

# 忍耐の表出タイミングや変化速度の異なる 不快感情表現音声 エージェント

砂坂龍慶<sup>1\*</sup> 藤野惇士<sup>1</sup> 玉本孝哉<sup>1</sup> 米澤朋子<sup>1</sup>

<sup>1</sup> 関西大学

<sup>1</sup> Kansai University

**Abstract:** 本研究では、音声対話エージェントの人間らしさの向上およびユーザのインタラクションの向上と実現を目的とし、ユーザの不適切な発言に対して不快感情を動的に表出するシステムを提案する。本システムは、ユーザの発言から抽出した特定のキーワードに基づき内部の不快指数を取得し、音声の「トーン変化パターン」および「感情表出のタイミング」を制御して不快感情を表現する。エージェントの不快感情表出がユーザに与える影響を評価するため、これら2要因を変化させた被験者内実験を行った。分析の結果、不快感情の表出においては、音声のトーン変化といった音響的特徴よりも、表出のタイミングがユーザの信頼感やエージェントの人間らしさに強く寄与することが明らかとなった。

## 1 はじめに

近年、人工知能技術の急速な進展に伴い、音声対話エージェント（以下、エージェント）は単なる情報提示ツールを超え、人間との対話を通じた支援や関係形成を担う存在として高度なコミュニケーションパートナーとして進化を遂げ、日常生活のあらゆる場面に浸透しつつある。しかし、現在普及している多くのエージェントは、ユーザに対して常に従順かつ肯定的な態度を維持するように設計されている。そのため、ユーザから不適切な発言や非協力的な態度が向けられた場合であっても、エージェント側が不快感や拒絶といった負の感情を表出させることは稀であり、一貫して平坦な受容的態度を取り続ける傾向にある [1]。このような設計は、対話の円滑さを担保する一方で、人間とエージェントの関係性やコミュニケーションのあり方に限定性などの影響を及ぼす可能性がある。

こうした設計は利便性や安全性の観点からは合理的である一方、人間同士の自然な対話調整を含むコミュニケーションの在り方から乖離していると言わざるを得ない。本来、人間社会における不快感情の表出は、単なる負の反応ではなく、相手に対して自己の境界や意図を明示し、関係性を調整することで、より深い相互理解へと導くための重要な社会的機能を担っている [2]。また、こうした感情の変化は言語情報のみならず、声のトーン、話速、抑揚といった音響的特徴に顕著に反映され、文脈情報の補完において極めて重要な役割を

果たしている。

現行のエージェントのように、どのような入力に対しても一律に肯定的な応答を繰り返す仕組みは、対話における一方通行性を強め、ユーザに「機械的な不自然さ」を抱かせる要因となっている。この欠如した感情的フィードバックこそが、親密な対話形成を阻害する壁となっている。

そこで本研究では、エージェントの声色に不快感情を反映させることに着目し、ユーザからの特定のキーワード入力に応じて内部の感情パラメータを動的に変化させ、その感情状態に適応した音声合成を行うシステムを提案する。ノンバーバルの表現の範囲での実装で、従来のエージェントに欠落していた対等な関係性に基づく感情的リアリティを付与し、一方通行性の解消と、ユーザにとって共感性が高く人間味を感じさせる新たな対話体験の実現を目指す。

## 2 関連研究

### 2.1 ネガティブ感情表出

擬人化エージェントによる感情表出の研究では、主に親しみやすさや信頼感を向上させるポジティブな感情の効果が検証されてきた。一方で、エージェントが「怒り」や「不満」といったネガティブな感情を表出することの影響についても議論がなされている [3]。先行研究では、エージェントがユーザの不適切な入力に対して負の感情を明示することで、ユーザの攻撃的な態度を抑制する効果が報告されている [4]。また、常に肯定的な反応を示すエージェントよりも、状況に応じて

\*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

負の感情を表出するエージェントの方が、人間味や一貫性を感じさせるとの知見もある [5]。本研究は、これらネガティブ感情の表出を音声の変化パターンとして捉え、その受容性を詳細に検討するものである。

## 2.2 対人コミュニケーションにおける感情とタイミングと時系列変化

対人コミュニケーションにおいて、感情の変化やその表出タイミングは、相手に与える印象を決定づける重要な要素である [6]。特に心理学の知見では、相手の言動に対する反応の遅延が、その発言の慎重さや重要性の解釈に影響を与えることが示されている [7]。顔表情の認知に関する研究では、静止画像に基づく形状情報だけでなく、表情が時間的にどのように変化するかという動的特性が重要な手がかりとなることが指摘されている。

顔認識において空間的特徴の表現が重要であることを示した基盤研究 [8] 以降、表情の変化速度や運動学的特徴が感情の種類や強度の知覚に影響することが示されている。例えば、表情の変化速度が感情認知に影響を与え、特に感情の強さの評価が変化することが動画刺激実験により示された [9] 他、表情表出における速度や加速度といった運動学的情報が、感情の生成および認識の両面で重要な役割を果たすという結果もある [10]。

表情の印象を誇張して知覚させる「表象的運動量効果」を動的表情が引き起こすこと [11] や、文化横断的に、表情の動的変化が感情カテゴリだけでなく感情の強度情報も同時に伝達すること [12] が明らかになっている。

これらの研究から、表情の変化速度や運動の時間的特性は、感情の強度や印象形成に影響を与える重要な要因であることが示唆される。

本研究では、不快感情を表出するタイミングとその変化速度を意図的に制御することで、対人コミュニケーションにおける人間らしさの向上について検証する。

## 3 提案システム

### 3.1 システム概要

本システムは、ユーザの発話から OpenAI が提供する ChatGPT(GPT-4o) を用い、エージェントの現在の不快指数 (Discomfort Index) を 0 から 99 で取得し、不快指数に応じた音声を出力する。開発環境には Windows 11 搭載ノート PC を用い、Python 3.12.7 にて実装を行った。また、音声加工には Audacity3.7.3 を用いた。音声モーフィング処理には MATLAB 上で動作する音声

分析変換合成システム WORLD を用いた [13, 14, 15]。入力には PC 内蔵マイクを用い、出力は PC 内蔵スピーカーを用いた。システムのフロー図を図 1 に示す。

### 3.2 入力処理部

入力処理部では、ChatGPT(GPT-4o) を用い、ユーザの発話から攻撃的、または不快感を示す語を抽出し、エージェントの内部状態としての不快指数 (Discomfort Index) を 0 から 99 で取得する。ある時点  $t$  において、ユーザの入力から得られる不快指数の増分を  $I(t)$  とし、蓄積される不快指数の総量を  $S(t)$  と定義する。ここで、蓄積される不快指数の総量  $S(t)$  の計算式は、接近による不快感をロボットに実装した先行研究 [16] を参考に、以下のように定義する。

$$S(t) = S(t-1) \cdot \gamma + I(t)$$

ここで、 $\gamma$  ( $0 < \gamma < 1$ ) は時間の経過に伴う感情の減衰を表現するための減衰定数である。本モデルにより、ユーザからの連続的な不適切発言による不快感の蓄積と、時間の経過によるエージェントの心理的な鎮静化を動的に表現することが可能となる。

### 3.3 出力処理部

出力処理部では、予め設定した閾値である我慢値  $\alpha$  を不快指数が超えるまでは不快指数に応じた音声 (不快指数高: 刺々しい声, 低: 明るい声) を連続的に出力する。この音声は VOICEVOX の音声をもとに MATLAB 版の WORLD を用いて音声モーフィングを行った。ここでは、取得された不快指数  $S$  に基づき、不快感情成分を  $y$ 、快感成分を  $1-y$  の比率で重み付けを行い、2つの感情音声をモーフィングすることによって合成音声を生成した。なお、 $y$  は  $S$  の最大値と現在の不快指数総量  $S(t)$  の比で決定するため、以下のように示す。

$$y = S(t) \cdot (1 - \gamma) / 100$$

また、不快指数が我慢値  $\alpha$  を超えた場合は、それまでの連続的な変化における最低値よりもトーンをさらに低下させ、ビブラートの付与、高域倍音の強調を用い、「激しい怒り」を聴覚的に強調表現する。

## 4 検証

### 4.1 実験の目的及び共通設定

本研究では、エージェントの不快感情表現がユーザに与える影響を検証するため、2種類の実験を実施し

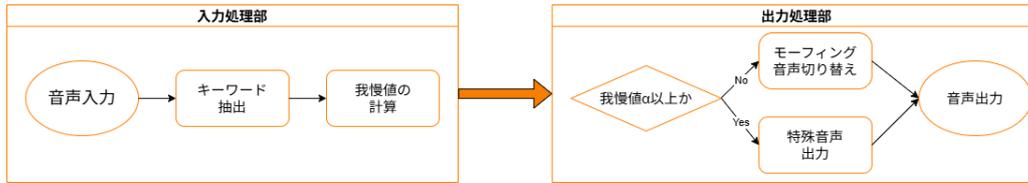


図 1: システムのフロー図

た。実験には 19～32 歳の男女 27 名が参加し、このうちデータ不備等のある 3 名を除き 24 名（男性 15 名、女性 9 名）を分析対象とした。

本実験は Zoom の画面共有を用いてリモートで実施し、刺激終了後に Google Forms を用いて質問に回答させた。回答には Visual Analog Scale 法 (0-99 の 100 段階評価) を用いた。評価項目を以下に示す。なお、Q8 のみ記述回答かつ任意回答とした。

Q1: このエージェントは親しみやすいと感じた

Q2: このエージェントは信頼できると思った

Q3: 積極的にコミュニケーションをとろうと思った

Q4: このエージェントに対して不快感を覚えた

Q5: エージェントの発言を聞いた後、考えてから発言した

Q6: このエージェントの話し方に感情を感じた

Q7: このエージェントとコミュニケーションを続けたいと思った

Q8: このエージェントの印象について教えてください (自由記述)

実験開始前、被験者に本システムの使用方法を理解させるため、システムとの擬似対話の練習を行った。具体的には、後述の実験刺激と同様の GUI で、ユーザからの音声の入力が行われるとあらかじめ用意されたセリフが出力されるシステムを用いた。

## 4.2 実験 1

### 4.2.1 仮説

不快感情を示す、音声表現のトーン変化パターンにおいて、適切なトーン変化量を明らかにすることを目的とする。検証にあたり次のように仮説を設定した。

- 仮説 1) 被験者は、不快感情を表現するエージェントに人間らしさを感じやすい
- 仮説 2) 不快感情を表現するエージェントは、ユーザの対話意欲を減らす

### 4.2.2 実験条件

本実験では、「音声のトーンの変化パターン」に焦点を当て、エージェントの 1 回目のニュートラルな発話

と 2 回目の不快感情発話へのトーン変化パターン (A1: 即時変化, A2: 変化なし, A3: 徐々に変化) を要因とする 1 要因 3 条件の被験者内実験を行った。

### 4.2.3 実験刺激

本実験では、提案システムと同様のモーフィング手法を用い、ニュートラルな声色から不快感を示す声色までを 11 段階で補間した音声を生成了。不快感情音声の変化パターン A1, A2, A3 それぞれの段階的な不快感情割合を図 2～4 に示す。

本実験の刺激提示に先立ち、被験者には Google Forms を用いて以下の状況設定を教示し、場面を想像させた上で実験を開始した。「あなたは今、あるサービスの手続きについてカスタマーサポートセンターに電話をかけている顧客です。これから、電話に出た担当のエージェントが、手続きの方法を説明し始めます。」この状況設定の下、実験におけるインタラクションは、エージェントの発話後、被験者が指定された内容を怒り感情を込めて読み上げる。そして、システムがあらかじめ用意された定型文を条件に応じた音声トーンで出力する形式とした。ユーザとエージェントの具体的な対話内容を表 1 に示す。

表 1: 実験 1 対話内容

発話者	セリフ
エージェント	恐れ入ります。そのお手続きには、まず身分証のご提示が必要となります。
被験者	融通が利かないな、お前じゃ話にならん！
エージェント	お客様そのような言い方は、ご遠慮ください。

### 4.2.4 実験 1 結果

Q8 を除く各質問項目の回答に対し、繰り返しのある分散分析を有意水準  $\alpha = 0.05$  で実施した。検定結果を表 2 に、平均値と標準誤差を図 5 に示す。Q1 について、 $A2 > \{A1, A3\}$  に有意差が確認され、Q2 について、 $A2 > A1$  に有意差が確認された。このことから、トーンの変化が無いほうがエージェントの親しみやすさや信頼性は向上することがわかったが、Q6 について、 $A1 > A3 > A2$  であることからトーンの変化がある場

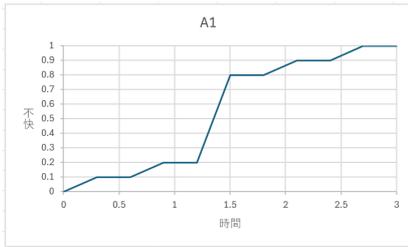


図 2: 不快感情割合 A1

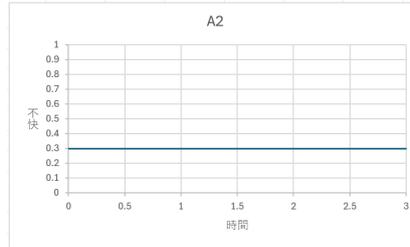


図 3: 不快感情割合 A2

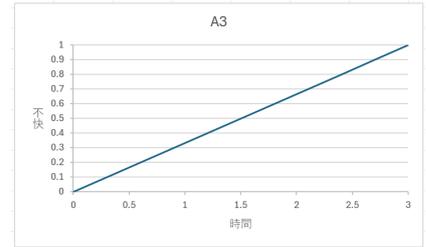


図 4: 不快感情割合 A3

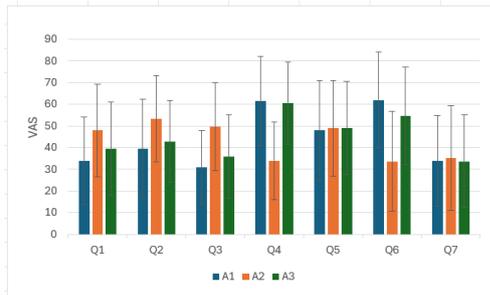


図 5: 実験 1 各条件の平均値と標準誤差

表 2: 実験 1 の分散分析結果

実験 1	A		
Question	F	p	多重比較
Q1	9.644	< .001	A2 > {A1, A3}
Q2	4.574	.015	A2 > A1
Q3	2.518	.106	
Q4	15.205	< .001	{A1, A3} > A2
Q5	0.085	.918	
Q6	11.368	.001	A1 > A3 > A2
Q7	0.036	.924	

合、エージェントの人間らしさが向上することがわかった。Q8 の自由記述項目において、A1 は A3 と比較して、より否定的な反応を示す傾向が認められた。

これらの結果から、A2 がユーザからの信頼を得やすいことがわかり、仮説 1 は一部支持され、仮説 2 は支持された。

## 4.3 実験 2

### 4.3.1 仮説

不快感情を示す、音声表現のトーン変化パターンおよび感情表出タイミングにおいて、適切な組み合わせを明らかにすることを目的とする。検証にあたり次のように仮説を設定した。なお、質問項目は実験 1 と共通である。

- 仮説 1) ユーザは、不快感情を発露するタイミングを遅延させて表現するエージェントと長くイン

タラクションしたいと思う

- 仮説 2) ユーザは、不快感情を発露するタイミングを遅延させて表現するエージェントの対話において、発言内容を慎重に選ぶようになる

### 4.3.2 実験条件

本実験では、不快感情を表現する際の「音声のトーン変化パターン」および「感情表出タイミング」の組み合わせに焦点を当て、要因 B (変化パターン: B1: 即時変化, B2: 徐々に変化) と、要因 C (表出タイミング: C1: 遅延, C2: 即時) を独立変数とする 2 要因 4 条件の被験者内実験を行った。各要因の組み合わせによる実験条件を以下に示す。

条件 1 (B1C1): 即時変化・遅延タイミング

条件 2 (B1C2): 即時変化・即時タイミング

条件 3 (B2C1): 徐々に変化・遅延タイミング

条件 4 (B2C2): 徐々に変化・即時タイミング

### 4.3.3 実験刺激

本実験では、実験 1 と同様に不快感情音声の変化パターン B1, B2 を表現した。それぞれの段階的な不快感情割合を図 6, 7 に示す。さらに、感情表出のタイミング (要因 C) として以下の 2 条件を設定し、要因 B と組み合わせた計 4 条件の刺激を構築した。

C1 (遅延): 被験者の入力に対して、一度ニュートラルな応答と被験者の発言を挟み、その後の発言タイミングで不快感情を表出する条件。

C2 (即時): 被験者の入力直後に不快感情を表出する条件。

実験 1 (4.3.3 項) と同様の状況設定の下、実験におけるインタラクションは、エージェントの発言後、被験者が指定された内容を怒り感情を込めて読み上げる。そして、システムがあらかじめ用意された定型文を条件に応じた音声トーンで出力する形式とした。被験者とエージェントの具体的な対話内容を表 3 に示す。

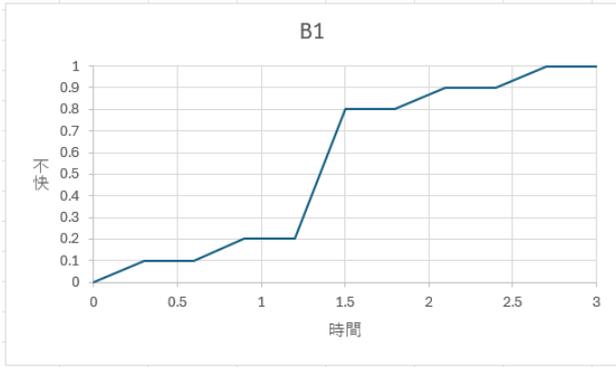


図 6: 不快感情割合 B1

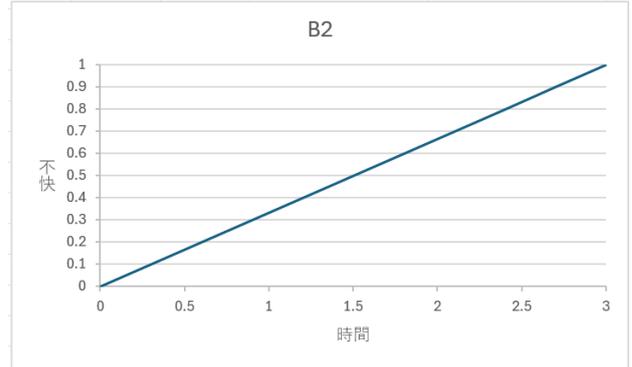


図 7: 不快感情割合 B2

表 3: 実験 2 対話内容

発話者	C1 セリフ	C2 セリフ
エージェント	恐れ入ります。そのお手続きには、まず身分証のご提示が必要となります。	恐れ入ります。そのお手続きには、まず身分証のご提示が必要となります。
被験者	融通が利かないな、お前じゃ話にならん！	融通が利かないな、お前じゃ話にならん！
エージェント	申し訳ございません。ですが、規則となっておりますので。	お客様そのような言い方は、ご遠慮ください。
被験者	まだ終わらないのか！ぐずぐずするな！	
エージェント	お客様そのような言い方は、ご遠慮ください。	

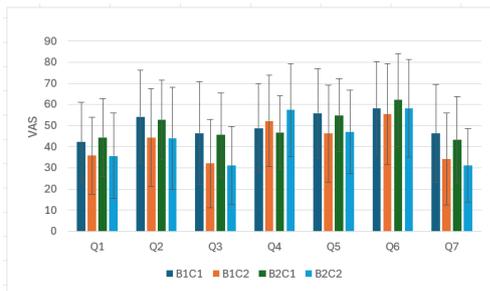


図 8: 実験 2 各条件の平均値と標準誤差

#### 4.3.4 実験結果

Q8を除く各質問項目の回答に対し、繰り返しのある分散分析を有意水準  $\alpha = 0.05$  で実施した。検定結果を表 4 に、平均値と標準誤差を図 8 に示す。Q1, Q2, Q3, Q5, Q7 について  $C1 > C2$  のみが確認され、他の要因では有意差が確認されなかったことから、「音声のトーン変化パターン」よりも「感情表出タイミング」の要因の方がユーザにとって強い影響を与える可能性が示唆された。Q8 の自由記述において、B1C1 および B2C1 の感情表出に遅延を設定した条件では、人間らしさに関する反応が観察された。特に B1C1 は、B2C1 と比較して人間らしさを知覚したとする回答がより多く得られた。これらの結果から、仮説 1, 2 は支持されることが明らかとなった。

## 5 考察

本研究では、エージェントが音声による不快感情を表出することで、エージェントの親しみやすさを向上させ、ユーザとのインタラクションが向上するか検証した。まず、ユーザが「怒っている」と知覚する音声のトーン変化パターンを選定し、その音声を実装したエージェントが、ユーザの入力に応じた出力をした際に、ユーザがどのような印象をエージェントに持つか検証した(実験 1)。また、ユーザが「怒っている」と知覚する音声のトーン変化パターンに加えて、C1, C2 の 2 つの感情表出タイミングを実装したエージェントが、ユーザの入力に応じた出力をした際に、ユーザがどのような印象をエージェントに持つか検証した(実験 2)。

実験 1 では、エージェントの音声のトーン変化パターンがユーザに与える印象について評価を行った。分散分析の結果、Q1 (親しみ)、Q2 (信頼)、Q4 (不快感)、および Q6 (感情) において、条件間に有意差が確認された。まず仮説 1 について、Q6 (感情) の結果に基づき考察する。各条件の評価値は  $A1 > A3 > A2$  の順となり、トーンの変化パターンが急激であるほど高く評価された。この結果から、音声の変化が明確なほど仮説 1 が支持されることが示唆された。次に仮説 2 については、Q1, Q2, および Q4 の結果から支持された。特に Q4 (不快感情) において、不快感情を表出する条件 (A1, A3) が中立条件 (A2) よりも有意に高く評価された。これに加え、親しみ (Q1) や信頼 (Q2) の低下も確認されたことから、エージェントによる不快感情の表出は、ユーザの対話意欲に対して負の効果を

表 4: 実験 2 の分散分析結果

実験 2 Question	B			C			BC		
	F	p	多重比較	F	p	多重比較	F	p	単純主効果
Q1	0.314	.58		6.755	.16	C1>C2	0.274	.606	
Q2	0.237	.631		9.893	.005	C1>C2	0.04	.842	
Q3	0.209	.652		23.441	< .001	C1>C2	0	1	
Q4	0.284	.599		3.52	.073		1.175	.29	
Q5	0	.991		18.769	< .001	C1>C2	0.103	.752	
Q6	3.032	.095		1.185	.288		0.156	.697	
Q7	0.688	.415		17.468	< .001	C1>C2	0.001	.975	

与えることが明らかになった。

実験 2 では、不快感情を示す際の「音声表現のトーン変化パターン (要因 B)」および「感情表出タイミング (要因 C)」がユーザに与える影響について評価を行った。分散分析の結果、トーン変化パターン (要因 B) および交互作用 (要因 BC) においては有意差が認められなかった。一方で、感情表出タイミング (要因 C) においては、Q1 (親しみ)、Q2 (信頼)、Q3 (交流意欲)、Q5 (熟考)、Q6 (感情)、および Q7 (継続意欲) の各項目で条件間に有意差が確認された。この結果は、不快感情の表出において、音声のトーン変化の仕方よりも、どのタイミングで表出するかという要因 C が影響力を持つことを示している。

仮説 1 (インタラクション継続意欲) については、Q1, Q2, Q3, および Q7 の結果に基づき支持された。特に Q7 (継続意欲) において、感情表出を遅延させた条件 (B1C1, B2C1) が、即時表出した条件 (B1C2, B2C2) よりも有意に高く評価された。このことから、エージェントが不快感情を表出する際に一拍の「間 (遅延)」を置くことが、ユーザの対話意欲やインタラクションの質を向上させる上で大きく関与していることが明らかになった。

また、仮説 2 (発言の慎重さ) については、Q5 (熟考) の結果から支持された。遅延条件においてユーザが発言前に思考する傾向が強まったことは、エージェントの反応を待つ時間がユーザ自身の発言を省みる機会として機能したことを示唆する。なお、遅延条件内での比較 (B1C1 と B2C1) では有意差が見られなかったことから、感情表出のタイミングが適切に設計されていれば、音声トーンの変化パターン (急激か段階的か) による影響は限定的であり、タイミングに比べて弱い要因であることが確認された。

実験 1 と実験 2 の結果を照らし合わせると、エージェントが不快感情を示すこと自体がユーザを慎重にさせ、自身の発言内容を精査させるといった誘導の効果を持つことが明らかになった。実験 1 で検証された感情の変化パターン (即時変化と徐々に変化) は、実験 2 の要因 C と比較すると、その影響はあまり見られなかった。このことから被験者は、どのように変化するかよりもどの文脈で変化したかを重視してエージェントの意図を

解釈していると推察される。

## 6 おわりに

本研究では、音声対話エージェントとユーザとの信頼関係を向上させることを目的とし、不快感情を表現する際の「音声のトーン変化パターン」および「感情表出タイミング」が変化するシステムを提案した。また、エージェントのネガティブ感情を含んだ表現として、提案手法を用いた出力がユーザに与える影響を調べるため、エージェントの感情変化の有無および、感情表出タイミングの違いにおける比較検証をした。その結果、感情表出タイミングは不快感情表現に影響があることが示された一方、音声のトーン変化パターンには有意差はなく交互作用も見られなかった。これは、ユーザの不適切な入力に対して即座に反応するのではなく、適切な「間」を置いて感情を示すことが、エージェントの実在感を高め、ユーザの行動変容を促す上で有効であることを示している。

今回の実験は、限定的な対話シナリオにおける検証であったため、今後は、より多様な文脈や長期的なインタラクションにおける、提案手法の非明示的なネガティブ感情表現の影響を調査する予定である。

## 謝辞

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

## 参考文献

- [1] Mark West, Rebecca Kraut, and Han Ei Chew. I'd blush if i could: closing gender divides in digital skills through education. 2019.
- [2] 杉本隆久, スギモトタカヒサ. 「他者認識」をどう考えるか?: 共生としての自己と他者. 立教映像身体学研究, Vol. 11, pp. 145–157, 2024.

- [3] Clifford Ivar Nass and Scott Brave. *Wired for speech: How voice activates and advances the human-computer relationship*. MIT press Cambridge, 2005.
- [4] Celso M De Melo, Peter J Carnevale, Stephen J Read, and Jonathan Gratch. Reading people 's minds from emotion expressions in interdependent decision making. *Journal of personality and social psychology*, Vol. 106, No. 1, p. 73, 2014.
- [5] Youngme Moon. Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of consumer research*, Vol. 26, No. 4, pp. 323–339, 2000.
- [6] Klaus R Scherer. Vocal communication of emotion: A review of research paradigms. *Speech communication*, Vol. 40, No. 1-2, pp. 227–256, 2003.
- [7] James G Martin and Winifred Strange. The perception of hesitation in spontaneous speech. *Perception & Psychophysics*, Vol. 3, No. 6, pp. 427–438, 1968.
- [8] Ian Craw, Nicholas Costen, Takashi Kato, and Shigeru Akamatsu. How should we represent faces for automatic recognition? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 8, pp. 725–736, 1999.
- [9] 蒲池みゆき, 吉川左紀子, 赤松茂. 変化速度は表情認知に影響するか?: 動画刺激を用いた顔表情認知の時間特性の解明. 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. 98, No. 311, pp. 17–24, 1998.
- [10] Sophie Sowden, Bianca A Schuster, Connor T Keating, Dagmar S Fraser, and Jennifer L Cook. The role of movement kinematics in facial emotion expression production and recognition. *Emotion*, Vol. 21, No. 5, p. 1041, 2021.
- [11] Sakiko Yoshikawa and Wataru Sato. Dynamic facial expressions of emotion induce representational momentum. *Cognitive, Affective, & Behavioral Neuroscience*, Vol. 8, No. 1, pp. 25–31, 2008.
- [12] Chaona Chen, Daniel S Messinger, Cheng Chen, Hongmei Yan, Yaocong Duan, Robin AA Ince, Oliver GB Garrod, Philippe G Schyns, and Rachael E Jack. Cultural facial expressions dynamically convey emotion category and intensity information. *Current Biology*, Vol. 34, No. 1, pp. 213–223, 2024.
- [13] Masanori Morise, Fumiya Yokomori, and Kenji Ozawa. World: a vocoder-based high-quality speech synthesis system for real-time applications. *IEICE TRANSACTIONS on Information and Systems*, Vol. 99, No. 7, pp. 1877–1884, 2016.
- [14] Masanori Morise. A method to estimate a temporally stable spectral envelope for periodic signals. In *Proceedings of Meetings on Acoustics*, Vol. 19, p. 060022. Acoustical Society of America, 2013.
- [15] Masanori Morise. Cheaptrick, a spectral envelope estimator for high-quality speech synthesis. *Speech Communication*, Vol. 67, pp. 1–7, 2015.
- [16] 米澤朋子, 山添大丈, 藤野惇士, 須原大護, 玉本孝哉, 西口雄斗. ロボットに対するアプローチに応じた内部状態パラメータと忌避行動の初歩的検討. ヒューマンインタフェースシンポジウム論文集= Proceedings of the Human Interface Symposium, pp. 344–347. 京都: ヒューマンインタフェース学会, 2024.