

表情・音声・テキスト情報から隠している緊張状態を検出する手法の検討

Investigation of a Method for Detecting Hidden Nervous States from Facial Expression, Voice, and Spoken Text Data

入江 陸右
IRIE Rikuu

広島市立大学 言語音声メディア工学研究グループ
Language and Speech Media Engineering Group, Hiroshima City University

概要 本研究では、ユーザが隠している緊張を検出するために必要な特徴量として、従来研究で相関が示されているバイタルサインの代わりに非接触で取得できる表情と発話音声の有効性について有効性を検討する。本稿では、機械学習を行うための「ユーザが緊張を隠す」データセットを作成するために必要な実験手法、実験手順について述べる。

1 はじめに

近年メンタルヘルスの重要性に注目されるようになり、会社や学校でもそのような相談を受けられる環境が整いつつある。しかし、強いストレス状態にもあるにも関わらず周囲に気を使って平常を装うことで、周りの人々が本人の精神的不調に気づけないという事態も起こっている。もし周りの人々や本人が早期に精神的不調に気づくことができれば、医療機関やカウンセリングの受診を奨めることで重篤な事態を避けられる可能性がある。

そのような背景から、いくつかの従来研究において緊張状態とバイタルサインの関係性について分析が行われている。西條ら[1]は、緊張状態と皮膚温度の低下に相関があることを指摘している。また、橋口ら[2]は、平静時と比べ緊張状態の時に心拍数が高くなったという実験結果を示している。しかし、バイタルサインを計測するには接触型デバイスが必要であり、社会実装するうえで利便性に難がある。

そこで本研究では、バイタルサインの代わりに非接触型デバイスで取得できる表情、音声、テキスト情報を用いて、人が隠している緊張状態を推定する手法を提案する。また、表情、音声、テキスト情報とバイタルサインの相関についての分析も行う。本稿では、緊張状態推定手法の構成および実験用データの収集手法について述べる。

2 緊張状態推定手法

ユーザが隠している緊張状態を、表情、音声、テキスト情報から検出するシステムの概要を図1に示す。①学習フェーズでは、まずデータ提供者の発話から音響情報およびテキスト情報を取り出し、同時にデータ提供者の表情変化をカメラで収録する。それらにデータ提供者自身による緊張度を正解ラベルとして付与したものを学習データとして機械学習に学習させる。入力データの詳細については3節で説明する。次に②推定フェーズでは、学習に用いていないユーザの発話音声と表情を入力として、緊張度を推定する。さらに③分析として、表情、音声、テキスト情報とバイタルサインの相関についての分析を行う。

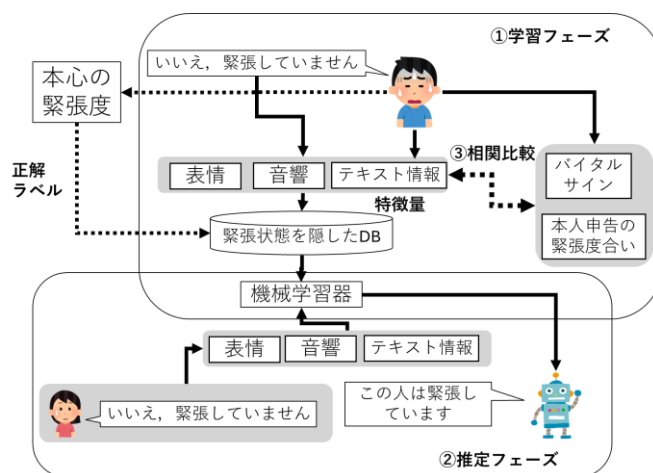


図1: 提案手法のシステム構成

表1: データセット作成に使用する機材一覧

収集するデータ	使用する機材
表情	Web カメラ
発話音声, 発話テキスト	高指向型コンデンサマイク
バイタルサイン	e4 リストバンド

今回は非接触での緊張検出を目指すほか、バイタルサインとの関係性について考察する。そのため、機械学習器による推定結果および自己申告した緊張度と、収録時のバイタルデータとの間で比較分析を行う。加えて、バイタルサインと非接触データとの相関についても分析する。

3 緊張状態の推定に用いる特徴量

提案手法を実現する機械学習器を構築するには、緊張状態において表情、音声、バイタルサインを計測したデータが必要であるが、そのようなデータセットが公開されていないため、自作する必要がある。そこで本節では、データセットの詳細やデータの収集方法について述べる。

表1はデータを取る際に使用する機材の一覧である。本実験では、表情を録画するためにWebカメラを使用する。この機材にはマイクも搭載されているが、今回はより性能の良い高指向型コンデンサマイクを別途使用する。発話音声の音響特徴量はopenSMILEを用いて算出する。また発話テキストは音声認識結果を手で修正したものを用いる。

また今回は表情、音響の他にバイタルサインも抽出するためe4リストバンド[3]を使用する。e4リストバンドの仕様については3.1節で詳細に記述する。



図 2 : e4 リストバンド[1]



図 3 : 実験データ収集実験の実施環境

3.1 バイタルサイン

バイタルサインとは生命徴候を意味し、大きく脈拍、呼吸、体温、血圧の 4 項目を基本とする指標である。e4 リストバンドはそれらの指標を光学式心拍センサ、皮膚電位センサ、3 軸加速度計、光学温度計の 4 種類のセンサを用いて計測する (図 2 参照)。さらにそれらの計測結果を用いて容積脈波、心拍間隔値、皮膚電気活動、XYZ 軸加速度、皮膚温度の 5 種類のデータを算出することができる。

関連研究の結果[2, 3]より、緊張状態の時、人間は皮膚温度の低下および心拍数の上昇といったバイタルサインの変化がみられる。そこで本研究では、e4 リストバンドで得られる 5 種類のデータのうち、容積脈波と皮膚温度に注目することでバイタルサインから緊張を読み取ることができる。この緊張とバイタルサインの相関は、隠れた緊張を検出するための非接触データの有効性を調べる際に用いる。

4 緊張状態データの収集手順

データ収集の手順を以下に記す。本研究では人為的に緊張状態を作るのではなく、日常生活において緊張するシチュエーションの前後にデータ収集を行う。緊張するシチュエーションは人によって様々だが、本研究では“大勢の前でスピーチをするとき”と、“目上の人と会話するとき”の 2 つの場面を対象とする。これ以降、本研究で対象とする緊張するシチュエーション (大勢の前でスピーチするとき、目上の人と会話するとき) を本番と表す。今回は実験参加者として大学生を想定しているため、データを取る具体的な場面として、“講義での研究成果プレゼンテーション”と“大学教員との個人面談”を予定している。

本研究では本番前および本番後にデータを取る。本番前には実験参加者が緊張を隠しているシチュエーションを集めるため、本番後では実験参加者が緊張していない状態を集めるためである。

実験データ収集手順は以下のとおりである。

1. 本番前直前の実験参加者に質問 (図 4) をして、実験参加者の様子を収録する
2. 収録後、実験参加者に現在の緊張度合いを質問する

質問者	今緊張していますか？
実験参加者	いいえ、緊張していません。
質問者	本当に大丈夫ですか？
実験参加者	はい、大丈夫です。
質問者	本番に向けての心意気はありますか？
実験参加者	(自由回答) (例) 夜通し練習してきたので自信満々です。

図 4 : 本番前の実験参加者に質問する台本

質問者	今緊張していますか？
実験参加者	いいえ、緊張していません。
質問者	本当に大丈夫ですか？
実験参加者	はい、大丈夫です。
質問者	本番が終わってどうですか？
実験参加者	(自由回答) (例) 緊張しましたが、非常にうまくいきました。

図 5 : 本番前の実験参加者に質問する台本

3. 本番を終えた直後の実験参加者に質問 (図 5) し、実験参加者の様子を収録する
4. 収録後、実験参加者に現在の緊張度合いを質問する

図 3 にデータを収集する際の実施環境を示す。実験参加者の正面の机の上に表情と音声を記録するためのノートパソコン、Web カメラ、高指向型コンデンサマイクを配置し、右腕に e4 リストバンドを着用し記録を行う。記録は 1 対 1 の対面で行い、質問者は実験参加者の隣に座り決められた質問をする。

図 4 は本番前の実験参加者が緊張を隠している様子を収録するための台本である。一つ目の質問には緊張しているか否かに関わらず大丈夫だと答えてもらう。これにより、緊張を隠している際の応答状況を収録する。二つ目の質問は実験参加者により緊張感を与えるために大丈夫かと念を押す質問をする。この質問にも本人の状態関係なく大丈夫だと答えてもらう。さらに、自由発話によるテキスト情報

の違いを収集するため、本番に向けての心意気を聞く。この時のセリフは、言葉の癖や言葉のポジティブ、ネガティブを検出するため、セリフ固定ではなく自由にしゃべってもらう。ただし、緊張を隠している状態であることが重要であるため、緊張していないふりの発言をしてもらう。データ収集後、最後に実験参加者が実際にどのくらい緊張しているかについて、緊張度合いを「緊張している」、「少し緊張している」、「あまり緊張していない」、「緊張していない」の4つの選択肢から回答してもらう。ここでは実験参加者の本心で答えてもらう。

図5は本番後の実験参加者への質問の台本である。基本的に図4の本番前の実験参加者への質問と変わりはないが、最後の質問では本番前に喋ってもらった心意気とは別にその時の思ったことをそのまま喋ってもらう。データ収集後の緊張度合いの調査も、本番前と同様に行う。

5 今後の方針

本報告では、現段階で計画している実験に用いる機器や実験手順についてまとめた。現在本実験について倫理審査申請中のため、審査が通り次第本稿の内容で50件程度のデータを収録する予定である。

データ収録後は、表情、音響、テキスト、バイタルサインの4つの特徴量を基に実験参加者が隠している緊張を考察、分析をしていく予定である。表情画像からは顔がこわばったり、ひきつっているような表情、音声からは声の震え、裏返り、早口など、発言からはネガティブな言葉などの特徴量が検出でき、それらに基づいて隠された緊張状態が検出できると想定している。また、バイタルサインを用いた推定との性能比較や、バイタルサインと相関がある特徴量についても調査を行う予定である。

参考文献

- [1] 西條淳夫, 安田昌司, 大隈正人: 皮膚温による感覚推定, 日本赤外線学会誌, Vol. 8, No. 1, pp.23-30 (1998) .
- [2] 橋口泰武: 緊張時における心拍数の分析, 千葉体育学研究, Vol. 14, pp.49-54 (1991) .
- [3] empatica: E4 wristband technical specifications, empatica (online), available from <<https://support.empatica.com/hc/en-us/articles/202581999-E4-wristband-technical-specifications>> (accessed 2024-6-15).