

「あアラウド法」における感情観測手法の改善と 分析手法の検討

高山 英里^{1,a)} 能丸 天志² 安中 勇貴¹ 山岸 丈留¹ 渡邊 恵太^{2,b)}

概要: 「あアラウド法」は、実験参加者が「あ」のみを発声することで、実験参加者のリアルタイムな感情を観測する手法である。この手法は、明確な評価基準を定義しにくいエンタテインメントコンテンツの評価に役立つ可能性がある。一方で、発声方法や感情分析の妥当性については、十分な検討がなされていない。そこで本研究では、「あアラウド法」を感情観測手法として確立するため、発声方法が結果に与える影響と、感情分析の実行可能性についての実験を行い、適した実験設計と有用性について考察した。

1. はじめに

エンタテインメントコンピューティング (EC) 分野には、システムの有用性や改善点の評価手法を確立できていないという課題がある。この課題に対して水口ら [1] は、「心の動き」からエンタテインメントシステムを評価する枠組みを提案した。そこで川島ら [2] は、実験参加者のシステム体験中の感情をリアルタイムに得る手法として、思考発話法 [3] を応用した「あアラウド法」を提案した (図 1)。

あアラウド法は、システムを体験する実験参加者に「あ」のみを発声させ、その声の大きさや抑揚から感情を観測する手法である。実験の結果、あアラウド法によってタスク中に思考を言語化する負担と反応性^{*1}を思考発話法よりも低減すること、実験参加者の「あ」の発声と感情に関連性があることを確認した。そしてこの結果から、実験参加者の言語化負担が小さい、実験参加者の突発的な感情を直接把握できる可能性があるなど、あアラウド法の有用性を考察した。

一方で、あアラウド法を感情観測手法として確立するには様々な議論が不足している。主な論点は、あアラウド法に適した「あ」の発声方法と、あアラウド法で実験者が感情を正しく分析できるのかが明らかでないことである。

第一に、あアラウド法に適した「あ」の発声方法が明らかでない。「あ」の発声方法は、感情が生じたタイミングで発声する方法と、感情の生起に関わらず「あー」と発



図 1 あアラウド法では、実験参加者がタスク中に感情を「あ」のみで表現する

声し続ける方法がある。1 回ごとの発声が長いほど、実験参加者が微細な感情変化を表現しやすくなる可能性があるが、実験参加者の身体的な負担は増加する。

第二に、「あ」の発声から実験者が正しく分析できるのかが明らかでない。川島らは発声と感情の関連付けの際、発声とタスクの状況に加えて、実験参加者への聞き取り内容も参考にしており、実験記録のみから感情の推定が行えることの検証は行っていない。また、上述の発声方法の違いによって、実験者が実験参加者の感情を推定する精度が変化する可能性がある。

そこで本研究では、各論点に対応する 2 つの実験を行った。実験 1 では、3 つの発声方法を用いて、実験参加者の発声に与える影響を比較した。実験 2 では、実験 1 と同一の発声方法を用いて、実験者が感情を推定する精度を比較した。また 2 つの実験結果から、あアラウド法に適した実験設計、感情観測手法としての有用性について議論した。

¹ 明治大学大学院 先端数理科学研究科

² 明治大学 総合数理学部

a) cs222026@meiji.ac.jp

b) keita_w@meiji.ac.jp

*1 “reactivity: 測定それ自体が対象者に影響を与え、後の測定結果を左右すること [4]”

2. 関連研究

2.1 エンタテインメントシステムにおける感情観測の試み

エンタテインメントシステムは何をもって有用と評価できるのか定義が曖昧であり、評価基準や実験手法を確立できていない。そこで、心の動きを可視化することでエンタテインメント性を評価する試みがある。水口ら [1] は EC の本質を「心の動き」によって楽しさを生むことだという仮説に基づき、クオリフィケーション制度を提案した。そのほか、楽曲における緊張-弛緩構造を分析するためのフレームワークの提案 [5][6] や、物語における感情曲線と幸福度による演出の定量化 [7] など、時系列表現における心の動きを評価する研究もある。一方、それらのフレームワークによって「心の動き」を正しく評価できるかは明らかでない。本研究では、「あ」の抑揚から心の動きを可視化することで、リアルタイムな感情観測を試みる。

2.2 様々な感情観測手法

時系列の感情を観測する手法として、表情、発話、ジェスチャ、心拍変動、皮膚電位、脳波など生体反応の計測が広く利用されている [8][9]。生体反応は、表出の難しい感情の認識や、自動的かつマルチモーダルな分析ができる利点がある。一方で、各観測手法にも、実験コストや感情分析の正確性などの課題がある。

表情による感情観測は、カメラにより非接触で行える利点がある。Ekman ら [10] は、表情から 10 の感情を識別できることを明らかにした。また、表情を定量的に測定する技術が開発され [11][12]、深層学習を利用して表情から感情を自動検出する取り組み [13] がある。一方で、表情は容易に改ざんが可能であり、動きの激しいタスクには不向きである。また Ekman はその後の研究で、表情と感情の関係には文化的背景や個人差の影響がある可能性を指摘した [14]。

皮膚電位 [15] や脳波 [16] を用いた感情観測は、実験参加者が改ざんしにくい利点がある。一方で、これらの観測手法は驚きやリラックスなどの特定の感情を取得するのに特化しており、全ての情動や感情を評価するのは難しい。加えて、これらの手法は計測した数値から間接的に感情を推察しているに過ぎない。また、コストが高くシステムが複雑になりやすい。

発話を用いた感情観測として、レトロスペクティブ法や思考発話法など [17] がある。思考発話法は、タスクを実行しながら、そのときの思考を話すことで、どのようなことを考えていたかを時系列で観測できる。一方で、タスクと並行して言語化を行うため、実験参加者の負担が大きく、本来のプレイや思考と異なった結果を生むことがある。

本研究では、実験コストの低減と感情分析の正確性の向

上を行うため、「あ」という単音のみを用いて感情観測を行う。「あ」のみの音声データを収集することにより、高価な機材の装着やテキストマイニングによる分析の必要がなく、分析が容易になる可能性がある。

2.3 発話を用いた感情評価の実験デザイン

発話から感情を評価する実験では、一般に感情ラベルや感情モデルを用いる。

感情ラベルを用いた評価では、実験者が各ラベルに対応する発話サンプルを収集し、実験参加者が収集したサンプルを分類する。ラベルの選択範囲は研究ごとに異なり、怒りや悲しみなどの代表的な感情のみを用いた研究 [3] や、30 種類のラベルを用いた研究 [18] などがある。実験者参加者は、提供したラベルから最も適切なものを選択する手法 [19][20] や、各感情の強さをリッカート尺度で評価する手法 [21] を用いて感情を分類する。また、これらの手法を組み合わせる [22] ことや、自由記述を追加する [18] こともある。

感情モデルを用いた評価では、全ての感情に対して最小単位での表現を試みる。これまでには多くの感情モデルが提案されており、代表例として Russel の円環モデル [23] や Millenson の 3 次元モデル [24]、Pluchik の感情の輪 [25] などがある。なかでも、Russel の円環モデルを用いた感情評価手法が複数提案されている [26][27]。Witvliet と Vrana[27] は、Russel に代表される快の次元と覚醒の次元を 45 度回転し、それぞれを喜怒哀楽の感情空間として表現する方法を提案した。

本研究ではこれらを踏まえ、喜怒哀楽の 4 感情を実験参加者に選択させる方法を用いる。ラベル数を最小限にし、実験者の負担を減らしつつ感情観測が可能であることを調査する。また、喜怒哀楽の 4 感情は、Russel の円環モデルを用いた分析に応用できる点を考慮した。

2.4 韻律と感情の関係

発声から感情を最も定量的に評価できる手法として、韻律の分析がある。韻律の分析では、周波数、音量、持続時間、リズム、声質などの音響的なパラメータを用いる [28]。さまざまな研究により、韻律と感情に関連があることを示した。幸せ、怒り、恐怖を表す発話は平均振動数が高く振幅の変動が大きい一方で、悲しみを表す発話は平均振動数が低く変動が小さい傾向がある [28][29][30]。また、怒りを表す発話は悲しみを表す発話に比べて速度が速くなることが明らかである [31][32]。

以上の点から、「あ」のみの韻律でも感情を分析できる可能性がある。Sauter ら [20] は、笑いや悲鳴のような非言語的感情表現は、音響特性から感情の分類と予測が可能であることを明らかにした。しかし、これは多くのサン

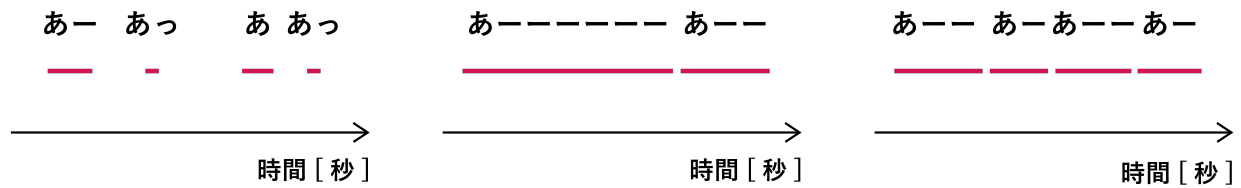


図 2 発声パターンによる発声時間の違い（左）都度発声，（中）耐久発声，（右）断続発声

ルを必要とするため、実験コストが高い。そこで本研究では、実験者が発話音声とタスクの状況を照らし合わせる方法で、感情を分析しうるかを調査する。

3. 実験 1：感情表現に適した発声方法の探索

実験 1 は、実験参加者にとって「あ」の発声による感情表現が容易な発声方法を明らかにすることが目的である。川島ら [2] の実験に基づき、4 項目について 3 つの発声パターンの順位づけを行った。順位付けには一対比較法を用いた。

3.1 実験参加者

実験参加者は 18 歳～24 歳の男性 16 名、女性 3 名の計 19 名の日本人である（平均 20.3、標準偏差 1.4）。

3.2 実験計画

3.2.1 3 つの発声パターン

本実験では、以下の 3 つ発声パターンを用いた。発声パターンは、川島ら [2] の論文を参考にした。3 つの発声パターンのイメージを図 2 に示す。

都度発声

感情変化が起きたときのみ「あ」と発声する

耐久発声

感情変化がなくとも、息が続く限り「あ」と発声し続ける

断続発声

感情変化がなくとも、息が苦しくない程度に区切って「あ」と発声し続ける

都度発声は、“頭に浮かんだことを「あ」だけで語ってください [2]”と指示していた。ただし、頭に浮かんだことを「あ」に変換する負担を考慮し、「あ」の発声のみを重視する指示に変更した。

耐久発声と断続発声については、川島らが“連続的に「あー」と発声し続けてください [2]”と指示していた方法を、2 つのパターンに細分化した。これは、息継ぎのタイミングが集中度に影響を及ぼす可能性があるという先行研究の記述をもとにした。

3.2.2 一対比較アンケート

本実験では、1 人の実験参加者に対して全ての組み合わせ

を提示する「Scheffe の一対比較法（浦の変法）[33]」を用いた。これは、複数種類の刺激の好ましさを順位付けるため、発声の好ましさには個人差があると判断したためである。

一対比較アンケート項目では、発声による実験参加者のストレスを測るため、「発声の負担」と「発声への抵抗感」を調査した。また、感情の観測に適しているかを測るため、「感情表現の難しさ」と「タスクへの集中阻害度」を調査した。発声の負担と感情表現の難しさの 2 項目は、川島らの“言語化の負担 [2]”という項目をもとにした。発声への抵抗感は、思考発話法において緊張感や他者の存在が発言数に影響を及ぼす可能性 [34] があるという事実をもとにした。タスクへの集中阻害度は、川島らの“反応性 [2]”という項目をもとにした。

3.2.3 発声パターンと感情表現の教示ビデオ

本研究では、実験参加者が耐久発声と断続発声の違いを明確に理解できるよう、発声パターンを教示するビデオを作成した。この手法は、複雑な内容や専門用語を含む実験において文字情報だけでなく図表やアニメーションを用いることで実験参加者の理解度が向上するという事実に基づく [35]。

さらに、予備実験から感情が「あ」の音声的特性に反映されにくい実験参加者がいることを確認した。これに対処するため、実験者が喜怒哀楽を「あ」で表現する様子を示す教示ビデオも用意した。教示ビデオにおける喜怒哀楽の表現は、撮影者とは別の実験者 1 名が各感情を適切に観測できることを確認した。なお、教示ビデオのナレーションは音読さん*2で作成した。

3.2.4 使用するゲーム

教示ビデオ内でプレイするゲームには、オープンソースでパラメータ変更が容易なオープンビデオゲームライブラリ [36] から Sliding Penguin*3を用いた。Sliding Penguin は、感情表現がしやすい可能性があるという先行研究の記述を元にした。

一方、実験参加者のタスクには、同ライブラリから Hunter Cameleon*4を用いた。Hunter Cameleon は、1 回のゲー

*2 <https://ondoku3.com/ja/>

*3 <https://open-video-game-library.github.io/SlidingPenguin/>

*4 <https://open-video-game-library.github.io/HunterChameleon/>

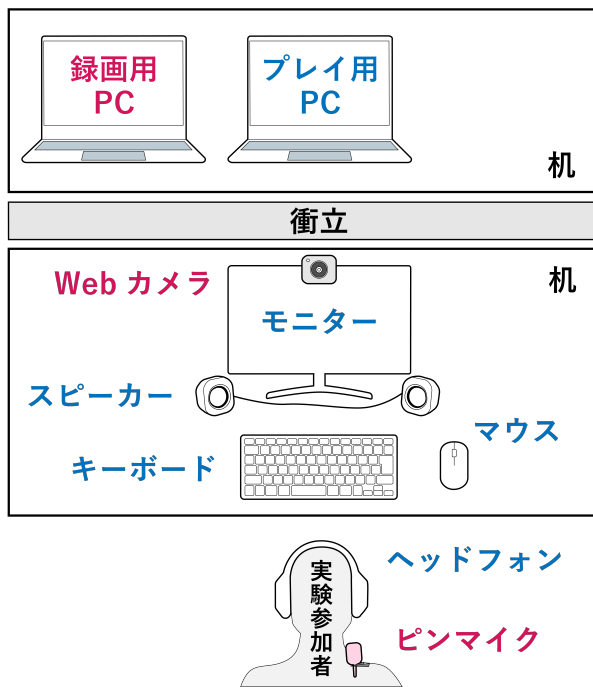


図 3 実験参加者はプレイ用 PC に繋いだ機器を操作しながら「あ」で感情を表現し、実験者は録画用 PC でその画面や声を録画する

ム時間が一定、難易度が調整可能、操作方法が直感的であるため選定した。

ビデオゲームの練習用に調整したパラメータとその値を以下に示す。練習用の値は、操作方法の習得を目的とし、難易度を低めに設定した。

- ハエの出現頻度：10 秒に 3 回
- リングの出現頻度：10 秒に 2 回

本番用に調整したパラメータとその値を以下に示す。本番用の値は、練習用と比較して難易度を高めに設定した。これは、実験参加者がタスク中にゲームに慣れ、感情変化が減少することを防ぐためである。

- ハエの出現頻度：2 秒に 3 回
- リングの出現頻度：2 秒に 3 回

3.3 実験設備

実験全体の構成図を図 3 に、実験参加者が機材を使用し、実験する様子を図 4 に示す。

使用したモニターは、サイズが 27 インチ、画面解像度が 4k、リフレッシュレートが 60Hz の液晶ディスプレイである。ビデオゲームプレイ用 PC には、Windows OS の PC を使用し、録画用 PC と分けた。これは、WebGL 版のビデオゲームの処理が重くなるのを防ぐためである。プレイ用 PC には、ゲームの操作に必要な機材を全て有線接続した。プレイ用 PC に接続した機材を、図 3 内で青色文字で示す。

録画用 PC には、実験参加者の顔を映す Web カメラ、

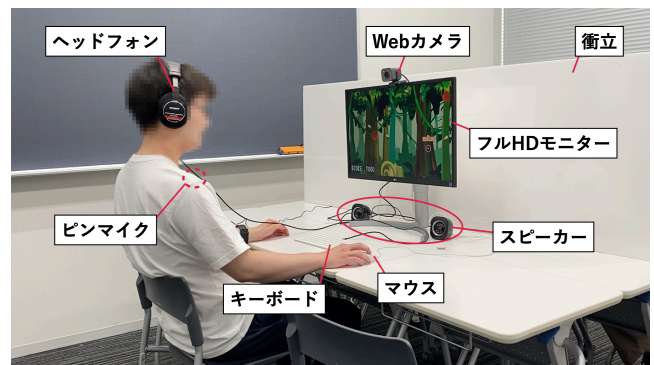


図 4 実験参加者は衝立で実験者と区切られ、周りの音を遮断しながらあアラウド法を行う

「あ」の発声状態を記録するピンマイク、プレイ用 PC を有線接続した。録画用 PC に接続した機材を、図 3 内で赤色文字で示す。

入力した映像や音声の録画には、Multi View Recorder[37]を使用した。Multi View Recorder は、Web 上で複数の映像や音声を 1 つの画面で録画できるオープンソースの録画ソフトウェアである。また、実験参加者はタスクに集中できるようにヘッドフォンを装着し、実験者と実験参加者の間には衝立を設置した。

3.4 実験手続き

実験の流れを以下に示す。

- (1) 実験同意書に記入する
- (2) 実験の説明動画を視聴する
- (3) ビデオゲームのルールや操作の説明を受ける
- (4) 実験で使用するビデオゲームの練習を 1 回行う（難易度は本番とは異なる）
- (5) 指示された発声パターン A で感情表現をしながら、ビデオゲームをプレイする
- (6) 指示された発声パターン B で感情表現をしながら、ビデオゲームをプレイする
- (7) 発声パターン A と B を比較する一対比較アンケートに回答する
- (8) (5) ～ (7) を全順列の計 6 回行う
- (9) 実験終了後、口頭アンケートに回答する

実験の説明動画では、実験の流れと教示ビデオによる各発声パターンの説明を行った上で、「あ」でわかりやすく感情表現を行うように指示した。一対比較アンケートでは、パターン A と比べたパターン B の発声方法を 5 段階評価 (1～5) で回答した。発声が容易（負担が小さい）であれば 1、困難（負担が大きい）であれば 5、同等であれば 3 と回答した。口頭アンケートでは、各パターンについての感想や実験全体で気になった点などを自由に回答した。

3.5 結果

3.5.1 分散分析

BellCurve 社のエクセル統計^{*5}を用いて分散分析を行い、主効果、個人差、発声パターンの順序、組み合わせによる影響の有意性を明らかにした。発声の負担、感情表現の難しさ、発声への抵抗感、タスクへの集中阻害度それぞれにおける分散分析表を表 1 に示す。

発声の負担では、主効果と主効果×個人において有意差 ($p < .01$) があった。感情表現の難しさでは、主効果と主効果×個人において有意差 ($p < .01$) があった。発声への抵抗感では、主効果×個人において有意差 ($p < .01$) があった。順序効果において有意差 ($p < .05$) があった。タスクへの集中阻害度では、主効果と主効果×個人において有意差 ($p < .01$) があった。組み合わせ効果において有意差 ($p < .05$) があった。

3.5.2 平均嗜好度

発声の負担、感情表現の難しさ、発声への抵抗感、タスクへの集中阻害度それぞれにおける平均嗜好度を表した尺度図を図 5 に示す。尺度図では、推定幅以上に離れた要素間では嗜好度に有意差があるといえる。

発声の負担では、平均嗜好度が耐久発声、断続発声、都度発声の順で高かった。また、耐久発声と都度間 ($p < .05$)、断続発声と都度発声間 ($p < .05$) で有意差があった。感情表現の難しさでは、平均嗜好度が断続発声、耐久発声、都度発声の順で高かった。また、断続発声と都度発声間 ($p < .05$)、耐久発声と都度発声間 ($p < .05$) で有意差があった。発声への抵抗感では、平均嗜好度が都度発声、断続発声、耐久発声の順で高かった。また、どの要素間にも有意差はなかった。タスクへの集中阻害度では、平均嗜好度が耐久発声、断続発声、都度発声の順で高かった。また、耐久発声と都度発声間 ($p < .05$)、断続発声と都度発声間 ($p < .05$) で有意差があった。

3.5.3 口頭での自由記述アンケート

発声の負担では、都度発声で「一番感情表現できた感覚」、「突然来るものにすぐ発声できる」という意見と、耐久発声と断続発声で「声を出していればよく、抑揚で喜びや微細な感情を出しやすい」という意見が複数あった。感情表現の難しさでは、「都度発声はハードルが低く、疲れにくい」、「耐久発声より断続発声の方が辛いことはなかった」など、発声時間に比例して疲れたという意見が複数あった。一方で、都度発声で「感情変化はあったが、声が出せなかったときがあった」、耐久発声と断続発声で「上手く感情表現できていたかわからない」という意見もあった。発声への抵抗感では、「都度発声では『表現しなきゃ』という意識が起きてしまう」、「パターン間で恥ずかしさに差はなかった」という意見があった。タスクへの集中阻害度では、「は

じめは都度発声がタスクに集中できたが、後半になると断続発声や耐久発声でも集中できるようになった」という意見が複数あった。断続発声では「区切らないといけない」という意識が働き、むしろやりにくい感じがあった」という意見が複数あった。

教示ビデオに対する感想として、「実験の内容がわかりやすかった」、「大げさにやるというイメージがついた」、「声の出し方 (抑揚とか) のイメージはついた」、「発声練習パートもあってよかった」など、教示ビデオがあることに好意的な意見が複数あった。ただし、「事前にイメージを掴みやすかった」が、「プレイ中は教示ビデオを意識する余裕がなかった」という意見も複数あった。また、普段の発声に近い都度発声に関して、「ビデオがないほうが、自然に声が出せたのではないかと思う」という意見が 1 名あった。

発声の慣れについて、「はじめは都度発声の方がタスクに集中できたが、後半は断続発声や耐久発声でも集中できるようになった」という、普段慣れていない発声方法に徐々に慣れていったという意見が複数あった。また、プレイ回数が増えるほど断続発声と耐久発声の区別がつきにくくなったという意見が複数あった。

普段のゲームプレイや動画視聴時に発声するかどうかについては、「あまりゲームをしないので、声を出すこともない」、「1 人のときは声を出すことは全くないが、友達とやるときは出すこともある」というように、声を出すことに慣れていない実験参加者が半数以上いた。

ヘッドフォンを装着したゲームプレイに関しては、実験参加者 19 名全員が問題なくタスクに集中できたと回答した。一方で、「骨伝導イヤホンの方が自分の声が聞こえるので変化があるのではないかと思う」といった自分の発声がかっこいいかどうか、感情表現の上手さに影響する可能性を指摘した実験参加者もいた。

3.6 考察

図 5 より、発声の負担、感情表現の難しさ、タスクへの集中阻害度の 3 項目において、都度発声の平均嗜好度が他の 2 パターンよりも有意に低かった。一方で、表 1 より、発声の負担、感情表現の難しさ、タスクへの集中阻害度の 3 項目において、結果には個人差があった。また、口頭アンケートでは、都度発声において感情と発声のタイミングのズレを感じたり、喜びといった突発ではない感情は声が出しにくかったりすることがわかった。

これらのことから、都度発声が最も実験参加者に適した発声パターンであるといえる。ただし、実験参加者に複数回プレイしてもらうことや、喜びや楽しさなどの覚醒度の低い感情 [23]、微細な感情の変化を得たい場合には、耐久発声や断続発声を取り入れることも必要である。

そのほか、教示ビデオの視聴によって実験内容と発声イ

^{*5} <https://bellcurve.jp/ex/ac.html>

表 1 発声パターンの分散分析表：

(左上) 発声の負担, (右上) 感情表現の難しさ, (左下) 発声への抵抗感,
(右下) タスクへの集中阻害度

変動要因	平方和	自由度	平均平方	F値	P値	判定
主効果	28.8947	2	14.4474	12.8171	$P < 0.001$	**
主効果×個人	92.4386	36	2.5677	2.2780	0.0028	**
組み合わせ効果	0.7105	1	0.7105	0.6304	0.4306	
順序効果	0.4298	1	0.4298	0.3813	0.5394	
順序×個人	17.4035	18	0.9669	0.8578	0.6278	
残差	63.1228	56	1.1272			
全体	203.0000	114				

変動要因	平方和	自由度	平均平方	F値	P値	判定
主効果	16.6491	2	8.3246	6.6331	0.0026	**
主効果×個人	96.6842	36	2.6857	2.1400	0.0051	**
組み合わせ効果	0.2193	1	0.2193	0.1747	0.6775	
順序効果	0.4298	1	0.4298	0.3425	0.5607	
順序×個人	12.7368	18	0.7076	0.5638	0.9103	
残差	70.2807	56	1.2550			
全体	197.0000	114				

変動要因	平方和	自由度	平均平方	F値	P値	判定
主効果	1.1053	2	0.5526	0.6097	0.5471	
主効果×個人	74.8947	36	2.0804	2.2954	0.0025	**
組み合わせ効果	0.0789	1	0.0789	0.0871	0.7690	
順序効果	4.6404	1	4.6404	5.1199	0.0275	*
順序×個人	9.5263	18	0.5292	0.5839	0.8963	
残差	50.7544	56	0.9063			
全体	141.0000	114				

変動要因	平方和	自由度	平均平方	F値	P値	判定
主効果	28.1228	2	14.0614	14.0131	$P < 0.001$	**
主効果×個人	93.2105	36	2.5892	2.5803	$P < 0.001$	**
組み合わせ効果	4.6404	1	4.6404	4.6244	0.0359	*
順序効果	1.9737	1	1.9737	1.9669	0.1663	
順序×個人	14.8596	18	0.8255	0.8227	0.6667	
残差	56.1930	56	1.0034			
全体	199.0000	114				

* : $p < .05$, ** : $p < .01$

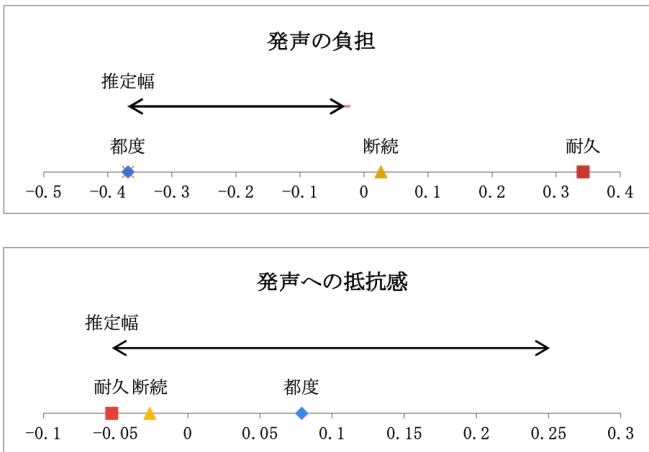


図 5 発声パターンの尺度図：

(左上) 発声の負担, (右上) 感情表現の難しさ, (左下) 発声への抵抗感,
(右下) タスクへの集中阻害度

メージの理解が進んだことがわかった。また、声を出してゲームをプレイすることに慣れていない場合、発声パターンごとの反応性には慣れの影響が出やすい可能性がある。よって、本番タスク前に教示ビデオを用いて発声練習を長めに行い、反応性を低減させることが望ましい。周りの音を遮断するヘッドフォンは、タスクに集中できる一方で、感情表現の上手さに影響を及ぼす可能性がある。よって、実験 2 では開放型イヤホンを利用する。

4. 実験 2：あアラウド法が実験評価として利用できるかの検証

実験 2 は、あアラウド法によって、実験者が実験参加者の感情を適切に観測できるかを明らかにすることが目的である。感情が生起および終了した時間、喜怒哀楽の中で該

当する最も近い感情、都度発声の簡単な経緯の 3 点について、実験者が実験参加者の回答を比較した。

4.1 実験参加者

実験参加者は 19 歳～22 歳の男性 4 名、女性 1 名の計 5 名の日本人である（平均 21、標準偏差 1.22）。

4.2 実験計画

4.2.1 感情観測シート

実験では、実験参加者の感情観測精度を分析する手段として、感情観測シートを用いた。感情観測シートの具体的な記入項目を以下に示す。

感情が生起した時間の計測

実験協力者の感情が変化したら、その開始時

間と終了時間を記録した。ゲーム画面上に表示された残り時間を記録した。

喜怒哀楽の評価

実験参加者はそのときの感情が「喜」「怒」「哀」「楽」のいずれの感情に最も近いと感じたかを記入した。

コメントの記入

その感情がどういった感情だと感じたかを言葉で記入した。

4.2.2 実験に使用した動画と解答

本実験では、実験協力者がオープンビデオゲームライブラリ内の Sliding Penguin をアアラウド法を用いてプレイした様子を撮影した動画を使用した。実験協力者は、感情表現が特に優れた人を、実験1の参加者から我々が選んだ。

動画の撮影には OBS Studio^{*6}を利用し、実験協力者のゲーム画面のキャプチャと、「あ」の発声音声を合わせたものを録画した。

解答は、感情表現シートを用いて作成した。実験協力者は、ゲーム終了後すぐに録画した動画を確認し、感情観測シートに則して記録した。

4.3 実験設備

実験の様子を図6に示す。モニターはサイズが27インチ、画面解像度が4k、リフレッシュレートが60Hzの液晶ディスプレイを使用した。実験動画の再生には Windows OSのPCを用いた。実験参加者は開放型の bluetooth イヤホンを着用し、紙に筆記用具で記入した。

4.4 実験手続き

実験の流れを以下に示す。

- (1) 実験同意書に記入する
- (2) 実験内容についての説明を受ける
- (3) 3つの発声パターンのうち、実験者が指定した動画を視聴しながら、タスクを行う
- (4) (3) をすべての発声パターンの動画で行う
- (5) 実験終了後、口頭アンケートに回答する

口頭アンケートは、半構造化で行なった。質問内容を以下に示す。

- 3つの発声パターンでやりやすかったもの・やりづらかったものはどれか
- どのように分析をしたか
- 分析で難しかったのはどういった点か
- その他難しかった点はあるか



図6 実験参加者はタスク中に動画の一時停止や巻き戻しを任意に実行できる

4.5 結果

4.5.1 感情生起の観測

実験協力者が感情が生起した時間として回答したものを模範解答として、実験参加者の回答と比較した。実験協力者の回答した感情の生起から終了の前後1秒を含む期間と、実験参加者が回答した時間の重なりを調べた。これは、感情の生起から画面を止めるまでの時間誤差を考慮したためである。

平均適合率は耐久発声が96.4%で最も高く、続いて都度発声が85.2%、断続発声が79.2%だった。有意水準を5%として、フリードマン検定を行ったところ $p > .05$ で差はなかった。平均再現率は耐久発声が91.7%で最も高く、続いて断続発声が90.0%、都度発声が80.7%だった。有意水準を5%として、フリードマン検定を行ったところ $p > .05$ で差はなかった。また2つのバランスを示すF値は耐久発声が0.94で最も高く、続いて断続発声が0.84、都度発声が0.80だった。

4.5.2 喜怒哀楽の観測

実験協力者が回答した感情が生起した時間を解答として、実験参加者の回答と比較した。感情生起が正しく観測できたものの中で、実験協力者の解答と実験参加者の回答が一致したものを喜怒哀楽の観測ができたとした。

平均適合率は耐久発声が75.0%で最も高く、続いて断続発声が63.2%、都度発声が39.1%だった。平均再現率は耐久発声が73.3%で最も高く、続いて断続発声が40.0%、都度発声が36.7%だった。有意水準を5%として、フリードマン検定を行ったところ $p < .05$ で有意差が認められた。Scheffeの比較法で多重比較を行ったところ、各水準ごとの有意差は確認されなかった。

4.5.3 コメントの分析

感情生起を正しく識別できたものについて、コメントを比較し調査した。喜怒哀楽の観測ができているもののうち、「怒」の感情は「コースから落ちそうになったことによる、焦りや怒り」などの比較的わかりやすい状況では観測

^{*6} <https://obsproject.com/>

できた。また、ゲームクリアやゲームオーバーなどの終了時の感情も観測できた。一方で、スピードの出すぎによる焦りや怒りなどの「怒」は、スピードが出たことによるスリルを楽しんでいる「喜」と間違えることがあった。

4.6 考察

結果から、あアラウド法によって、体験中の感情生起の観測ができる可能性が高い。また、感情生起の観測と感情の推定において、耐久発声が最も適している可能性が高い。

また、本実験の実験参加者の人数が少ないため、フリードマン検定で有意差が出なかった可能性がある。そのため今後、実験者を増やして再実験を行い、どの方法が最も分析に適しているかを調べる。

喜怒哀楽の観測の精度が低いことに関しては、実験に使用したゲームが「焦り」の感情が発生しやすかったことが影響した可能性が高い。焦りは喜怒哀楽での分類が難しく、感情の観測に影響した可能性がある。そのため今後、焦り以外の感情も偏りなく発生するように、多様なタスクを利用して実験を行う。

また、実験参加者によって分析結果にばらつきがあった。特に実験参加者のうち1名は、実験協力者の怒の発声のほとんどを誤って喜と回答した。これは、教示方法が曖昧であった可能性がある。そのため、今後は実験参加者がある程度の統一性をもって分析できる教示方法を検討する。

5. 議論

5.1 発声方法の検討

我々は2つの実験から、耐久発声と都度発声のそれぞれの、利点と課題を明らかにした。

実験2の結果は、耐久発声が最も感情を観測する精度が高く、感情観測手法として適する可能性を示唆した。実験参加者が5人と少ないため、今後、人数を増やして検証する必要がある。

一方で、実験1より、耐久発声は実験参加者の負担が大きく感情を表現しにくいことがわかった。今後は、実験設計を工夫し、実験参加者の負担の軽減を試みる。また実験1では、実験参加者が「あ」の抑揚を適切に変化させることができているかを確認していない。特に耐久発声は実験参加者への負担が大きく、タスクの妨げになったり、発声への負担が実際の感情表現に影響を及ぼしたりする可能性がある。今後は、耐久発声を用いて、実験参加者が実際に感情を表現できるかを実験する。

5.2 分析手法の検討

実験2では、実験者が実験参加者の感情が生起するタイミングと喜怒哀楽の4感情をある程度正確に観測できる可能性を示唆した。さらには、実験者が特殊な訓練を受けず

とも実験参加者の感情を観測できる可能性を示唆した。

一方で、全ての感情を喜怒哀楽で定義するのは難しい。そのため、本研究の結果だけでは全ての感情を観測できるかは明らかでない。また、人によって喜怒哀楽の基準が異なる可能性がある。今後は、感情ラベルを追加する、感情モデルを活用するなどの分析手法を検討する。さらには、音声情報以外のモダリティを追加してより強固で確実な感情分析を行う方法や、機械学習によって音声特徴を抽出して感情分析を自動化する方法などを調査する。

5.3 評価対象とするコンテンツの選定

実験1の口頭アンケートにおいて、感情変化の頻度が高かった、ゲームの難易度が高くプレイに集中しすぎた、喜びの感情の生起が少なかったという指摘があった。このことから、あアラウド法では評価するコンテンツが発声や分析に影響を与える可能性が高い。

実験1の口頭アンケートを考慮すると、本手法は、感情の動きが明確で頻度が低いコンテンツの評価に適している可能性がある。一方で、操作負荷が高く感情変化が連続的に起きるレースゲームや格闘ゲームなどの評価には不向きな可能性がある。今後は、本手法の利用に適したコンテンツの特徴や利用場面について調査を進め、EC研究への適用性を検討していく。

5.4 あアラウド法の有用性と展望

我々は2つの実験から、実験参加者が自身の感情と対応した発声を問題なく行えること、実験者が特殊な装置や機械学習などを用いずに実験参加者の感情を推測できることを明らかにした。このことから、あアラウド法は実験参加者と実験者の両観点から、感情観測手法として利用できる可能性が高い。

さらにあアラウド法は、意図しない不意の発声から得られる情動と、意図的な感情表現から得られる感情をどちらも観測できる可能性がある。鈴木[38]らは、外部から客観的に観測不可能で言語化が必要な主観的経験や感覚を「感情」、身体反応から客観的に観測可能な主観的経験や感覚を「情動」と区別している。一般的な感情評価方法である身体計測機器[25]は情動の評価に該当し、アンケートは感情の評価に該当する。あアラウド法は、これらを区別せずに評価できる可能性がある。

本研究では、発声と分析の開始前に、喜怒哀楽の発声例とゲームプレイの実演による教示ビデオを用いた。既存の教示ビデオにおいても、喜怒哀楽の大まかな感情を分析できる可能性を示したが、より細かい感情分析が可能であるかは不明である。今後は、実験参加者が感情表現しやすい教示方法と、実験者が分析しやすい教示方法をそれぞれ検討し、観測手法としての精度を高める。

5.5 言語の影響

河野ら [39] は、日本語の感嘆詞において「あ」は表現の種類が多く、驚きや感嘆などの快感情と落胆や否定などの不快感情の両方を表現できることを示した。英語において日本語の「あ」と似た用法である感嘆詞は「oh」であるため、非日本語話者にも同様の結果が得られるかは調査が必要である [40]。また、「ミスした時は『あ』が出やすかったが、他の発声もしたかった」「喜びや楽しさの感情表現がしにくかった」というコメントを踏まえ、「あ」に加えて驚きや感嘆の表現が得意な「お」を併用して発声させることで、より多くの感情表現を促せる可能性がある。

6. おわりに

本稿では、発声方法が実験参加者の負担や反応性に与える影響を調査し、実験者が実験参加者の感情を適切に分析できるかを検証した。この結果、感情変化が起きたときに「あ」と発声させると実験参加者は容易に感情を表現でき、息が続く限り「あ」と発声させると実験者が正しく感情を分析できる可能性を示唆した。

今後は、これらの結果と考察を踏まえ実験手順を整理し、あアラウド法を実験手法として確立していく。そして、あアラウド法を用いて実際にエンタテインメントシステムを評価できるかを検証し、コンテンツの適用範囲と課題を明らかにしていく。

謝辞 本研究は JSPS 科研費 JP22K12336 の助成を受けたものです。

参考文献

- [1] 水口 充, 片寄晴弘: エンタテインメントコンピューティング研究における評価問題の解決に向けての施策の実践, インタラクション, Vol. 2019, pp. 141–150 (2019).
- [2] 川島拓也, 渡邊恵太: あアラウド法: 体験中の心理プロセスを「あ」の音声情報で評価する手法の提案と検証, エンタテインメントコンピューティングシンポジウム 2022 論文集, Vol. 2022, pp. 178–183 (2022).
- [3] Ericsson, K. A. and Simon, H. A.: *Protocol analysis: Verbal reports as data (Revised Edition)*, the MIT Press (1993).
- [4] 山口 洋: 反応性の問題と予測妥当性の自己発生-意図の測定の実動への影響, 社会学部論集, No. 45, pp. 83–91 (2007).
- [5] 片寄晴弘, 橋田光代, 飯野なみほか: Emotion Movement Design Annotator による感動デザインの分析-M. Rigolo のバランス芸を例として, エンタテインメントコンピューティングシンポジウム 2019 論文集, Vol. 2019, pp. 267–274 (2019).
- [6] 橋田光代, 片寄晴弘ほか: 心を動かす情報学: 時系列表現 WG の活動について, 研究報告エンタテインメントコンピューティング (EC), Vol. 2020, No. 1, pp. 1–4 (2020).
- [7] 智也番庄, 晴弘片寄: 物語の演出の抽出手法の考察, エンタテインメントコンピューティングシンポジウム 2020 論文集, Vol. 2020, pp. 82–89 (2020).
- [8] Egger, M., Ley, M. and Hanke, S.: Emotion Recognition from Physiological Signal Analysis: A Review, *Electronic Notes in Theoretical Computer Science*, Vol. 343, pp. 35–55 (2019). The proceedings of AmlI, the 2018 European Conference on Ambient Intelligence.
- [9] Scherer, K. R. and Moors, A.: The Emotion Process: Event Appraisal and Component Differentiation, *Annual Review of Psychology*, Vol. 70, No. 1, pp. 719–745 (online), DOI: 10.1146/annurev-psych-122216-011854 (2019).
- [10] Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E. et al.: Universals and cultural differences in the judgments of facial expressions of emotion., *Journal of personality and social psychology*, Vol. 53, No. 4, p. 712 (1987).
- [11] COHN, J. F., ZLOCHOWER, A. J., LIEN, J. and KANADE, T.: Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding, *Psychophysiology*, Vol. 36, No. 1, p. 35–43 (online), DOI: 10.1017/S0048577299971184 (1999).
- [12] Ekman, P. and Friesen, W. V.: Facial action coding system, *Environmental Psychology & Nonverbal Behavior* (1978).
- [13] Li, S. and Deng, W.: Deep Facial Expression Recognition: A Survey, *IEEE Transactions on Affective Computing*, Vol. 13, No. 3, pp. 1195–1215 (online), DOI: 10.1109/TAFFC.2020.2981446 (2022).
- [14] Ekman, P.: Facial expression and emotion., *American psychologist*, Vol. 48, No. 4, p. 384 (1993).
- [15] Nakasone, A., Prendinger, H. and Ishizuka, M.: Emotion recognition from electromyography and skin conductance, *Proc. of the 5th international workshop on biosignal interpretation*, Citeseer, pp. 219–222 (2005).
- [16] Li, X., Zhang, Y., Tiwari, P., Song, D., Hu, B., Yang, M., Zhao, Z., Kumar, N. and Marttinen, P.: EEG based emotion recognition: A tutorial and review, *ACM Computing Surveys*, Vol. 55, No. 4, pp. 1–57 (2022).
- [17] Ericsson, K. A. and Simon, H. A.: *Protocol analysis: Verbal reports as data (Revised Edition)* (1993).
- [18] Cowen AS, Elfenbein HA, L. P. K. D.: Mapping 24 emotions conveyed by brief human vocalization, *The American psychologist*, Vol. 74, No. 6, pp. 698–712 (2019).
- [19] Leinonen, L., Hiltunen, T., Linnankoski, I. and Laakso, M.-L.: Expression of emotional-motivational connotations with a one-word utterance, *The Journal of the Acoustical society of America*, Vol. 102, No. 3, pp. 1853–1863 (1997).
- [20] Sauter, D., Eisner, F., Calder, A. and Scott, S.: Perceptual Cues in Nonverbal Vocal Expressions of Emotion, *Quarterly journal of experimental psychology (2006)*, Vol. 63, pp. 2251–72 (online), DOI: 10.1080/17470211003721642 (2010).
- [21] Scherer, K. R., Banse, R. and Wallbott, H. G.: Emotion inferences from vocal expression correlate across languages and cultures, *Journal of Cross-cultural psychology*, Vol. 32, No. 1, pp. 76–92 (2001).
- [22] Cao, H., Cooper, D. G., Keutmann, M. K., Gur, R. C., Nenikova, A. and Verma, R.: Crema-d: Crowd-sourced emotional multimodal actors dataset, *IEEE transac-*

- tions on affective computing, Vol. 5, No. 4, pp. 377–390 (2014).
- [23] Russell, J. A.: A circumplex model of affect, *Journal of personality and social psychology*, Vol. 39, No. 6, p. 1161 (1980).
 - [24] Millenson, J. R. and Leslie, J. C.: *Principles of behavioral analysis*, Macmillan New York (1967).
 - [25] Plutchik, R.: *The emotions: Facts, theories and a new model*, University Press of America (1962).
 - [26] Russell, J. A., Weiss, A. and Mendelsohn, G. A.: Affect grid: a single-item scale of pleasure and arousal, *Journal of personality and social psychology*, Vol. 57, No. 3, p. 493 (1989).
 - [27] Witvliet, Charlotte vanOyen Vrana, S. R.: Psychophysiological responses as indices of affective dimensions., *Psychophysiological responses as indices of affective dimensions. Psychophysiology*, Vol. 32, No. 5, pp. 436–443 (1995).
 - [28] Banse, R. and Scherer, K. R.: Acoustic profiles in vocal emotion expression., *Journal of personality and social psychology*, Vol. 70, No. 3, p. 614 (1996).
 - [29] Sobin, C. and Alpert, M.: Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy, *Journal of psycholinguistic research*, Vol. 28, pp. 347–365 (1999).
 - [30] Huttar, G. L.: Relations between prosodic variables and emotions in normal American English utterances, *Journal of Speech and Hearing Research*, Vol. 11, No. 3, pp. 481–487 (1968).
 - [31] Fairbanks, G. and Pronovost, W.: An experimental study of the pitch characteristics of the voice during the expression of emotion, *Communications Monographs*, Vol. 6, No. 1, pp. 87–104 (1939).
 - [32] Johnson, W. F., Emde, R. N., Scherer, K. R. and Klinnert, M. D.: Recognition of emotion from vocal cues, *Archives of General Psychiatry*, Vol. 43, No. 3, pp. 280–283 (1986).
 - [33] 浦 昭二: 1 対比較実験の解析, 品質管理, Vol. 16, pp. 78–80 (1959).
 - [34] 海保博之, 原田悦子: プロトコル分析入門 発話データから何を読むか, 新曜社 (1993).
 - [35] Montalvo, W. and Larson, E.: Participant Comprehension of Research for Which They Volunteer: A Systematic Review, *Journal of Nursing Scholarship*, Vol. 46 (online), DOI: 10.1111/jnu.12097 (2014).
 - [36] 岡 拓也, 川島拓也, 林 大智, 渡邊恵太: 研究利用しやすく標準性を目指したビデオゲームの設計と開発, エンタテインメントコンピューティングシンポジウム論文集, Vol. 2021, pp. 181–186 (2021).
 - [37] 岡 拓也, 川島拓也, 築瀬洋平, 渡邊恵太: HCI 研究における評価実験用ビデオゲームの要件探究とオープンビデオゲームライブラリを用いたケーススタディ, 研究報告ヒューマンコンピュータインタラクション (HCI), Vol. 2022-HCI-196, pp. 1–8 (2022).
 - [38] 鈴木はる江: 感覚と情動から心身相関を考える, 心身健康科学, Vol. 5, No. 1, pp. 8–14 (オンライン), DOI: 10.11427/jhas.5.8 (2009).
 - [39] 河野 武: 日本語における感動詞とイントネーションの交響: 関連性モダリティの風情, 大妻女子大学紀要. 文系 = Otsuma Women's University annual report. Humanities and social sciences, Vol. 51, pp. 226–207 (2019).
 - [40] Heritage, J.: A Change-of-State Token and Aspects of its Sequential Placement, In J. Maxwell Atkinson and John Heritage (eds.), *Structures of Social Action* (1984).