

BCIを用いた感情分類による感情的な擬似発声システム

井上 しいな^{*†} 平野 怜旺^{*‡} 渡邊 恵太[†]

概要. 本研究では、ブレインコンピュータインタフェース (BCI) を用いた感情分類による感情的な擬似発声システムを提案する。本システムは、脳波解析によりリアルタイムでユーザの感情状態を分類し、分類結果を合成音声の感情パラメータにマッピングすることで、ユーザの感情状態を反映した合成音声を実現する。さらに、大規模言語モデル GPT-4o を用いて、ユーザが入力した文章を感情状態に応じて補正することで、感情表現の幅を広げる。ユーザは脳波機器を装着し、簡易的な文章を入力することで、合成音声による感情的な擬似発声を行うことができる。本稿では、本システムの実用性と応用可能性を検証するために、健常者を対象としたユーザスタディを行った。その結果、喜怒哀楽の4感情に対する推定精度が一定の水準に達したことを確認した。さらに、BCIによる感情分類と GPT-4o を組み合わせることで、言語表現の多様性が向上する可能性を示唆した。

1 はじめに

人間の感情状態は、日常生活の意思決定において重要な役割を果たす。言語障害を持つ人は、自身の感情を表出することは非常に困難である。特に、失感情症と呼ばれる症状は、感情を認識し、それを言葉で表現する能力が著しく低下する [43, 56]。自身の感情状態の変化に気づき、それを言葉にすることは、葛藤やフラストレーションに対処するために不可欠であり、健康維持に直結する重要な課題である。

言語障害を持つ人がコミュニケーションを取る手段として、合成音声を用いた手法がある [21, 27]。合成音声を用いることにより、発声が困難な人や状況であっても、音声を介したコミュニケーションを可能にする。また、合成音声の声色を変化させることで、感情表現を可能にした技術も開発されている [5, 34]。このような合成音声技術を活用することにより、言語障害を持つ人であっても、他者との交流を維持し、社会的孤立感を軽減できる [69]。さらに、自身の考えや感情をより自然に伝えることができ、精神的な健康や生活の質が向上する可能性がある [68]。

感情パラメータを搭載した合成音声を用いることにより、ユーザは感情的な擬似発声システムとして利用できる。感情パラメータは、感情ラベルや再生速度、ピッチなどで構成され、これらを設定することにより、合成音声は感情的な音声を再生できる [37, 29]。例えば、感情パラメータを搭載した市販の合成音声である VOICEPEAK¹ は、喜怒哀楽の感情ラベルの設定を 100 段階の数値で決定できる。これにより、ユーザは感情の強度を細かく調整し、よ

り自然で表現豊かな感情的音声を生成できる。

しかし、失感情症の人は、自身の感情状態の認識が難しいため、合成音声の感情パラメータを調整することが困難である。音声を再生する度にパラメータを設定する必要があり、失感情症の人にとっては設定する手間や苦勞が大きい課題となっている。一方、失感情症の人の感情状態を脳波信号を用いて感情状態を推定できることが示唆されている [2]。そのため、リアルタイムで脳内で生起している感情を推定し、その結果に基づいて自動的にパラメータを設定するシステムが有用である。

本研究では、ブレインコンピュータインタフェース (BCI) を用いた感情分類による感情的な擬似発声システムを提案する (図 1)。本システムは、脳波解析により、リアルタイムでユーザの感情状態を分類する。そして、分類結果を合成音声の感情パラメータにマッピングすることで、ユーザの感情状態を反映した合成音声を実現する。また、大規模言語モデル GPT-4o² を用いて、入力した文章をユーザの感情状態を反映した文章に補正する。そのため、ユーザは脳波機器を装着し、簡易的な文章を入力することで、合成音声を用いた感情的な擬似発声を行うことができる。

本稿では、本システムの実用性と応用可能性を検証するため、本システムを用いて、インタビュー形式の会話タスクを行った。その結果、ユーザの感情状態をリアルタイムで反映する擬似発声システムの可能性を示唆した。また、喜怒哀楽の感情推定において、4 分類の分類精度が一定の精度を示し、一般化により、分類精度を向上できる可能性を示唆した。さらに、BCIによる感情分類と GPT-4o を組み合わせることにより、GPT-4o の表現の幅を広げることができる可能性を示唆した。

Copyright is held by the author(s).

* Authors contributed equally

† 明治大学総合数理学部

‡ 明治大学大学院先端数理科学研究科

¹ <https://www.ah-soft.com/voice/6nare/index.html>

² <https://openai.com/api/>

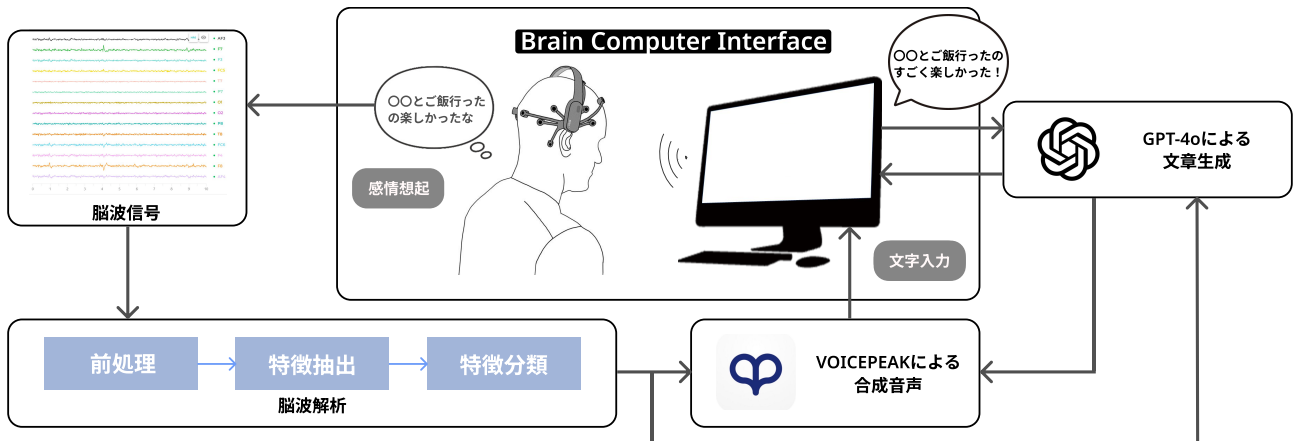


図 1. 本システム構成図. MATLAB を用いてリアルタイムで脳波データを解析し, ユーザの感情状態を喜怒哀楽の 4 つの感情ラベルに分類する. 感情分類結果とユーザの入力文章を GPT-4o に入力し, リアルタイムの感情を反映した文章に補正する. そして, VOICEPEAK に GPT-4o が補正した文章と感情分類結果を入力し, 感情的な擬似発声を行うことができる.

2 関連研究

2.1 言語障害と合成音声支援

言語障害を持つ人や自分の声を失った人は, 自身の思いや感情を言葉で表現できないため, 他者とのコミュニケーションが困難である [59, 49]. 合成音声技術は, 音声を通じた交流を可能にするため, コミュニケーションの代替手段として機能する [9]. また, 研究の進展により合成音声の品質が向上しているが [25, 58], 自然なイントネーションや感情表現に依然として課題が残されている [50].

感情は会話を魅力的で自然な対人コミュニケーションにするために不可欠である [62]. 感情的な合成音声の実現により, 言語障害を持つ人の自己表現の幅を広げることができる [25, 58]. Zhu らは, 感情状態を喜びと怒りの 2 つに分類し, ユーザが感情強度を 3 段階で表現できる音声合成モデルを実現した [73]. また, Yoon らは, 入力テキストから感情推定を行い, 感情的な音声を生成するシステムを提案した [70]. このシステムは, GPT-3 を活用した感情予測手法により, 自然で表現力豊かな感情的な音声を生成できる可能性を示した. これらの研究は, 言語障害者のコミュニケーション支援として有用である.

本研究では, BCI を用いた感情分類と合成音声を組み合わせたシステムを開発した. このシステムは, 脳波データから感情状態を推定し, ユーザの感情状態をリアルタイムに反映した音声を生成する. そのため, ユーザは簡易的な文章を入力することで, 自身の感情を反映した擬似発声を行うことができる. 本システムは, 言語障害者の感情表現の幅を広げ, より自然で豊かなコミュニケーションを可能にするものである.

2.2 言語障害と感情表現

言語障害は大きく分けて, 言葉を表現することを苦手とする失語症と, 発声や発音が上手にできない構音障害がある [63]. 失語症の中でも失感情症の人は, 自身の感情を認識し, 説明することが困難であり, 感情処理の多様な側面に影響を与える [60].

Kessler らは, 失感情症と感情調節の困難さとの間に強い相関関係があることを示した [26]. また, Grynberg らは, 失感情症を持つ人は, 感情的な表情の認識が困難であり, 他者の感情を認識し, 適切に反応することが苦手であることを示した [19]. これらの特徴は, 他者とのコミュニケーションや共感性の低さにつながり, 日常生活における対人関係の構築や維持, 職場でのパフォーマンス, そして全体的な生活の質に影響を及ぼす可能性がある.

失感情症の治療法としては, 認知行動療法 [57] や感情認識トレーニング [48] などが行われているが, 脳波信号を用いた新たなアプローチも注目されている. Laura らは, 脳波信号を用いて喜びや哀しみの感情を分類する方法を提案した [2]. この研究では, 失感情症を持つ人を対象に実験を行ったところ, 感情分類の精度が 94.37% を達成し, 脳波信号を用いた感情推定が失感情症の理解と治療に役立つ可能性を示した.

本研究では, 失感情症を持つ人のコミュニケーション支援を目的とした, BCI を用いた感情的な擬似発声システムを開発した. 本システムは, 失感情症を持つ人の感情表現の困難さを軽減し, より自然で感情豊かなコミュニケーションを支援することを目指す. これにより, 失語症や構音障害といった言語障害を持つ人であっても, 自身の感情を表現し, 他者とのコミュニケーションをより円滑に行えるようになることが期待できる.

2.3 感情分類

感情観測は、心理指標・行動指標・生理指標の3つに分類できる [74]. 心理指標は内面的な心理状態を、行動指標は外面的な行動やパフォーマンスを、生理指標は生理的な反応や身体状態を測定する. また、心理指標は、アンケートやインタビューを通じて行われ、他の客観的な指標と組み合わせて使用できる.

行動指標を用いた感情分類では、様々な手法で感情分類が行われている. 例えば、Issaらは音声による感情分類において、8分類(哀しみ, 幸せ, 怒り, 穏やか, 恐れ, 驚き, 中立, 嫌悪)で71.61%, 7分類で86.1%, 4分類で64.3%の精度を達成した [23]. また、Jayalekshmiらは、画像から自動的に表情を認識する手法を提案し、7分類で90.14%の分類精度を達成した [24]. さらに、Limらは、バーチャル環境における感情分類において、瞳孔位置を使用する手法を提案し、最大で59.19%の精度を達成した [30].

一方、生理指標を用いた感情分類では、脳波信号が注目されている. 脳波信号は無意識的な生理信号であり、他の手法と比較して、意識的に感情分類結果を操作することが難しい特徴がある [31]. さらに、脳波は高い時間分解能を持っており、ミリ秒単位で感情状態を検出することができる [6, 11]. Balconiらは、動画を用いて実験参加者に特定の感情を誘発したところ、感情誘発をしてから150〜250ミリ秒の間で脳波が変化し、脳波が感情の変化に敏感であることを示した [65]. また、Valentiらは、脳波データを用いて4分類(面白い, 嫌悪, 哀しい, 中立)で97.2%の高い精度を達成した [66].

本研究では、脳波データを用いてユーザの感情を喜怒哀楽の4つに分類した. 脳波を用いた感情分類は、他の手法と比較して高い精度と時間分解能が期待できるため、感情観測の有力な手段である. さらに、失感情症の人の感情状態を脳波信号を用いて推定できることが示唆されている [2]. そのため、自身の感情を認識することが難しい失感情症を持つ人にとって、脳波を用いた感情分類は最適なアプローチである.

3 システム

3.1 システム構成

本システムでは、EMOTIV EPOC X³を用いて脳波計測を行い、MATLABを用いて脳波解析を行った. EPOC Xのセンサー配置は、国際的な10-20システムに基づいており、サンプリング周波数は256Hzに設定した(AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4) [3, 12]. 取得した脳波データは、無線通信によりコンピュータ

に送信し、MATLAB上でデータの前処理、特徴抽出、特徴分類の一連の処理を実行した. 分類結果は、リアルタイムで1秒ごとに出し、ユーザの文章入力中の複数の分類結果から、喜怒哀楽の感情の割合を計算し、感情パラメータの数値を決定した.

さらに、Pythonを用いて、GPT-4oとVOICEPEAKを統合することで、リアルタイムでユーザの感情を反映する擬似発声システムを開発した. 感情パラメータの数値は、UDP通信によりMATLABからPythonに送信され、Python上で、受信した数値データをGPT-4oとVOICEPEAKに送信する. 文章生成にはGPT-4oを使用し、感情パラメータとユーザの入力文章をGPT-4oに送信することで、その感情状態を考慮し、GPT-4oがユーザの感情状態を反映した文章を生成する. 最後に、VOICEPEAKは感情パラメータに基づき、GPT-4oが生成した文章を音声で出力する.

3.2 BCIシステム

3.2.1 前処理

本システムでは、脳波データの解析に、MATLABを使用した [13]. 生の脳波データに対して前処理を実施した [64]. 重要な脳波成分を包括的に捉える1Hzから40Hzの周波数帯域を対象とするバンドパスフィルタを適用した [45, 42]. これにより、信号中のノイズ成分を効果的に低減した [36]. フィルタリング後、データを各刺激イベントに対応するエポックに分割することで [20], 特定の刺激に対する脳の反応を精密に抽出し、解析における正確性を担保した [39, 55].

3.2.2 特徴抽出

特徴抽出には、共通空間パターン(CSP)アルゴリズムを採用した [46]. CSPは、異なる脳の状態や反応を表す信号パターンを効果的に識別する [8]. 収集した脳波データの共分散行列を算出し、固有値分解することで、信号の空間的な分布を最適に表現する固有ベクトルを得る [16]. これらの中から、分類タスクに最適なものを選択し、CSPフィルターとして利用する. フィルタリング後のデータに対して、対数分散を計算し、特徴量として使用した [32].

3.2.3 特徴分類

特徴量では、サポートベクターマシン(SVM)を採用した [41]. SVMは、高次元空間における最適な決定境界を見出す機械学習アルゴリズムである [67]. 非線形の問題に対してもカーネル法を用いて効果的に対応できる [54]. 本システムでは、脳波データから抽出した特徴量をSVMに入力し、分類器の学習を行った.

³ <https://www.emotiv.com/epoc-x/>

3.3 感情擬似発声システム

本研究では、BCIを用いた感情分類により、リアルタイムでユーザの感情状態を反映した擬似発声システムを開発した。画面下部に入力部分を表示し、画面上部にユーザが入力した文章とユーザの感情状態を反映した文章を表示した。ユーザは、脳波機器を装着し、文章を入力することで、自身の感情を効果的に伝えることができる。ユーザのシステム使用例を図2に示す。この例では、ユーザは「最近楽しかったこと」についてシステムを使用して回答した。このとき、ユーザの感情状態の分類結果の割合は（喜、怒、哀、楽）で（15, 0, 0, 85）であった。そのため、ユーザが「昨日は友達とご飯に行った」と入力すると、「楽しくて、美味しかった!」といったポジティブな表現がGPT-4oによって補正される。さらに、感情状態の分類結果をVOICEPEAKの感情パラメータにマッピングすることで、感情状態を反映した音声が出力する。このシステムにより、ユーザは感情的な擬似発声を行うことができる。

4 ユーザスタディ

4.1 感情誘発動画

本実験では、感情を誘発するために、8本のショート動画をを用いた[18, 35, 61, 7]。それぞれの動画は、喜怒哀楽の4つの感情を誘発しやすい内容にし、23分の長さで構成した。参加者の正しい感情を引き出すために、7つの条件（感情の誘発性、既知性の低さ、視覚的なシンプルさ、多様性、感情の特異性、長さ、理解しやすさ）を基に慎重に選定した[17, 18]。

4.2 参加者

参加者は、大学生および大学院生の計10名（男性6名、女性4名、年齢:19-24、平均年齢:20.8）である。本実験では、感情分類のための分類器を作成するデータ収集セッションを行った。ユーザ自身の感情認識を必要とするため、言語障害を持たない参加者を対象とした。参加者のプライバシーを最大限に尊重し、全ての個人情報匿名化して保存した。

4.3 手順

本研究では、実験の信頼性と参加者の理解を確保するために、実験全般の詳細な説明を行い、実験を開始した。まず、参加者にEPOC Xを装着し、脳波データの質を確認した。脳波の安定を確認後、参加者は感情誘発動画を視聴した。この時、同時に脳波データを取得した。次に、参加者は本システムを用いて、20分間の会話タスクに取り組んだ。実験終了後、口頭での自由記述インタビューを実施し、参加者の主観的な意見や感想を収集した。

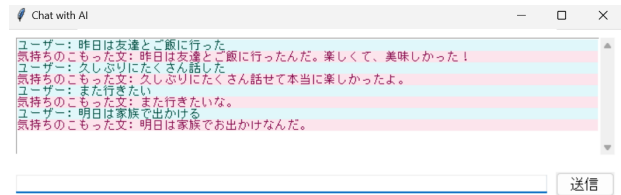


図 2. 入力システム画面。画面下部に入力部分を表示し、画面上部にユーザが入力した文章とユーザの感情状態を反映した文章を表示。ユーザは簡易的な文章を入力することで、自身の感情を効果的に伝えることができる。

4.4 データ収集

データ収集セッションでは、参加者は感情誘発動画を視聴しながら、コントローラーのジョイスティック部分を操作した[72]。感情を定量的に評価するため、感情モデルにラッセルの円環モデルを採用した。このモデルは人の感情が活性・不活性 (arousal) と快・不快 (valence) の2次元の連続量で記述されるという考えに基づく[47]。横軸を valence、縦軸を arousal とすると、その符号によって離散化し、感情の分類を行うことができる。このとき、第1象限は喜び、第2象限は怒り、第3象限は哀しみ、第4象限は楽しみの感情を示す(図3)。

本実験では、覚醒している、快適・心地よい、落ち着いている、不快の4つの指標を用いて、感情を4分類する(1:覚醒しているかつ快適, 2:快適かつ落ち着いている, 3:落ち着いているかつ不快, 4:不快かつ覚醒している)。コントローラーの座標位置と喜怒哀楽を対応させ、参加者は動画を視聴しながら、感情状態に合わせてコントローラーを操作した。この操作を、8本の動画に対して行った。

感情誘発動画を視聴している参加者の脳波データを収集した。バイアスを防ぐため、参加者には動画がどの感情を誘発するかは事前に伝えず、順序効果を考慮するために動画の提示順序をランダム化した。収集した脳波データをもとに、SVM分類器を作成した。

4.5 会話タスク

データ収集セッション後、参加者は本システムを用いて、特定のテーマに沿って会話タスクを行った。このタスクの目的は、システムの実用性と有効性を評価することである。会話テーマは4つの感情に基づいて設定し(1.最近嬉しかったこと, 2.最近怒ったこと, 3.最近哀しかったこと, 4.最近楽しかったこと)、各テーマについて、5分間インタビュー形式で会話を行った[28, 51, 15]。これらの質問は、参加者の個人的な経験や感情を引き出すように設計し、単純な「はい」「いいえ」では答えられない開放型の質問とした。例えば、「最近嬉しかったこと」のテ

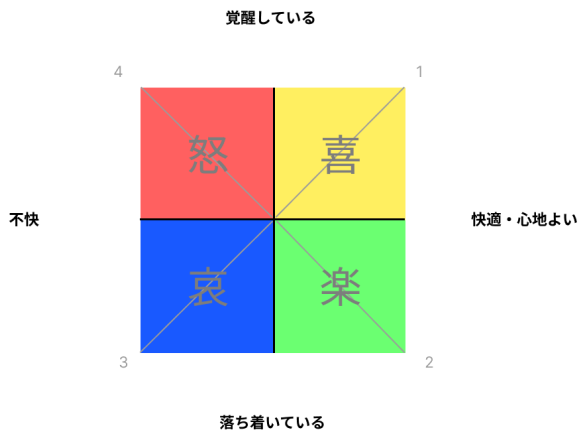


図 3. ラッセルの円環モデルに基づく感情分類. 第 1 象限は喜び, 第 2 象限は怒り, 第 3 象限は哀しみ, 第 4 象限は楽しみの感情を示す.

マでは、「あなたが最近経験した嬉しい出来事について教えてください」といった質問を含む。参加者は、各質問に対して本システムを使用して返答した。

5 結果と考察

5.1 分類精度

本研究では、本システムの感情分類性能を検証した。その結果、喜怒哀楽の 4 クラス分類において、個人の分類精度は 44.25% から 59.18% の範囲であり、一定の分類性能を示した。今回のユーザスタディでは、分類器を一般化せずに個人最適化した。これは、少ない実験参加者数で分類器を一般化すると、汎化能力が低下するためである。したがって、より多くの脳波データを収集し、一般化した分類器を開発することで、汎化能力が向上する可能性がある [1, 38]。また、感情データを取得することなく本システムが使用可能になれば、言語障害者であっても本システムを利用できる可能性がある。言語障害者は、自身の感情状態が認識できず、感情データの取得が困難な場合が多いため、感情データ取得が不要なシステムは有用である。今後の課題として、より多くの実験参加者を対象とした研究を行い、分類器の汎化能力を向上させる必要がある。さらに、実際の言語障害者を対象に本システムの有用性を検証していくことが重要である。

5.2 感情認識

感情認識に関して、ポジティブな感情の認識と伝達に関して肯定的な意見を得た。一方で、ネガティブな感情の表現に関して否定的な意見があり、会話タスクの設計について改善の必要性を示唆した。参加者からは、「楽しい感情と嬉しい感情が正確に認識され、効果的に伝えられた」、「擬似発声の感情表現は

豊かで、抑揚が自然で人間らしさを感じた」といった回答を得たが、「悲しみと怒りの感情表現が不自然だった」といった意見も得た。また、「システムを使用しているだけで、新しいものを使うワクワク感を感じた」という意見もあり、会話タスクにおいて参加者のネガティブな感情を十分に誘発できなかった。

一方で、「ずっと怒りの音声ばかり出力されていた」といった意見も得た。この理由として、「脳波計を付けていることにストレスを感じていた」といった回答があったことから、脳波計の使用自体が参加者のストレスとなり、怒りの感情が誘発されやすくなった可能性がある。会話タスクの設計や計測機器の使用方法に配慮し、より自然な感情表現を実現することが課題である。

5.3 応答文

本研究では、BCIによる感情分類と GPT-4o を組み合わせることにより、入力文章にユーザの感情状態を反映し、感情的な文章に補正した応答文を出力した。ユーザスタディの結果、BCIによる感情分類と GPT-4o を組み合わせることにより、GPT-4o の表現の幅を広げることができる可能性を示唆した。

応答文に関して、感情を効果的に伝える文章生成ができていたことを示したが、表現の不自然さや誇張に対する改善の必要性も示唆した。「思ったまま入力した文章を伝わりやすいように 2 文にしていたのがよかった」、「メモ書きで打ったものを文章で言い換えてくれていた」という意見を得た。

一方で、「応答文が意図と異なる言い回しになっていたり、語尾が変わっていたりするのが不自然に感じた」、「感情が誇張されすぎていたり、思っていない文章が追加されたりした」といった否定的な意見も得た。感情を強調するために入力文から変更された表現が、一部の参加者は不自然に感じたと考察する。

6 議論

6.1 感情誘発の方法

本研究では、感情擬似発声システムの有用性を評価するために、参加者に特定の感情を想起させ、その感情を会話タスクを通じて誘発する実験を行った。実験の感情誘発方法としては、参加者自身の経験を想起させるアプローチを採用した。その結果、ポジティブな感情に対してはシステムが適切に感情を反映できたが、ネガティブな感情に対しては期待した反応が得られなかった。この結果には複数の要因を考察する。まず、システムの使用自体が参加者にとって新鮮で楽しい体験となり、ネガティブな感情を引き出すことができなかった可能性がある。また、一部の参加者にとっては、怒りや悲しみといったネガティブな感情を想起することが難しかった可能性もある。今回の実験では、感情誘発の方法として参加

者自身の経験の想起を用いたが、この方法では参加者間で誘発される感情の質や強度にばらつきが生じやすい。より統制された感情誘発を行うためには、感情誘発をしやすい動画や音楽、シナリオなどの刺激を用いることが重要である。

今後の研究では、ユーザの感情状態を均等に誘発するための感情誘発手法の改良が必要である。特に、ネガティブな感情を効果的に喚起するための新たなアプローチを検討することが重要である。また、ユーザの感情状態をより正確に反映するために、システムのより多様な感情表現方法についても検討する必要がある。

6.2 ユースケース

本研究では、BCIを用いた感情分類による感情的な疑似発声システムを提案した。リアルタイムの感情分類を合成音声の感情パラメータにマッピングしたことにより、ユーザの感情を発声や認識を伴わずに表現できる可能性を示したことが、本研究の主要な貢献である。本システムは様々なケースで応用できる可能性がある。

6.2.1 言語障害支援や診療への応用

本システムは、言語障害を持つ人のコミュニケーション支援とリハビリテーションに活用できる可能性がある。自身の感情状態に対応する感情的な音声を聴くことで、感情状態と声質の関連性の理解が促進される。これは、Mayerらが提唱した感情知能理論に基づく感情理解の向上に寄与し、適切な発声練習を通じて感情表現能力が向上することが報告されている[40]。さらに、Bänzigerらが提案したマルチモーダル感情認識テスト[4]に、本システムを活用することで、患者の感情表現スキルの向上や、セラピストの感情認識能力の訓練が可能となる。また、Schlegelらが提案したジュネーブ感情認識テスト[53]と同様に、本システムも感情認識能力の測定・訓練ツールとしての活用が期待できる。

6.2.2 インタラクティブ性向上のための応用

本システムは、リモート会議やリアルタイムチャットにおいて、離れた相手に感情を伝え、会話のインタラクティブ性を向上できる可能性がある。コンピュータを介したコミュニケーションにおいて感情伝達は重要な役割を果たす[14]。リモートでの会話は、通信環境やデバイスの状態に左右されやすいが、合成音声を使用することで、安定して感情的な音声を提供することができる。また、移動中やダブルワークなど声を発せられない場面においても本システムは有効である。

オンラインゲームやバーチャル空間における会話シーンにおいても本システムの応用が期待できる。ゲーム中にプレイヤーが表す多様な感情をゲーム内

で可視化することで、没入感や楽しさが向上する可能性が示されている[71, 22]。そのため、これらの感情を他者と共有することで、インタラクティブ性の向上が期待できる。バーチャル空間内においても、ユーザの感情をアバターの表情に反映させることで、より自然なコミュニケーションを実現できる[44]。このように、現実世界と異なる環境でも、感情がコミュニケーションの質を向上させる重要な要素となり得る。本システムは、感情を効果的に伝えるため、他者とのインタラクティブな会話を可能とする。また、合成音声を使用するため、非現実的空間において匿名性を保つことができる。

6.2.3 エンタテインメント分野への応用

本システムは、エンタテインメント分野での応用も期待できる。例えば、演劇において、感情を適切に表現することは観客に内容を効果的に伝えるために重要であり、演技における感情表現は観客の没入感と理解を深める上で重要な要素である[52]。本システムにより、合成音声に感情的な音声を発させることで、演者は自身の感情を正確に表現する方法を効果的に練習できる可能性がある。

また、VtuberやAItuberといったバーチャルキャラクターが人気を集めている。これらは新しいエンタテインメントの形態であり、視聴者にエンタテインメントやコミュニケーションを提供する[10, 33]。本システムにより、キャラクターに自分の感情を乗せて話すことで、視聴者によりインタラクティブな体験を提供できる。これにより、視聴者の関与度を高めるだけでなく、感情的な共鳴を促進することで、物語やキャラクターに対する理解や共感を深めることが期待できる。

今後の研究では、さらなるアプリケーションのために、より高度な感情分類アルゴリズムの開発や、リアルタイムでの応答性の向上を検討する。これにより、よりユーザの感情状態に適した発声を実現することを目指す。

7 おわりに

本研究では、脳波データを用いた感情分類による感情疑似発声システムを提案し、その有用性について調査した。本システムは、合成音声のパラメータに脳波データによる感情分類を用いることで、ユーザの発声や感情認識を伴わずに、ユーザの感情状態をリアルタイムで反映する合成音声を実現した。実験結果は、本システムがユーザの感情状態を適切に表現できることを示唆しており、言語障害者のコミュニケーション支援ツールとして有用である可能性を示している。また、本システムはコンピュータを介したコミュニケーションのインタラクティブ性の向上やエンタテインメント分野への応用も期待できる。

参考文献

- [1] A. Apicella, P. Arpaia, Giovanni D'Errico, D. Marocco, G. Mastrati, N. Moccaldi, R. Prevede. Toward cross-subject and cross-session generalization in EEG-based emotion recognition: Systematic review, taxonomy, and methods. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2022.
- [2] S. D. ArulDass and P. Jayagopal. Identifying Complex Emotions in Alexithymia Affected Adolescents Using Machine Learning Techniques. *Diagnostics*, 12(12):3188, 2022.
- [3] N. A. Badcock, P. Mousikou, Y. Mahajan, P. de Lissa, J. Thie, and G. McArthur. Validation of the Emotiv EPOC(®) EEG gaming system for measuring research quality auditory ERPs. *PeerJ*, 1:e38, 2013.
- [4] T. Bänziger, D. Grandjean, and K. R. Scherer. Emotion recognition from expressions in face, voice, and body: the Multimodal Emotion Recognition Test (MERT). *Emotion*, 9(5):691–704, 2009.
- [5] H. Barakat, O. Turk, and C. Demiroglu. Deep learning-based expressive speech synthesis: a systematic review of approaches, challenges, and resources. *EURASIP Journal on Audio, Speech, and Music Processing*, 2024:Article 11, 2024.
- [6] M. Y. Bekkedal, J. Rossi, and J. Panksepp. Human brain EEG indices of emotions: Delineating responses to affective vocalizations by measuring frontal theta event-related synchronization. *Neuroscience Biobehavioral Reviews*, 35(9):1959–1970, 2011. Pioneering Research in Affective Neuroscience: Celebrating the Work of Dr. Jaak Panksepp.
- [7] R. F. S. Benjamin H. Detenber and G. G. B. Jr. Roll 'em!: The effects of picture motion on emotional responses. *Journal of Broadcasting & Electronic Media*, 42(1):113–127, 1998.
- [8] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-r. Müller. Optimizing Spatial filters for Robust EEG Single-Trial Analysis. *IEEE Signal Processing Magazine*, 25(1):41–56, 2008.
- [9] H. T. Bunnell. Speech Synthesis: Toward a “Voice” for All. *Acoustics Today*, 2021.
- [10] P. Chinchilla and J. Kim. VTuber for Streamers: Exploring the Role of Social Presence in the Visual Representation of Streamers. *Communication Studies*, 0(0):1–17, 2024.
- [11] P. R. Davidson, R. D. Jones, and M. T. R. Peiris. EEG-Based Lapse Detection With High Temporal Resolution. *IEEE Transactions on Biomedical Engineering*, 54(5):832–839, 2007.
- [12] S. Debener, F. Minow, R. Emkes, K. Gandras, and M. De Vos. How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology*, 49(11):1449–1453, 2012.
- [13] A. Delorme and S. Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, 2004.
- [14] D. Derks, A. H. Fischer, and A. E. Bos. The role of emotion in computer-mediated communication: A review. *Computers in Human Behavior*, 24(3):766–785, 2008. Instructional Support for Enhancing Students' Information Problem Solving Ability.
- [15] P. Dilley. Interviews and the Philosophy of Qualitative Research. *The Journal of Higher Education*, 75:127–132, 2003.
- [16] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Elsevier, 2nd edition, 1990.
- [17] S. Girard, T. Beesley, S. Hudon, B. Lau, and J. Infante. Database of Emotional Videos from Ottawa (DEVO). *Collabra: Psychology*, 6(1):10, 2020.
- [18] J. J. Gross and R. W. Levenson. Emotion elicitation using films. *Cognition and Emotion*, 9(1):87–108, 1995.
- [19] D. Grynberg, B. Chang, O. Corneille, P. Maurage, N. Vermeulen, S. Berthoz, and O. Luminet. Alexithymia and the processing of emotional facial expressions (EFEs): systematic review, unanswered questions and further perspectives. *PLOS ONE*, 7(8):e42429, 2012. Epub 2012 Aug 23.
- [20] T. C. Handy. Event-related potentials: A Methods Handbook. In *Proceedings of the Workshop on Event-Related Potential Techniques*, 2005.
- [21] M. S. Hawley, S. P. Cunningham, P. D. Green, P. Enderby, R. Palmer, S. Sehgal, and P. O'Neill. A voice-input voice-output communication aid for people with severe speech impairment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(1):23–31, Jan 2013. Epub 2012 Aug 3.
- [22] K. Isbister. *How Games Move Us: Emotion by Design*. MIT Press, Cambridge, MA, 2016.
- [23] D. Issa, M. Fatih Demirci, and A. Yazici. Speech emotion recognition with deep convolutional neural networks. *Biomedical Signal Processing and Control*, 59:101894, 2020.
- [24] J. Jayalekshmi and T. Mathew. Facial expression recognition and emotion classification system for sentiment analysis. In *2017 International Conference on Networks Advances in Computational Technologies (NetACT)*, pp. 1–8, 2017.
- [25] S. Kayte, B. Ambedkar, M. R. Mundada, and C. N. Kayte. A Review of Unit Selection Speech Synthesis. In *Proceedings of the International Conference on Communication and Signal Processing*, 2015.
- [26] H. Kessler, M. Kammerer, H. Hoffmann, and H. Traue. Regulation von Emotionen und

- Alexithymie: Eine korrelative Studie [Regulation of emotions and alexithymia: a correlative study]. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 60(5):169–174, May 2010. In German.
- [27] G. Krishna, C. Tran, Y. Han, M. Carnahan, and A. H. Tewfik. Speech synthesis using EEG. *arXiv preprint arXiv:2002.12756v2*, 2020.
- [28] S. Kvale and S. Brinkmann. *InterViews: Learning the Craft of Qualitative Research Interviewing*. SAGE Publications, 1996.
- [29] Y. Lei, S. Yang, X. Wang, and L. Xie. MsEmoTTS: Multi-scale emotion transfer, prediction, and control for emotional speech synthesis. *arXiv preprint arXiv:2201.06460v1*, 2022.
- [30] J. Z. Lim, J. Mountstephens, and J. Teo. Exploring Pupil Position as An Eye-Tracking Feature for Four-Class Emotion Classification In VR. *Journal of Physics: Conference Series*, 2129(1):012069, dec 2021.
- [31] Y.-J. Liu, M. Yu, G. Zhao, J. Song, Y. Ge, and Y. Shi. Real-Time Movie-Induced Discrete Emotion Recognition from EEG Signals. *IEEE Transactions on Affective Computing*, 9(4):550–562, 2018.
- [32] F. Lotte and C. Guan. Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms. *IEEE Transactions on Biomedical Engineering*, 58(2):355–362, 2011.
- [33] Z. Lu, C. Shen, J. Li, H. Shen, and D. Wigdor. More Kawaii than a Real-Person Live Streamer: Understanding How the Otaku Community Engages with and Perceives Virtual YouTubers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.
- [34] R. d. A. M. J. C.-B. V. N. M. A. R. J. M.-M. Lucía Gómez-Zaragozá, Óscar Valls. Speech emotion recognition from voice messages recorded in the wild. *arXiv preprint arXiv:2403.02167*, 2024.
- [35] S. J. Luck. Ten Simple Rules for Designing ERP Experiments. *Behavior Research Methods, Instruments, & Computers*, 36:530–533, 2004.
- [36] S. J. Luck. An Introduction to the Event-Related Potential Technique. In *Proceedings of the Workshop on Event-Related Potential Techniques*, 2005.
- [37] X. Luo, S. Takamichi, T. Koriyama, Y. Saito, and H. Saruwatari. Emotion-Controllable Speech Synthesis Using Emotion Soft Labels and Fine-Grained Prosody Factors. In *Proceedings of the APSIPA Annual Summit and Conference*, pp. 794–799. APSIPA, 2021. Authorized licensed use limited to: Cambridge University. Downloaded on August 18, 2024, from IEEE Xplore. Restrictions apply.
- [38] V. T. Maja Stikic, Robin R. Johnson and C. Berka. EEG-based classification of positive and negative affective states. *Brain-Computer Interfaces*, 1(2):99–112, 2014.
- [39] S. Makeig, S. Debener, J. Onton, and A. Delorme. Mining event-related brain dynamics. *Trends in Cognitive Sciences*, 8(5):204–210, 2004.
- [40] J. D. Mayer, P. Salovey, and D. R. Caruso. Emotional Intelligence: Theory, Findings, and Implications. *Psychological Inquiry*, 15(3):197–215, 2004.
- [41] K.-R. Müller, M. Tangermann, G. Dornhege, M. Krauledat, G. Curio, and B. Blankertz. Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring. *Journal of Neuroscience Methods*, 167(1):82–90, 2008. Brain-Computer Interfaces (BCIs).
- [42] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti. ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, 48(2):229–240, Feb 2011.
- [43] Y. Moriguchi, T. Ohnishi, R. D. Lane, M. Maeda, T. Mori, K. Nemoto, H. Matsuda, and G. Komaki. Impaired self-awareness and theory of mind: An fMRI study of mentalizing in alexithymia. *NeuroImage*, 32(3):1472–1482, 2006.
- [44] A. Neviarouskaya, H. Prendinger, and M. Ishizuka. EmoHeart: Conveying Emotions in Second Life Based on Affect Sensing from Text. *Advances in Human-Computer Interaction*, 2010:Article ID 209801, 13 pages, 2010.
- [45] E. F. L. Niedermeyer and F. H. L. da Silva. *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*, Fourth Edition. In *Proceedings of the 4th International Conference on Clinical Applications of Electroencephalography*, 1998.
- [46] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE transactions on rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, 8 4:441–6, 2000.
- [47] J. A. Russell. Affective space is bipolar. *Journal of Personality and Social Psychology*, 37:345–356, 1979.
- [48] D. Samur, M. Tops, C. Schlinkert, M. Quirin, and S. L. Koole. Four decades of research on alexithymia: moving toward clinical applications. *Frontiers in Psychology*, 4:861, 2013.
- [49] M. Sato. For Understanding of Aphasia and Aphasics. *Higher Brain Function Research*, 2020.
- [50] N. Savage. Thinking deeply to make better speech. *Communications of the ACM*, 60:15 – 17, 2017.

- [51] L. W. Sayrs. Interviews : an introduction to qualitative research interviewing. *Evaluation and Program Planning*, 3:287–288, 1996.
- [52] R. Schechner. *Performance Studies: An Introduction*. Routledge, London, 4th edition, 2020.
- [53] K. Schlegel, D. Grandjean, and K. R. Scherer. Introducing the Geneva emotion recognition test: an example of Rasch-based test development. *Psychological Assessment*, 26(2):666–672, 2014. Epub 2013 Dec 2.
- [54] B. Scholkopf and A. Smola. Learning with Kernels: support vector machines, regularization, optimization, and beyond. In *Adaptive computation and machine learning series*, 2001.
- [55] N. B. Shamlo, T. R. Mullen, C. Kothe, K. Su, and K. A. Robbins. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, 9, 2015.
- [56] P. E. Sifneos. The prevalence of "alexithymic" characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics*, 22:255–262, 1973.
- [57] V. Spek, I. Nyklíček, P. Cuijpers, and V. Pop. Alexithymia and cognitive behaviour therapy outcome for subthreshold depression. *Acta Psychiatrica Scandinavica*, 118(2):164–167, Aug 2008. Epub 2008 May 22.
- [58] X. Tan, T. Qin, F. K. Soong, and T.-Y. Liu. A Survey on Neural Speech Synthesis. *ArXiv*, abs/2106.15561, 2021.
- [59] Y. Tanaka, M. Shindo, Y. Hashimoto, and K. Kaga. Consideration of Communicative Problems in Aphasic Patients. *The Japan Journal of Logopedics and Phoniatrics*, 31:404–411, 1990.
- [60] G. J. Taylor and R. M. Bagby. New trends in alexithymia research. *Psychotherapy and Psychosomatics*, 73(2):68–77, Mar-Apr 2004.
- [61] J. Theeuwes. Top-down and bottom-up control of visual selection. *Acta Psychologica*, 135(2):77–99, 2010.
- [62] A. Triantafyllopoulos, B. W. Schuller, G. İymen, M. Sezgin, X. He, Z. Yang, P. Tzirakis, S. Liu, S. Mertes, E. André, R. Fu, and J. Tao. An Overview of Affective Speech Synthesis and Conversion in the Deep Learning Era. *Proceedings of the IEEE*, 111(10):1355–1381, 2023.
- [63] Unknown. Aphasia. *The New England Journal of Medicine*, 326(8):506–511, 1992. Downloaded from nejm.org at Cambridge University Library on August 17, 2024. For personal use only. No other uses without permission. Copyright © 1992 Massachusetts Medical Society. All rights reserved.
- [64] J. A. Urigüen and B. G.-Z. Soto. EEG artifact removal—state-of-the-art and guidelines. *Journal of Neural Engineering*, 12, 2015.
- [65] A. Uusberg, R. Thiruchselvam, and J. J. Gross. Using distraction to regulate emotion: Insights from EEG theta dynamics. *International Journal of Psychophysiology*, 91(3):254–260, 2014.
- [66] S. Valenzi, T. Islam, P. Jurica, and A. Cichocki. Individual Classification of Emotions Using EEG. *Journal of Biomedical Science and Engineering*, 7(08):604–620, 2014.
- [67] V. N. Vapnik. Statistical Learning Theory. In *Proceedings of the Workshop on Statistical Learning Theory*, 1998.
- [68] J. Yamagishi. Advanced speech synthesis technologies for vocal disabilities. *Journal of Information Processing and Management (情報管理)*, 57(12):882–889, 2015.
- [69] J. Yamagishi, C. Veaux, S. King, and S. Renals. Speech synthesis technologies for individuals with vocal disabilities: Voice banking and reconstruction. *Acoustical Science and Technology*, 33:1–5, 2012.
- [70] H.-W. Yoon, O. Kwon, H. Lee, R. Yamamoto, E. Song, J.-M. Kim, and M.-J. Hwang. Language Model-Based Emotion Prediction Methods for Emotional Speech Synthesis Systems. In *Interspeech 2022*, pp. 4596–4600. International Speech Communication Association (ISCA), 2022.
- [71] M. H. Yun, J. H. Lee, H. joo Lee, and S. Cho. Classification of Bluffing Behavior and Affective Attitude from Prefrontal Surface Encephalogram During On-Line Game. In *Proceedings of the International Conference on Biometrics (ICB 2006)*, Vol. 3832 of *Lecture Notes in Computer Science*, pp. 706–712, Berlin, Heidelberg, 2005. Springer-Verlag Berlin Heidelberg.
- [72] T. Zhang, A. El Ali, C. Wang, A. Hanjalic, and P. Cesar. RCEA: Real-time, Continuous Emotion Annotation for Collecting Precise Mobile Video Ground Truth Labels. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–15, New York, NY, USA, 2020. Association for Computing Machinery.
- [73] X. Zhu and L. Xue. Building a controllable expressive speech synthesis system with multiple emotion strengths. *Cognitive Systems Research*, 59:151–159, 2020.
- [74] 泰子 黒澤, 彩加 瀧本, 駿太 前田, 充展 國見, 和哉 井上. 複数の指標を同時に測定することの意義・現状・課題. *The Proceedings of the Annual Convention of the Japanese Psychological Association*, 2021.