表 2: 実験 2-1 における分散分析および多重比較の結果

評価項目	要因 B（主効果）			要因 D（主効果）			交互作用 B × D		
	F 値	p 値	多重比較	F 値	p 値	多重比較	F 値	p 値	多重比較
Q1	1.855	.187	n.s.	1.923	.179	n.s.	0.449	.510	n.s.
Q2	2.947	.100	n.s.	0.134	.718	n.s.	0.072	.791	n.s.
Q3	7.855	$.010^{**}$	$B_2 > B_1$	0.083	.776	n.s.	0.123	.730	n.s.
Q4	0.513	.481	n.s.	2.934	.101	n.s.	0.218	.645	n.s.
Q5	1.005	.327	n.s.	1.852	.187	n.s.	0.116	.737	n.s.
Q6	8.896	$.007^{**}$	$B_2 > B_1$	0.367	.551	n.s.	4.191	$.053^{\dagger}$	$D_2 : B_2 > B_1$
Q7	6.349	$.020^{*}$	$B_2 > B_1$	0.005	.994	n.s.	1.864	.186	n.s.

$^{**}p < .01, ^{*}p < .05, ^{\dagger}p < .10, n.s. : notsignificant$

表 3: 実験 2-2 における分散分析および多重比較の結果

評価項目	要因 C（主効果）			要因 D（主効果）			交互作用 C × D		
	F 値	p 値	多重比較	F 値	p 値	多重比較	F 値	p 値	多重比較
Q1	15.985	$< .001^{**}$	$C_1 > \{C_2, C_3\}$	1.242	.277	n.s.	1.363	.266	n.s.
Q2	11.312	$< .001^{**}$	$\{C_2, C_3\} > C_1$	0.240	.593	n.s.	0.246	.783	n.s.
Q3	16.466	$< .001^{**}$	$\{C_2, C_3\} > C_1$	0.009	.925	n.s.	0.208	.813	n.s.
Q4	0.942	.397	n.s.	0.159	.694	n.s.	0.493	.614	n.s.
Q5	0.483	.620	n.s.	0.567	.459	n.s.	0.086	.918	n.s.
Q6	0.443	.645	n.s.	0.062	.806	n.s.	1.043	.361	n.s.
Q7	1.549	.224	n.s.	1.279	.270	n.s.	0.696	.504	n.s.

$^{**}p < .01, ^{*}p < .05, n.s. : notsignificant$

を検証した（実験 2）．なお，エージェントの抑制行動の実装にあたっては，ストレス値・我慢値・感情表出値・感情制御値の 4 パラメータからなる内部モデルを構築し，感情表出に遅延を発生させることで我慢を再現した．

実験 1 では，エージェントの表情刺激に対する評価を行った．その結果，Q1（不快感），Q3（抑制），Q4（自然さ），Q5（人間らしさ），Q6（我慢）において条件間に有意差が認められた．まず，仮説 1-i については，Q3 および Q6 の結果から支持された．特に，Q6（我慢）において混合表情条件（A5）が単一表情条件よりも有意に高く評価されたことは，ユーザが表情の中に含まれる感情の矛盾を「本音を隠そうとする意図」として読み取ることが可能であることを示している．Q5 の結果から，混合表情（A5）が単一の快表情等と比較して高く評価されたことで仮説 1-ii が支持された．これは，単一の感情のみならず，相反する感情が同時に存在する複雑な表情が，エージェントの内面や「心」の存在を感じさせた可能性があると考えられる．一方で，Q6 において当初想定していた「目と口で異なる感情を割り当てる（A1，A2）」条件よりも,「目と口のそれぞれで感情を混合させる（A5）」条件の方が高く評価されたため，仮説 2 は一部を支持する結果となった．この理由として，A1 と A2 のような顔の上下で明確に感情が分かれる表現は不自然さが際立つが，A5 のように各部位内で感情が混合している状態は,「感情を抑えようとする微細な葛藤」や「漏れ出る本音」としてリアリティのある「我慢」として知覚されたと考えられる．以上の結果から，A5 の表情が「我慢」を伝えるのに最も適切であると判断し，実験 2 の刺激として採用した．

実験 2-1 では，不快な状況における我慢（抑制）行動の有無が印象に与える影響を検証した．結果として，Q3（抑制），Q6（対話意欲），Q7（好意度）において我慢の有無（要因 B）の主効果に有意差が認められた．まず，B2 が B1 よりも Q6 および Q7 で高く評価されたことから，仮説 1-i および仮説 1-iii は支持された．これは，不快な出来事に対して即座に怒りを露わにせず，一旦，感情を飲み込んで抑制しようとするふるまいが，ユーザへの「配慮」や「社会的な礼儀正しさ」として好意的に解釈されたためと考えられる．また，Q6（対話意欲）において関係性との交互作用（BD）に有意傾向が見られた．具体的には，親しい関係（D1）よりも他人関係（D2）において，我慢の有無による評価差がある傾向にあった．これは，親しい間柄であれば多少の感情表出は許容されるが，関係の浅い「他人」であるほど社会的な距離感や礼儀が重視されるため，抑制行動が関係維持（対話継続）に対してより重要な役割を果たしたことを示唆している．一方で，Q5 には有意差が見られず，仮説 1-ii は支持されなかった．この要因として,「我慢する」という行動は，生物的な「人間らしさ」というよりも，性格的な「好ましさ」や「誠実さ」として評価された可能性が考えられる．もしくは，短時間のインタラクションであったため，人間らしさを評価するに至らなかった可能性もある．

実験 2-2 では，関係性に応じた我慢後の表現の違いによる印象の違いについて検証した．結果として，Q1・Q2・Q3 において要因 C の主効果に有意差があり，ユーザは表情の違いを正しく識別できていた．仮説 3 では，関係の浅い「他人」条件においては，ネガティブな感情を露わにするよりも社会的配慮として「笑顔による建前」を用いる方が好意的に評価されると予測していた．結果としては，Q5（人間らしさ）や Q7（好意度）では，要因間の有意な交互作用は見られなかった．特に C3 は関係性に関わらず一律に低く評価された．評価

が低かった理由として，人間同士のコミュニケーションであっても，明らかに不快な状況や拒絶すべき場面において「笑顔」を見せることは，必ずしも肯定的な「建前」としては機能していないと考えられる．むしろ，刺激のネガティブさと反応のポジティブさが矛盾している状態であり，状況を正しく理解していないような「不誠実さ」を感じさせた可能性がある． 本実験においても，不快刺激に対する C3 の反応は，社会的配慮というよりも，状況に即していない不自然な反応として知覚され，その結果として好意度の低下を招いた可能性がある．無理に偽りの表情を表出するよりも，フラットな態度を保つ方が，不誠実さを感じさせず，無難な対応として受け入れられたと考えられる．以上より，エージェントが社会的ふるまいを行う際は，単にポジティブな表情を提示すればよいわけではなく，実験 2-1 で示されたような「葛藤を含んだ抑制」を見せることが，ユーザとの信頼関係構築において重要となる可能性が示されている．

6 おわりに

本稿では，エージェントに対するハラスメント等の不快な状況において，関係性を破綻させずに拒絶の意思を伝えるため，日本的な「本音と建前」および「我慢」の概念を取り入れたエージェントを実装した．エージェントは，ストレス値・我慢値・感情表出値・感情制御値の 4 つのパラメータに基づき，不快刺激に対して即座に反応せず，閾値を超えるまでは「我慢」を保ち，限界を超えた時点で初めて「本音」を表出するという，時間的・心理的な葛藤プロセスを表現した．

これらのエージェントの感情表現がユーザに与える効果を検証した結果，一つの顔部位に異なる感情を表現させた混在表情は，ユーザにエージェントの内面的な葛藤や本音を隠そうとする意図を知覚させた．そして，不快刺激に対して即座に怒りを露わにせず，我慢のふるまいを見せるエージェントは，即座に怒りを表出するエージェントと比較して，ユーザの対話意欲や好意度を高めた．一方で，不快な状況下での安易な笑顔の提示は不誠実と捉えられる傾向が示された．このことから，葛藤を含んだ抑制表現が，ユーザへの社会的配慮として機能し，信頼関係の維持に寄与する可能性が考えられる．また，関係性の設定による差は示されなかったが，テキストのみでエージェントとの関係性を提示したことも影響した可能性がある．今後は実験における関係性の設定方法の再検討や，実際にインタラクションを通じた長期的な影響などの検証を行っていきたい．

謝辞

本研究は一部科研費 23K11278, 23K11202, 24K02977, 22K19792 の助成を受け実施した．

参考文献

[1] 塩見昌裕, 境くりま, 船山智, 港隆史, 石黒浩. ロボハラ (ロボットハラスメント) 人間酷似型ロボットに対する人間からのハラスメント事例. 人工知能学会全国大会論文集 第 39 回 (2025), pp. 3Q6GS801–3Q6GS801. 一般社団法人 人工知能学会, 2025.

[2] Naoto Yoshida and Tomoko Yonezawa. “honne and tatemae”: Expression of the agent’s hidden desire through physiological phenomena. *The Transactions of Human Interface Society*, Vol. 26, No. 2, pp. 249–258, 2024.

[3] Robert R Holt. On the interpersonal and intrapersonal consequences of expressing or not expressing anger. 1970.

[4] 木野和代. 日本人の怒りの表出方法とその対人的影響. 心理学研究, Vol. 70, No. 6, pp. 494–502, 2000.

[5] 木野和代. 対人場面における怒りの表出方法の適切性・効果性認知とその実行との関連. 感情心理学研究, Vol. 10, No. 2, pp. 43–55, 2004.

[6] 吉田琢哉. 感情表出に対する態度が親密で対等な二者関係における怒り感情制御方略および親密度に及ぼす影響. 社会心理学研究, Vol. 26, No. 3, pp. 211–218, 2011.

[7] 倫太今井. インタラクションの認知科学. 認知科学のススメ. 新曜社, 2018. 日本認知科学会 監修.

[8] 笹屋里絵. 表情および状況手掛りからの他者感情推測. 教育心理学研究, Vol. 45, No. 3, pp. 312–319, 1997.

[9] 藤村友美, 鈴木直人. 表情の表出過程および形態学的変化が感情認識に及ぼす影響：次元的観点に基づいた表情による検討. 認知心理学研究, Vol. 5, No. 1, pp. 53–61, 2007.

[10] 後藤みの理, 加納政芳, 加藤昇平, 國立勉, 伊藤英則. 感性ロボットのための感情領域を用いた表情生成. 人工知能学会論文誌, Vol. 21, No. 1, pp. 55–62, 2006.

[11] Paul Ekman and Wallace V Friesen. Facial action coding system. *Environmental Psychology & Nonverbal Behavior*, 1978.

[12] マーヴィン・ミンスキー, 安西祐一郎. 心の社会. 安西祐一郎訳, 東京: 産業図書, 1990.

[13] Jeff T Larsen and A Peter McGraw. The case for mixed emotions. *Social and Personality Psychology Compass*, Vol. 8, No. 6, pp. 263–274, 2014.

[14] 加納政芳, 後藤みの理, 加藤昇平, 中村剛士, 伊藤英則. ロボットの混合感情表出のための表情制御手法 (¡ 特集¿ エンタテインメントコンピューティング). 知能と情報, Vol. 17, No. 2, pp. 250–255, 2005.

[15] 目良和也, 青山正人, 大道博文, 黒澤義明, 竹澤寿幸. 顔部位の変化の不随意性に基づく本心感情を隠蔽している表情アニメーションの生成ーツンデレ表情を事例としてー. 知能と情報, Vol. 32, No. 6, pp. 944–955, 2020.

[16] 曹蓮, 杉森伸吉. 日本人は表情から本音と建前の感情をどのように認知するのか?ー喜びと怒りの段階的モーフィング表情画像を用いた検討ー. 感情心理学研究, Vol. 28, No. Supplement, pp. ps03–ps03, 2020.

[17] Miyuki Kamachi, Vicki Bruce, Shigeru Mukaida, Jiro Gyoba, Sakiko Yoshikawa, and Shigeru Akamatsu. Dynamic properties influence the perception of facial expressions. *Perception*, Vol. 42, No. 11, pp. 1266–1278, 2013.

[18] 大道博文, 林柚季, 目良和也, 黒澤義明, 竹澤寿幸. Action unit の組み合わせを用いた快感情抑圧表現アニメーションの生成. 人工知能学会全国大会論文集 第 33 回 (2019), pp. 4Rin114–4Rin114. 一般社団法人 人工知能学会, 2019.

[19] Cynthia Breazeal. Emotion and sociable humanoid robots. *International journal of human-computer studies*, Vol. 59, No. 1-2, pp. 119–155, 2003.

[20] 目良和也, 市村匠, 黒澤義明, 竹澤寿幸. 情緒計算手法と心的状態遷移ネットワークを用いた音声対話エージェントの気分変化手法. 知能と情報, Vol. 22, No. 1, pp. 10–24, 2010.

[21] Edmund T Hall and Edward T Hall. *The hidden dimension*, Vol. 609. Anchor, 1966.

[22] 八重澤敏男, 吉田富二雄. 他者接近に対する生理・認知反応-生理指標・心理評定の多次元解析. 心理学研究, Vol. 52, No. 3, pp. 166–172, 1981.

[23] Andrea Bönsch, Sina Radke, Heiko Overath, Laura M Asché, Jonathan Wendt, Tom Vierjahn, Ute Habel, and Torsten W Kuhlen. Social vr: How personal space is affected by virtual agents' emotions. In *2018 IEEE conference on virtual reality and 3D user interfaces (VR)*, pp. 199–206. IEEE, 2018.

[24] Wataru Sato and Sakiko Yoshikawa. Brief report the dynamic aspects of emotional facial expressions. *Cognition and Emotion*, Vol. 18, No. 5, pp. 701–710, 2004.