

令和 4 年 6 月 19 日現在

機関番号：32503

研究種目：基盤研究(C) (一般)

研究期間：2017～2021

課題番号：17K00160

研究課題名(和文)感情の連続的な変化を推定するリアルタイムマルチモーダル感情認識システムの構築

研究課題名(英文)Real-time multimodal emotion recognition system that estimates emotional dynamics

研究代表者

有本 泰子(Arimoto, Yoshiko)

千葉工業大学・情報科学部・准教授

研究者番号：60586957

交付決定額(研究期間全体):(直接経費) 3,600,000円

研究成果の概要(和文):マルチモーダルオンライン対話データに含まれる感情3次元の評価値を正解とし,計測した話者の表情・音声・自律神経系反応から3次元の感情(快-不快,覚醒-睡眠,支配-服従)を推定するモデルを構築し,その検証を行った。その結果,テキストチャット場面の感情推定モデルでは,正解となる感情の評価値と推定値との誤差が快-不快で0.32,覚醒-睡眠および支配-服従では0.36となった。また,音声チャット場面のモデルでは,平均二乗誤差(MSE)が快-不快で0.121,覚醒-睡眠で0.161,支配-服従で0.164となった。さらに,リアルタイムに話者の情報を取得し,話者の感情を推定するシステムの実装を行なった。

研究成果の学術的意義や社会的意義

本研究の成果により,オンライン対話場面において,欠落する感情を補完する仕組みを実現することが可能となる。例えば,インターネットを利用した音声チャットやテキストチャットにおける対話場面で,本システムの感情推定技術を用いれば,話者の感情を推定した結果を対話相手に提示することができる。新型コロナウイルスの蔓延により,顧客との打合せや大学の授業などもオンライン環境下で行うようになった。オンライン環境下で感情コミュニケーションに齟齬が生じ,本来の目的を達成することに困難を感じることも多い。本研究課題で構築するシステムは,こうした問題を解決しうる世界初のリアルタイムマルチモーダル感情推定システムとなる。

研究成果の概要(英文):Multimodal emotion recognition models to estimate 3-dimensional emotions (pleasant-displeased, arousal-sleep, and dominance-submission) using speaker's facial and vocal expression and autonomic nervous system responses during online conversation were constructed and evaluated. The results showed that the difference between the estimated emotional values of the correct emotional values in the text chat situation was 0.32 for pleasant-displeased, and 0.36 for both arousal-sleep and dominance-obedience. In the model for the voice chat situation, the mean square error (MSE) was 0.121 for pleasant-displeased, 0.161 for arousal-sleep, and 0.164 for dominance-submissiveness. In addition, real-time multimodal emotion recognition system was implemented for estimating speaker's emotional dynamics during online conversation.

研究分野: 音声言語情報処理

キーワード: 感情認識 マルチモダリティ 音声 表情 心拍 発汗 機械学習 テキスト

1. 研究開始当初の背景

我々のコミュニケーションは対面によるコミュニケーションからデジタルデバイスが介在するオンラインコミュニケーションへと変容している。オンラインコミュニケーションは現実のコミュニケーションと比較すると、映像・音声・テキストなど媒介するコミュニケーションメディアの違いにより、表情や声の抑揚などの非言語情報が消失し、伝達される情報に欠落が生じる。そのため、利用者が現実世界で直接コミュニケーションを取るときと同じ感覚でオンラインコミュニケーションを図っていても、話者の感情などが十分に伝達されているとはいえず、コミュニケーションに齟齬を生じさせている。つまり、オンラインコミュニケーション場面に暴露される人々は、対話相手の感情状態を適切に理解することができず、対話相手の感情に無関心な状態となってしまうたり、相手を怒らせまいと必要以上に警戒したりする状態になる可能性がある。こうしたコミュニケーションの齟齬を解消し、オンライン場面でも現実世界の対面場面と同等の感情コミュニケーションを実現するためには、非言語情報を利用してオンラインコミュニケーション手段を人間の拡張する新規な情報表現へと変換し、欠落する感情を補完する仕組みが必要である。本研究では、オンラインコミュニケーション場面で欠落する感情を補完するため、表情・音声・自律神経系反応など計測可能な話者の情報から感情をリアルタイムに推定するマルチモーダル感情推定システムを開発し、その評価を行うことを目的とする。

これまでもいくつかの感情推定システムが開発されてきた。しかし、その多くが表情のみや音声のみなどのシングルモダリティから感情を推定するシステムであった。また、マルチモーダルな情報から感情を推定するシステムが開発されていても、使用される情報は表情と音声の組合せがほとんどで、感情を正直に反映する自律神経系反応を組合せて感情を推定するシステムの開発はない。自律神経系反応を感情推定システムに組み込むことのメリットは、表情や音声など他者に感情を伝達する際に利用可能な外在化反応のみに頼らなくても良い点にある。オンラインコミュニケーション環境下では、感情を伝達するために利用する表情や音声などの情報が遮断されるため、表情や音声に対する話者の感情表出が意識的・無意識的に抑制され、それらの情報を感情推定に利用できない可能性がある。そのため、話者本人の正直な感情反応が表れる内在化反応のひとつである自律神経系反応を本研究課題で構築するリアルタイムマルチモーダル感情認識システムに利用し、感情の推定精度向上を目指す。本研究課題で目指すシステムの構築が実現すれば、世界初の音声・表情・自律神経系反応を利用したマルチモーダル感情推定システムとなるであろう。

2. 研究の目的

本研究では、表情・音声・自律神経系反応など計測可能な話者の情報から感情をリアルタイムに推定するマルチモーダル感情推定システムを開発し、その効果を評価することを目的とする。

a. オンラインコミュニケーション場面で感情を推定するために必要な計測情報の特定化

これまでに収録してきた3種類の異なるチャットを用いたマルチモーダル対話の分析を行い、コンピュータを介したコミュニケーション場面で表出する感情を推定するためのモデルを構築する。感情推定モデルを構築することで、オンラインコミュニケーション場面の感情推定に必要な特徴量を特定する。モデルの構築にあたり、感情が対話の遷移に沿って動的に変化することを考慮し、カルマンフィルタなどの状態空間モデルを念頭に分析を進める。

b. リアルタイムマルチモーダル感情推定システムの実装

aで構築した感情推定モデルをもとに、リアルタイムマルチモーダル感情推定システムを実装する。異なる3種類のオンラインコミュニケーション場面を想定しているため、感情推定部には複数の感情推定モデルを実装する予定である。リアルタイムマルチモーダル感情推定システムの客観評価として、感情推定モデルごとの推定率を求め、推定率を向上させるように推定器の精度向上を図る。本研究課題では、実際の対面コミュニケーションと同等の感情コミュニケーションを実現することを目標とするため、対面コミュニケーション場面のマルチモーダル対話データを利用して構築した感情推定モデルの感情推定精度と同等レベルの推定精度を目指す。

c. 主観評価実験によるリアルタイムマルチモーダル感情推定システムの有効性検証

主観評価実験を実施し、実装したリアルタイムマルチモーダル感情推定システムのユーザー評価を行って、その有効性を検証する。ここでも異なる3種類のオンラインコミュニケーション場面ごとに、システムの評価を行わせる予定である。

3. 研究の方法

マルチモーダル感情推定モデルの構築には、これまでに収録した100名50組のマルチモーダルオンライン対話データを使用する。収録済みのマルチモーダルオンライン対話データは、対話中に話者が表出した感情的反応（表情・音声・自律神経系反応（表情筋・心拍・発汗））と感情状

態の変化（主観評価による感情の動的評定値）が既に収録されており、本研究課題で行う機械学習によるマルチモーダル感情推定モデルの構築の歳の学習・評価用データとして最適である。マルチモーダルオンライン対話データに含まれる感情評定値を正解とし、計測した話者の表情・音声・自律神経系反応から感情を推定するモデルを構築する。感情が対話の遷移に沿って変化することから、モデル構築に使用する推定器は状態遷移をモデル化する機械学習を検討する。リアルタイムマルチモーダル感情推定システムの実装では、入力部として簡易なカメラとマイクおよび筋電図・心電図・皮膚電気反応が計測可能なセンサーから、話者の表情・音声・表情筋・心拍・発汗の情報を実時間で取得し、感情を推定した結果をリアルタイムに対話相手に呈示する枠組みを考案し、システムに実装する。

4. 研究成果

機械学習によって構築するマルチモーダル感情推定モデルの学習・評価用データの整備を行った。マルチモーダル感情推定モデルの構築には、これまでに収録した100名50組のマルチモーダルオンライン対話データを使用する。本データは、実際にシステムを応用する場面を想定し、友人同士の対話参加者に実際にオンラインチャットをさせ、その場面の話者の情報（表情・音声・自律神経系反応（表情筋・心拍・発汗））を記録している。コミュニケーションに利用可能なモダリティをコントロールし、比較するため、チャットの種類はビデオチャット・音声チャット・テキストチャットの3種類を用意し、参加者にはいずれかひとつのチャット環境で対話をさせている。さらに、対話参加者に自分自身と対話相手の感情3次元（快-不快、覚醒-睡眠、支配-服従）を対話の遷移に沿って、動的に評価をさせている。本対話データに含まれる大規模な情報を、感情推定モデルの入力値として利用する特徴量として計算し、マルチモーダル感情推定モデルの構築に利用可能な形に整備し直した。さらに、テキストチャットの会話内容を転記する作業と、自律神経系反応と映像・音声との同期をとる作業を行った。

マルチモーダルオンライン対話データに含まれる感情3次元の評価値を正解とし、計測した話者の表情・音声・自律神経系反応から3次元の感情（快-不快、覚醒-睡眠、支配-服従）を推定するモデルを構築し、その検証を行った。マルチモーダルオンライン対話データからチャットの種類にかかわらず27名のデータを使用した。抽出した特徴量は、音声6種類、表情の特徴量が9種類、心拍の特徴量が5種類、発汗の特徴量が6種類である。これらを使用して推定可能性を調べるため、個人に対してランダム効果を適用したマルチレベルモデルによって、モデルを構築し、その推定精度を求めた。その結果、決定係数が快-不快および覚醒-睡眠では0.45、支配-服従では0.47となった。

さらに、対話中の感情表出に表情筋が利用できるか分析を行なった。マルチモーダルオンライン対話データに含まれる31名のデータを利用した。ビデオチャット場面・音声チャット場面・テキストチャット場面の3種類のデータが含まれているため、これらを要因として表情筋（皺眉筋・大頬骨筋）の対話内平均値および標準偏差に対して分散分析を行なった。表情筋は表情形成時および発話時に強く反応が出るため、表情を形成する必要がなく、発話することもないテキストチャット時は反応が出ないことが予想された。分析の結果、大頬骨筋はビデオチャット・音声チャット・テキストチャットの順に大きく反応していること、テキストチャット場面でもある程度の反応が出ていることが明らかとなった。

テキストチャット場面における感情推定モデルの構築を行った。マルチモーダルオンライン対話データからテキストチャットで対話を行っている26名のデータを使用した。マルチモーダルオンライン対話データに含まれる感情3次元の評価値を正解とし、計測した話者の自律神経系反応（心拍・発汗）およびテキスト情報から3次元の感情（快-不快、覚醒-睡眠、支配-服従）を推定するモデルを構築し、その検証を行った。この時使用した感情評価値は、話者自身が自分の感情を評価したのではなく、対話相手からみてどのような感情状態にあるかを評価した値を用いている。また、使用したテキスト情報も、対話相手に示すためにチャットのタイムラインにあげるテキストではなく、タイムラインにあげる前にメッセージボックスに記入しているテキスト情報を利用した。メッセージボックスに記載されたテキストをWord2Vecを用いて単語ベクトルを取得した。心拍の特徴量はフレームサイズ4秒、シフト長50ミリ秒で6種類の特徴量を抽出した。発汗の特徴量も心拍と同様のフレームサイズおよびシフト長で6種類の特徴量を抽出している。言語情報は単語埋め込みを行ったのち、Long short-term memory (LSTM) を利用したネットワークで感情情報を推定するモデルを構築した。心拍・発汗の特徴量もLSTMで感情を推定するモデルを構築した。それぞれの出力を統合し、感情評価値の推定を行った。その結果、正解となる感情の評価値と推定値との誤差が快-不快で0.32、覚醒-睡眠および支配-服従では0.36となった。

音声チャット場面における感情推定モデルの構築を行った。マルチモーダルオンライン対話データから音声チャットによって対話を行っている18名のデータを使用した。マルチモーダルオンライン対話データに含まれる感情3次元の評価値を正解とし、計測した話者の自律神経系反

応（心拍・発汗）、対話中の音声信号および表情画像から3次元の感情（快—不快、覚醒—睡眠、支配—服従）を推定するモデルを構築し、その検証を行った。モダリティごとにCNN-LSTMのニューラルネットワークを構築し、まずはCNNで各モダリティの特徴量を抽出したのち、その後続くLSTM層で各特徴量の時間変化を捉えて感情を推定した。使用したモダリティはHR（R-R間隔）・SCL・表情・音声の4種類ある。HR・SCL・音声の各モデルでは、それぞれの信号を一次元CNNにより直接畳み込んだ。表情モデルでは、映像から顔の部分だけを抽出して、96×96ピクセルにリサイズしたカラー画像を使用し、CNNにはResNet50のネットワークと同等のものを使用した。LSTMのシーケンスの長さはいずれのモデルも200とした。ユニモーダルモデルの出力を統合する方法として、CNNからの出力を結合するfeature level fusionとLSTMから出力される情報を結合するdecision level fusionを実施し比較した。その結果、feature level fusionでは平均二乗誤差（MSE）が快—不快で0.101、覚醒—睡眠で0.205、支配—服従で0.228となった。また、decision level fusionではMSEが快—不快で0.121、覚醒—睡眠で0.161、支配—服従で0.164となった。

リアルタイムに話者の情報（表情、音声、自律神経系反応（心拍・発汗））を取得し、話者の感情を推定するシステムの実装を行なった。具体的には、マルチモーダル情報入力部・特徴量抽出部・感情推定部・推定結果表示部の実装を行なった。マルチモーダル情報入力部は簡易なカメラとマイクおよび心電図・皮膚電気反応が計測可能なセンサーから、話者の表情・音声・心拍・発汗の情報を実時間で取得する。特徴量抽出部は、話者の情報から感情を推定するために必要な特徴量の計算を行う。感情推定部の実装も行なったが、新型コロナウイルス感染拡大の影響から被験者を用いた評価実験が行えなかったため、ダミーのモデルを適用し、動作の確認を行うにとどめた。さらに、推定結果表示部では、話者の感情状態が時間の流れに沿って動的に変化する様子を呈示するため、3次元で構成される感情空間上に話者の感情状態の推定結果を示すマーカーを配置する3Dアニメーションを実装した。

本研究に関連する成果として、査読付き雑誌論文2件、国際会議論文5件（うち査読付き4件、招待講演1件）、国内学会発表19件（うち招待講演2件）がある。

5. 主な発表論文等

〔雑誌論文〕 計7件（うち査読付論文 7件/うち国際共著 0件/うちオープンアクセス 7件）

1. 著者名 有本 泰子、今西 利於、森 大毅	4. 巻 63
2. 論文標題 自然で表現豊かな笑い声合成に向けた感情情報からの笑い声の構成要素決定法	5. 発行年 2022年
3. 雑誌名 情報処理学会論文誌	6. 最初と最後の頁 1159 ~ 1169
掲載論文のDOI (デジタルオブジェクト識別子) 10.20729/00217618	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 永岡 篤、森 大毅、有本 泰子	4. 巻 73
2. 論文標題 感情音声コーパス共通化のための新たな感情ラベル推定における既存感情ラベル併用の効果	5. 発行年 2017年
3. 雑誌名 日本音響学会誌	6. 最初と最後の頁 682 ~ 693
掲載論文のDOI (デジタルオブジェクト識別子) 10.20697/jasj.73.11_682	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Takuto Matsuda and Yoshiko Arimoto	4. 巻 -
2. 論文標題 Acoustic discriminability of unconscious laughter and scream during game-play	5. 発行年 2022年
3. 雑誌名 Proc. Speech Prosody 2022	6. 最初と最後の頁 575-579
掲載論文のDOI (デジタルオブジェクト識別子) 10.21437/SpeechProsody.2022-117	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Hiroki Mori, Tomohiro Nagata, and Yoshiko Arimoto	4. 巻 -
2. 論文標題 Conversational and Social Laughter Synthesis with WaveNet	5. 発行年 2019年
3. 雑誌名 Proc. Interspeech 2019	6. 最初と最後の頁 520-523
掲載論文のDOI (デジタルオブジェクト識別子) 10.21437/Interspeech.2019-2131	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Yoshiko Arimoto and Hiroki Mori	4. 巻 -
2. 論文標題 Emotion Category Mapping to Emotional Space by Cross-Corpus Emotion Labeling	5. 発行年 2017年
3. 雑誌名 Proc. Interspeech 2017	6. 最初と最後の頁 3276-3280
掲載論文のDOI (デジタルオブジェクト識別子) 10.21437/Interspeech.2017-994	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Yoshiko Arimoto	4. 巻 -
2. 論文標題 Challenges of Building an Authentic Emotional Speech Corpus of Spontaneous Japanese Dialog	5. 発行年 2018年
3. 雑誌名 The International Conference on Language Resources and Evaluation (LREC2018)	6. 最初と最後の頁 6-13
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Arimoto Yoshiko, Horiuchi Yasuo, Ohno Sumio	4. 巻 -
2. 論文標題 Consistency of base frequency labelling for the F0 contour generation model using expressive emotional speech corpora	5. 発行年 2018年
3. 雑誌名 Proc. Speech Prosody 2018	6. 最初と最後の頁 398--402
掲載論文のDOI (デジタルオブジェクト識別子) 10.21437/SpeechProsody.2018-81	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計19件 (うち招待講演 2件 / うち国際学会 0件)

1. 発表者名 阿部将士, 有本泰子
2. 発表標題 Speech laugh が発生しやすい調音に関する対話ドメイン依存性の検証
3. 学会等名 日本音響学会2022年春季研究発表会講演論文集
4. 発表年 2022年

1. 発表者名 土井敦也, 有本泰子
2. 発表標題 WaveNetによる叫び声合成の実現に向けたコンテキストラベルの検討
3. 学会等名 日本音響学会2022年春季研究発表会講演論文集
4. 発表年 2022年

1. 発表者名 松田匠翔, 有本泰子
2. 発表標題 CNN-LSTM モデルを使用したゲーム場面の笑い声と叫び声検出モデルの構築
3. 学会等名 日本音響学会2022年春季研究発表会講演論文集
4. 発表年 2022年

1. 発表者名 松田匠翔, 有本泰子
2. 発表標題 ゲーム場面の笑い声・叫び声に対する音響的識別可能性の検証
3. 学会等名 日本音響学会2021年秋季研究発表会講演論文集
4. 発表年 2021年

1. 発表者名 福田樹人, 有本泰子
2. 発表標題 プレイヤーの無意識な笑い声に対して生成されるゲームイベントの効果に関する生理心理学的検証
3. 学会等名 電子情報通信学会HCGシンポジウム2021論文集
4. 発表年 2021年

1. 発表者名 有本泰子, 阿部将士, 津田逸成
2. 発表標題 Speech laughが発生しやすい調音に関する音声学的分析
3. 学会等名 日本音響学会2021年秋季研究発表会講演論文集
4. 発表年 2021年

1. 発表者名 有本泰子
2. 発表標題 オンライン時代の感情コミュニケーション
3. 学会等名 第12回産業日本語研究会・シンポジウム パネルディスカッション(招待講演)
4. 発表年 2021年

1. 発表者名 有本泰子, クー ユエン, 小山俊樹, 金子裕亮
2. 発表標題 Affect Burst検出に向けた笑い声および叫び声ラベリングとその音響分析に関する予備的検討
3. 学会等名 電子情報通信学会技術研究報告
4. 発表年 2019年

1. 発表者名 クー ユエン, 加茂直樹, 有本泰子
2. 発表標題 音響的特徴量によるゲームプレイ中の笑い声分類モデルの構築
3. 学会等名 日本音響学会2019年秋季研究発表会講演論文集
4. 発表年 2019年

1. 発表者名 小山俊樹, 有本泰子
2. 発表標題 ゲームプレイ中に表出した社会的な笑い声のラベリングとその音響分析
3. 学会等名 日本音響学会2020年春季研究発表会講演論文集
4. 発表年 2020年

1. 発表者名 金子裕亮, 有本泰子
2. 発表標題 音響的特徴量によるゲームプレイ中の叫び声の識別
3. 学会等名 日本音響学会2020年春季研究発表会講演論文集
4. 発表年 2020年

1. 発表者名 クー ユエン, 有本泰子
2. 発表標題 ゲームプレイ中の笑い声検出モデルの構築
3. 学会等名 日本音響学会2020年春季研究発表会講演論文集
4. 発表年 2020年

1. 発表者名 森大毅, 永田智洋, 有本泰子
2. 発表標題 WaveNetによる笑い声の合成
3. 学会等名 日本音響学会2018年秋季研究発表会
4. 発表年 2018年

1. 発表者名 有本泰子, 森大毅
2. 発表標題 対話中に表出した笑い声の声質分析
3. 学会等名 日本音響学会2018年秋季研究発表会
4. 発表年 2018年

1. 発表者名 有本泰子, 野村直人, 加茂直樹
2. 発表標題 キャラクターエージェントに対する他者性認知の客観的評価
3. 学会等名 日本音響学会2019年春季研究発表会
4. 発表年 2019年

1. 発表者名 有本泰子
2. 発表標題 感情音声コーパスの構築と利用：いかにして生の感情を対象に研究を行うか
3. 学会等名 音声資源活用シンポジウム2017（招待講演）
4. 発表年 2017年

1. 発表者名 有本泰子
2. 発表標題 感情知覚に対するセンシティビティが異なる評価者群による感情評定値の比較
3. 学会等名 日本音響学会2018年春季研究発表会講演論文集
4. 発表年 2018年

1. 発表者名 森大毅, 有本泰子, 永田智洋
2. 発表標題 複数の会話コーパスを対象とした笑い声イベントのアノテーション
3. 学会等名 日本音響学会2017年秋季研究発表会講演論文集
4. 発表年 2017年

1. 発表者名 佐久間結, 堀内靖雄, 有本泰子, 大野澄雄, 黒岩眞吾
2. 発表標題 藤崎モデルのパラメータ推定におけるマイクロプロソディ除去による効果の検討
3. 学会等名 日本音響学会2017年秋季研究発表会講演論文集
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------