PLEN Robotics

RimTech

感情・ストレス度の推定方法、その先にある発話音声解析の可能性とは?

目次

1 イントロダクション

2 感情・ストレス度の推定方法

3 発話音声解析の利用可能性

4 Care Cube搭載のMotivelの優位性

イントロダクション

ChatGPTが、2022年11月、OpenAIにより公開されたことで、生成AIに対する関心が非常に高まっています。日本からのChatGPT を提供するOpenai.comへのアクセス数は、2023年5月中旬に767万回/日に達したとされています。世界的に見ると、米国、インドに次いで、3番目に多いとされています。
また、ChatGPTを活用した関連サービスも多く登場しています。

一方、現状、生成AIが実現できていない領域には、ヒューマンセンシングによる感情・ストレス度の推定があります。生成AIは、テキストベースの情報処理に依拠していることから、文章から感情・ストレス度を推定することはできるとされているものの、ヒトの生体情報を用いた感情・ストレス度の推定はできないとされています。

感情・ストレス度の推定方法

現在、ヒトの感情・ストレス度を推定する方法には、様々なものが使用されています。これら方法を大別すると、直接的に、ヒトと接触してセンシングする方法（以下、「接触型」という。）と、接触しないでセンシングする方法（以下、「非接触型」という。）に分けられます。

1 接触型

接触型の方法には、手首（以下、「リスト」という。）に装着する端末、耳に装着する端末、頭に装着する端末、胸部等に装着する端末等が使用されています。特に、リストに装着する端末は、時計型の端末として広く普及しており、心拍数の計測データから感情・ストレス度を推定しています。また、睡眠の質を推定するサービスもあります。なお、胸部等に装着する端末も心拍数を計測します。

耳に装着する端末は、イヤホン型の端末として普及しており、脳波を計測することで、感情・ストレス度の推定を行うとしています。脳波を計測する観点では、頭に装着する端末も同じような役割を果たしています。

接触型のメリットは、身体に密着して計測するため、総じて、精度の高い計測ができる点にあります。逆に、デメリットには、端末を身に付ける行為（以下、「アクション」という。）が煩雑であること、端末の購入代金が負担になるという点が主に指摘されています。

2 非接触型

非接触型の方法には、映像解析、テキスト解析、発話音声解析等があります。映像解析は、カメラを使用して体表面の変化を計測し、感情やストレス度の推定を行っています。例えば映像脈波という技術は、血液中のヘモグロビンの波長495～570nmの緑色の可視光を吸収するという性質を利用して、映像の中から緑色輝度成分の時系列分析を行い、脈波信号を抽出します。こうして得られた体表面の変化から心拍数を推定し、その心拍数を用いて、感情・ストレス度を推定することが可能となっています。

また、顔の表情、筋肉の動き、視線や瞳孔の変化から感情を推定する技術もあります。テキスト解析は、電子メール、グループワーク用のアプリケーション（Teams、Slack等）、SNS等に書き込まれたテキスト情報を単語や文節に分割し、その出現頻度や相関関係などから、書き込んだヒトの感情・ストレス度の推定を行っています。発話音声解析は、ヒトの声に含まれる音声特徴量を抽出して、当該音声特徴量から感情・ストレス度の推定を行っています。

これら非接触型のメリットは、計測用の端末が身体に直接的に触れないため、ストレスが掛かりにくい点や、導入時の端末購入代金の負担が総じて小さい点が指摘されています。特に、発話音声解析は、必要なのは声だけであり、既存のスマートフォン、タブレット、パソコン等があればそれらをそのまま使用できるため、計測機器への追加投資負担が軽いというメリットが指摘されています。

		接触型		非接触型	
	時計型 端末	イヤホン型 端末	映像 解析	テキスト 解析	発話音声 解析
初期コスト	×	×	△	○	○
運用コスト	△	△	△	△	○
手軽さ	△	△	△	○	○
推定精度	○	○	△～○	△	△～○
多様性※	×	×	○	×	○

注）上表の作成は、リスク計測テクノロジーズ株式会社の調査に基づきます。
※　ユーザーの年齢幅、多言語対応等

表1　感情・ストレス度の推定方法の比較

発話音声解析の利用可能性

(1) 心身の状態の可視化

発話音声解析は、計測のために発声するというアクションが要求されるものの、声という生体情報を利用するため、たとえ発声時間が短時間であっても、発話者の感情・ストレス度に関して非常に多くの情報を得ることができます。そのため、こうした情報量の多さを利用して、他の状態の推定にも利用することができます。例えば後述するMotivelでは、活動意欲（以下、「モチベーション」という。）、集中力、ヒヤリハットリスクの検知、睡眠リスクの検知が行われています。モチベーションや集中力は、計測時点から2週間以内に、これら指標が低下する可能性を数値で示します。ヒヤリハットリスクの検知は、様々なヒヤリハット（例えば、仕事上のケアレスミス等）が発生する可能性を数値で示します。睡眠リスクの検知は、2時間以内に眠気を感じる可能性を数値で示します。

ここで、睡眠リスクの検知について、
説明します。発話音声を用いた睡眠リスクの検知機能は、2022-2023年にかけて、開発が行われました。睡眠リスクは、2時間以内に眠気を感じる可能性がどの程度あるのかを確率的に数値で示すものです。

この確率が高いほど、2時間以内に眠気を可能性が高いことを意味します。そのため、確率が高い場合には、眠気の発生を回避する対策を取ることで、効果的に眠気を防止し、眠気から生じる事故等の様々なリスクを回避することが期待されます。睡眠リスクの検知の開発では、眠気を感じることに個人差があるため、アンケートを通じた眠気の有無に関する回答では、客観的に眠気を評価できないという問題がありました。そこで、眠気に関する客観的な指標として、心拍数を採用し、声に含まれる音声特徴量と心拍数の関係性を調査しました。心拍数を採用した理由は、ヒトは、副交感神経が優位になると眠気を感じやすくなるという関係性を利用するためです。

ヒトは、この副交感神経が優位になると、通常時に比べて、心拍数が低下します。つまり、声の音声特徴量から、心拍数の低下を予測できれば、最終的に眠気を感じる可能性を予測できるとしています。

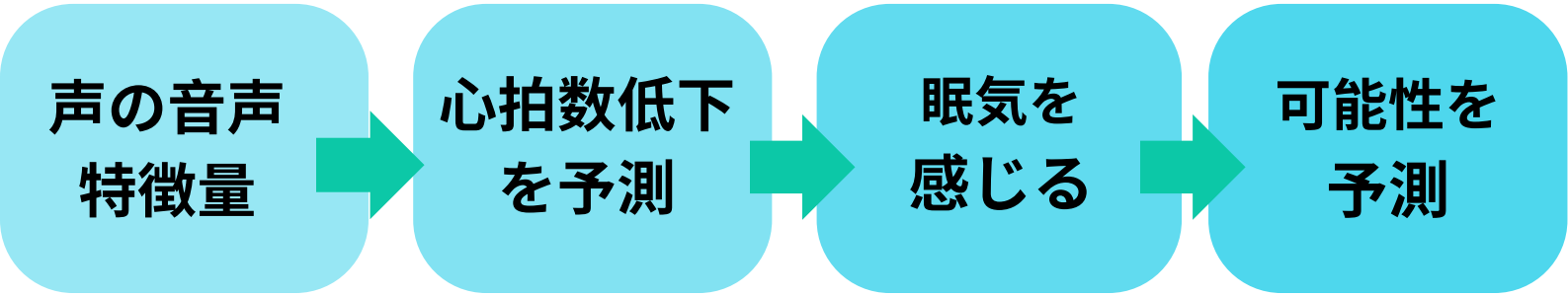

図1 Sleepy Meterの眠気リスクの検知方法

睡眠リスクの検知機能の開発にあたり、RimTechは2022年に横須賀市、小田原市、相模原市、厚木市、横浜市の消防局の協力を得て、消防局職員の発話音声データを収集しました。また、心拍数データの取得に向けて、ウェアラブル端末を常時身に付けた協力者からも、発話音声データを取得しました。その結果、4000回の発話音声データと、300万件を超える心拍数データを得ることができ、睡眠リスクの検知を行う音声解析エンジン（以下、「Sleepy Meter」という。）が完成しました。

(2)認知機能の動向予測

将来的な利用可能性としては、発話音声によるヒューマンセンシングでは、認知機能の低下を検知する領域でも活用が期待されています。高齢者の言葉にならない思いを知る技術として、大学の研究者や介護施設のスタッフの間で意見交換が行われています。また、幼児や子どもの状態を知る技術としても活用が期待されており、こちらも大学の研究者とディスカッションが行われており、今後の研究の進展が待たれます。

(3)エンターテーメント、マーケティング領域への応用

この他、音声解析はエンターテーメントやマーケティングの分野でも活用できる可能性があります。発話音声解析に必要な音声ファイルは、データ容量が非常に小さいことが特徴です。サンプリングレート16,000Hzの音声ファイルの容量は、3秒間で概ね107KB程度であるため、比較的に低速なインターネット通信環境（300Kbps～1Mbps）であっても、クラウド型のAPI（ソフトウェアやアプリケーション機能の共通ツール）で快適に計測ができます。

さらに、Care Cube等のエッジ側で計測する場合でも、大きな計算処理能力を必要としないため、軽快に計測ができます。これにより進行中のエンターテーメントや商談の際の顧客の感情変化や集中度を計測し、リアルタイムで対応を変化させることにより、顧客の満足度に大きな影響を与えることも可能になります。実例として、RimTechは、ヒトのセンシング技術でより人々の生活に即したプロダクト開発を行うSHIN4NY株式会社（以下、「SHIN4NY」という。）と協力し、クラウド上に設置した音声解析エンジンで高速かつ高頻度の計測を実現しています。

SHIN4NYは、2022年6月、eスポーツ大会となる「SHIN4NY CUP」を主催し、ゲーム実況者の声をリアルタイムに計測するだけでなく、YouTubeライブでの配信が行われました。こうした技術を応用することで、アバターを使用するメタバース空間上において、相手の状態を知るための方法としても活用することが期待できます。

Care Cube搭載の Motivelの優位性

Care Cubeには、RimTechが独自に開発し、提供を行っているMotivelが搭載されています。Motivelを構成する要素は、（1）音声解析エンジン、（2）データアナリティクス、（3）利用シーンに応じた運用ノウハウの3つがあります。これら（1）から（3）に関してMotivelの優位性について解説します。

1 音声解析エンジン

Motivelは2020年3月にリリースされ、ユーザーを通じて様々な環境下で使用されてきました。そうした中、環境ノイズ（空調機やエスカレーターの運転音、自動車の走行音等）の多い場所でも、高精度かつ安定した計測を実現したいという声が多く寄せられました。環境ノイズには生活騒音、道路交通騒音、建設騒音、工場騒音など様々なものがあり、そのレベルは夜間の居室内で30～40db、飲食店内で60～70db、地下鉄車内で70～80dbとされています。

こうした環境ノイズが音声データと混入すると音声解析の精度に悪影響を及ぼすことがあります。そこで、Motivelの開発では、ユーザーが実際に使用される場所の環境ノイズの調査を行い、2022年秋に高いノイズ耐性を獲得しました。この結果、ユーザーからは「安定した計測結果を得られる」「再現性が高い」「利用シーンが増える」などのコメントが寄せられています。

また、刺激が脳内に到達するのに視覚が20〜40ミリ秒要するのに対して、音声は8〜10ミリ秒という研究があります。聴覚は視覚の８倍の情報処理能力があるとも言われ、これに対応するためMotivelは、高速計測・高頻度計測に適した音声解析エンジンとなっています。Web-APIを使用した計測は、4G回線以上のインターネット通信環境であれば、1秒未満で計測結果を手元で確認することができます。なお、エッジ側で計測するエッジ版APIでも、同様の結果となります。また、3秒間の音声データをWeb-APIに送信することで、リアルタイムでの高頻度計測を実現しており、結果もリアルタイムで確認することが可能となっています。さらに重要な点は、Motivelは独自の特徴量抽出アルゴリズムにより、発話時間が短くても精度の高い結果を得ることができます。

2022年秋にリリースしたMotivelは、当初5秒間の録音データを必要としていましたが、現在は「声だけ3秒」を掲げて、3秒間の音声データでの計測を実現しています。なお、実際には3秒未満の音声データで高い精度を実現しています。そのため、「声だけ3秒」というのは、実際の利用シーンを想定した余裕のある音声データの時間となっています。

2 データアナリティクス

2020年にMotivelの最初のバージョンをリリースした後、60万件を超える音声データの解析が行われ、多くの専門家と共に当該音声解析の結果からヒトの状態を推定する試みを、継続しています。こうした実績を踏まえて、1回の計測結果から状態を推定するに留まらず、複数回の計測結果（時系列データ）から状態を推定することにも取り組んでいます。

例えば、ヒトは抑圧された状態（高ストレス状態）にあると、感情の起伏が乏しくなる傾向があることから、計測結果の推移変動を参照し、一定の水準以上の抑揚がなくなる場合には、高ストレス状態にある可能性が高いと推定しています（図2）。

Motivelは、実際の現場で計測した大量のデータを基礎に、ヒトの状態の推定に関する実践的な判断・解釈を、日々積み重ねています。こうした実践的な判断・解釈の記録（データ）が増加することも、大量のデータと同じく、Motivelの信頼性を支える重要な要素となっています。

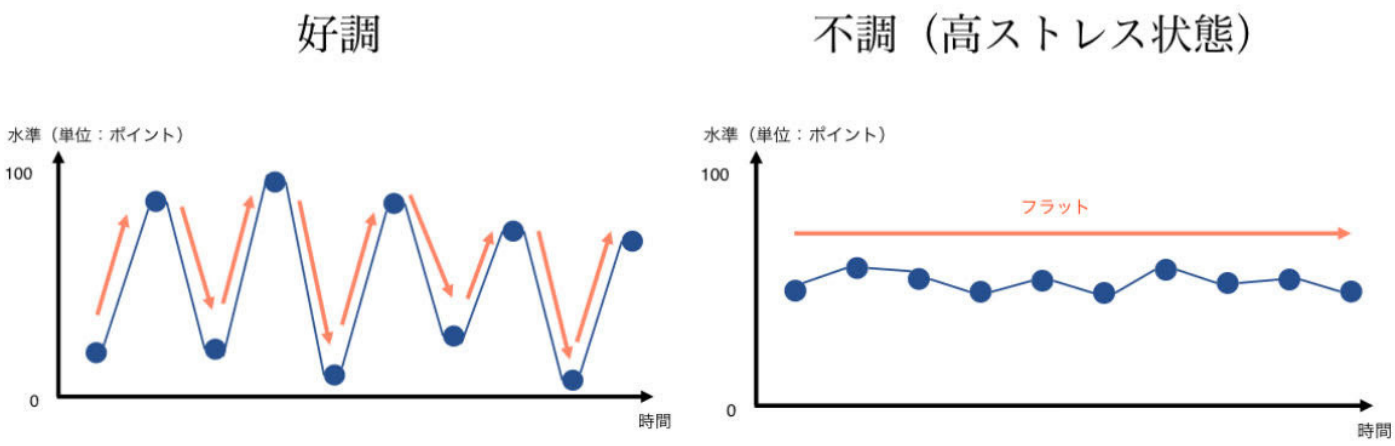

図2　任意の感情指標の推移変動から推定する好調・不調の判断

3 利用シーンに応じた運用ノウハウ

Motivelは、様々な用途、場所で利用されるため、その時々の条件に合わせたチューニングが必要になります。このチューニングは、ハードウェアで調整することもあれば、ソフトウェアで調整することもあり、日々、新たな課題に直面しています。例えば、音声を収録するマイクですが、世の中には数多くのメーカーから様々なタイプのマイクが発売されています。そして、それぞれのマイクの種類及びその特性に応じて、録音し易い周波数帯域は
異なります。こうしたマイクの特性も踏まえて、常に安定かつ高精度な計測結果が得られる運用ノウハウが蓄積されています。