

平成 29 年 6 月 28 日現在

機関番号：32619

研究種目：若手研究(B)

研究期間：2013～2016

課題番号：25871139

研究課題名(和文) マン マシン間対話における双方向型マルチモーダル感情コミュニケーションのモデル化

研究課題名(英文) Modeling bidirectional multimodal emotion communication for man-to-machine dialog

研究代表者

有本 泰子 (Arimoto, Yoshiko)

芝浦工業大学・工学部・准教授

研究者番号：60586957

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：本研究は、マン-マシン間の双方向的な感情インタラクションの実現を目指し、話者を対話に没入させ、感情表出の頻度・強度を人間同士の対話に近づけるコンピュータの振る舞いを明らかにすることを旨とした。そのために、まずは人間同士の対話を対象に、対話中の感情表出(表情・音声・自律神経系反応)の同調について分析を行った。その結果、感情経験では同調傾向/非同調傾向を示すことがあるものの、感情表出においては同調傾向しか示さず、感情経験とは無関係に表出行動の模倣がある可能性を示唆した。また、キャラクターエージェントを用いてマン-マシン間対話を実施し、同様の分析を行って、コンピュータに対する同調の傾向を調べた。

研究成果の概要(英文)：This research aims to realize bidirectional emotional interaction between man and machine, to elucidate the behavior of the computer which makes the speaker as expressive as in the dialog with human and which makes them immerse in the dialog with the computer. Using the human-to-human dialog corpus, the synchronization of the emotional expressions (facial expressions, speech, autonomic nervous system responses) between the interlocutors was analyzed. As a result, although both the synchronized / the asynchronized emotional experience were observed between the interlocutors, only the synchronized emotional expression/reaction was observed between them. The result suggested that the mimicry of emotional expression/reaction was independent from the synchronization of emotional experience. The other research was conducted with the man-machine dialog corpus using character agent, and those material were analyzed to examine the synchronization of emotional expression with computer.

研究分野：音声言語情報処理

キーワード：ヒューマンエージェントインタラクション マルチモーダル情報処理 感情 生理反応 表情

1. 研究開始当初の背景

現在の高齢化社会や核家族化社会において、人の感情を理解して共感する介護ロボットや感情コントロール教育を兼ねた子守りシステムなど、人の感情を利用したシステム開発に対する期待が高まっている。現在では、コンピュータによる感情認識や表情・感情音声合成などの一方向的な感情表出の研究は一定の成果をあげているものの、コンピュータとの対話においては、話者は対話に没入せず、また人間同士の対話と比較すると話者の十分な感情の喚起や表出も得られないため、マンーマシン間の感情表出および感情状態の変化を双方向的に分析する感情コミュニケーションの研究はあまり進んでいない。一方で、人間同士の対話場面では、話者間で表情や音声、しぐさなどのマルチモーダルな表出行動を意識的・無意識的に模倣しあうことで、話者間で情動が伝播 (emotional contagion) するという報告がある。つまり、人間同士の対話コミュニケーションにおいては、一方向的な感情知覚・表出だけでなく、感情表出の模倣などの双方向的 (インタラクティブ) な感情コミュニケーションが対話を促進し、表現豊かな対話コミュニケーションが実現されているのである。(図1)

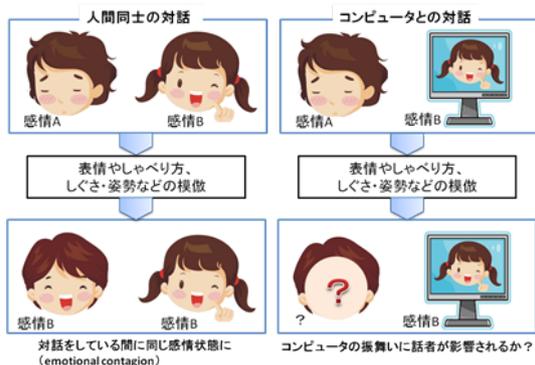


図1 人間同士の対話とコンピュータとの対話における感情インタラクション

本研究では、マンーマシン間の双方向的な感情インタラクションの実現を目指し、話者を対話に没入させ、感情表出の頻度および強度を人間同士の対話に近づけるコンピュータの振る舞いを明らかにする。まずは人間同士の双方向的な感情コミュニケーションのひとつである表出行動の模倣に着目して、話者のマルチモーダルな表出行動をコンピュータに模倣させることでマンーマシン間の双方向的な感情インタラクションを実現し、その効果を検証する。

2. 研究の目的

本研究では、マンーマシン間における話者の表出行動の模倣をコンピュータ側に導入し、対話への没入および感情表出への効果を検証するために、下記の基礎的検討を行う。

- (1) 人間同士の対話における表出行動の模倣と感情状態・表出の変化の分析

コンピュータとの対話における表出行動の模倣の切替えタイミングなどを規則化するために、現在までに収録してきた人間同士の対話の分析を行う。

- (2) (1)の結果をもとにマンーマシン間対話におけるコンピュータの振る舞いの規則化
コンピュータとの対話における表出行動の模倣の切替えタイミングなどを規則化する。本研究で実施するマンーマシン間対話においてはその規則に従って表出行動の模倣を呈示する。
- (3) コンピュータ側の表出行動の模倣が話者の感情状態に与える影響を検証
(2)で規則化した振る舞いを用いて、Wizard-of-OZ (WOZ) 法によるコンピュータとの疑似対話を表出行動の模倣ありとなしの2条件で行い、話者の感情状態の変化および対話への没入感に対する差があるか検証する。
- (4) コンピュータ側の表出行動の模倣が話者の感情表出に与える影響を検証
(3)における対話実験でコンピュータ側の模倣の有無により話者の外的・内的感情表出 (表情・音声・心拍・発汗など) に差があるか検証する。

3. 研究の方法

対象とする対話場面をマンーマシン間のしりとりゲーム場面とする。対話場面をしりとりゲームに限定することで、発話内容にある程度の制約を設けることができ、感情の表出を言語内容ではなく、表情や音声に集約することができる。そのため、マンーマシン間の感情インタラクションを発話内容に依存させず、表情と音声の操作だけでシステム側の感情表出を制御することが可能となる。また、感情の表出は表情や音声などの外的な指標にのみ表れるわけではなく、心拍や発汗など目に見えない内在的な生理反応にも表れる。こうした人間の内在的な表出に対してもシステム側がリアクションを取ることができるよう、対話中は話者の表情・音声・生理反応をリアルタイムでモニタリングする予定である。システム側の模倣の表出には簡便なキャラクターエージェントを使用する予定である。人間の顔を用いた表情合成システムではなく、キャラクターエージェントを使用することで、人の顔が持つ個人特性を排除でき、顔の個人特性が聞き手に与える知覚的な影響も排除可能である。

- (1) 人間同士の対話における表出行動の模倣のタイミングと感情の変化の分析
マンーマシン間のマルチモーダル対話データ収集に向けたコンピュータ側の振る舞いを決めるために、人間同士の対話における表情・音声・生理反応などの表出行動の模倣のタイミングと感情の変化について分析を行う。すでに収録済みの表情・音声・生理反応 (心拍・発汗) で構成されたしりとり対話データを分析対象とし、対話中の対話者間の表

出行動の模倣と感情状態の変化との関係を検証し、感情状態を変化させるシステム側の模倣のタイミングを規則化する。人間同士の対話データに関してはすでに整理済みであり、本研究にそった分析を行える状態となっている。

(2) 対話システムの表情・音声の制御を半自動で行う WOZ 対話システムの開発に着手。話者の表情・音声・生理反応をリアルタイムにモニタリングしながら対話システムの表情・音声の制御を半自動で行う WOZ 対話システムの開発に着手する。対話システムの出力としてキャラクタエージェントを実装する。キャラクタの表情生成については、簡易モーションキャプチャの Microsoft Kinect や音声インタラクションシステム構築ツールキットの「MMDAgent」を用いる予定である。Kinect を利用して取得した対話者の顔の特徴点の動きを参考に、キャラクタの模倣の表情を生成する。また合成音声に関しても、話者の音声の情報を MMDAgent から取得して主に韻律情報を操作した模倣の合成音声を呈示する。対話システムの発話生成に関しては、応答の即時性、しりとりゲームにおける回答の人間らしさの保持、発話内容が複雑化する可能性を考慮し、Text-To-Speech を利用して人手にて生成する。

(3) マンマシ間マルチモーダル対話データの収集

(2)にて開発した WOZ 対話システムを用いてマンマシ間のマルチモーダル対話データの収録を行う。収録場所は理化学研究所内にある防音室を利用する。システム側の振舞いは WOZ 対話システムによって記録を行い、話者の表情・音声・生理反応の収録は現在使用している表情分析用 API と音声キャプチャボードおよび生理反応計測装置を使用する予定である。40 名程度(男女 20 名ずつ)の被験者をシステム側の模倣あり条件となし条件の 2 条件に分けて対話データの収集を行う。収録後には対話に対する没入感に関する質問紙調査および収録した映像データを見ながら対話中の感情の主観評価実験を行わせ、話者の感情状態や対話への没入感に関する主観評価データを取得する。

(4) 表情・音声・生理反応などのマルチモーダル感情情報の分析と双方向型マルチモーダル感情コミュニケーションのモデル化
表出行動の模倣が話者の感情状態および感情表出に与える影響を検証する。感情状態に与える表出の模倣の影響の検証では、システム側の模倣状態と主観評価実験により得られた感情の主観評価値および没入感との関係を明らかにする。感情表出に与える影響の検証ではシステム側の模倣の有無により話者の外的・内的感情表出(表情・音声・心拍・発汗など)との関係を明らかにする。これらの分析により明らかにした知見から、マンマシ間における双方向型マルチモーダル感情コミュニケーションのモデル化

を行う。

4. 研究成果

(1) 人間同士の対話における表出行動の模倣のタイミングと感情の変化の分析

人間同士の対話データはすでに整理済みのマルチモーダルコーパスを使用した。被験者は 26 組 52 名の男女(平均年齢 21 歳、標準偏差 2.34)である。被験者は同性の友人同士で対話収録に参加させた。話者の表情およびしぐさを記録するため、話者の様子を 4 台の CCD カメラで撮影した。音声は 48kHz, 16bit で、映像は 30fps, 640×480pixel で記録した。また、皮膚コンダクタンス水準(SCL)の計測のため左足の人差し指と中指に電極を装着させた。さらに、心電図(HR)を測定するために胸部(左右の第 4 肋間付近と右鎖骨下 10cm 付近)にも電極を装着させた。SCL と HR はともに 1kHz でサンプリングを行っている。映像・音声・生理反応の同期を取るため、撮影開始のシャッタータイミングで映像音声記録装置から生理信号記録装置に TTL 信号を自動送信し、SCL と HR の記録開始のトリガーとした。誘発する情動をある程度コントロールするため、被験者どうしを競わせる競合課題と被験者どうしで協力して課題を進める協調課題を 10 分間行わせた。競合課題ではしりとりを行わせた。その際、ゲームとして相手を負かすよう戦略的にしりとりを進めるように指示した。協調課題では、何らかのルールに従って上がる点数を二人で協力して稼ぐように指示した。被験者にはルールは知らせていないため、点数の上がる様子から二人でルールを推測する必要がある。収録したデータに対し、被験者本人に情動の評価を行わせた。評価には GTrace[1]を用い、映像データを視聴させながら評価対象者の情動を動的に評価させた。評価項目は情動の次元説で用いられる覚醒-睡眠(Arousal)と快-不快(Valence)、および話者の態度に近い肯定的-否定的(Positivity)の 3 次元を評価させた。評価対象者は話者自身(自己評価)とその対話者(他者評価)である。本研究では自己評価のデータのみを用いる。情動の評価値は対話の開始から 10 秒間隔で平均値を求めている。先行研究[2-7]を参考に、計測した表出行動および生理反応から各特徴量を求めた。生理指標も情動の評価値と同様に 10 秒間隔で各特徴量を求めた。音声は 200ms を基準とした Inter-pausal unit(IPU)を発話単位とし、発話ごとに特徴量を抽出したのち、10 秒間隔で統計量を求め直した。表情の特徴量に利用する表情の特徴点は、FaceSDK を利用してフレームごとに 14 点を抽出し、アフィン変換をかけて顔の回転および大きさを被験者間で補正した。表情の個人差の影響を除去するため、補正後の各特徴点から標準顔のフレームの特徴点との差分を求め、表情の特徴量とした。以後の分析では、対話の開始および終了 1 分間を除去した 8 分間のデータを用

いて分析を行う。

ペア内の情動の同調傾向を調べるため、自己評価値の相関係数を対話ごとに求めた。相関係数が正の相関を示せばその情動は対話を通して同調していたことになる。逆に負の相関を示せば非同調していたことになる。求めた相関係数に対して、3 情動状態×2 性別×2 課題による 3 元配置分散分析(ANOVA)を行った。その結果、情動状態×課題、性別×課題について交互作用が認められた(それぞれ、 $F(2, 144) = 5.32, p < 0.01, F(1, 144) = 4.81, p < 0.05$)。相関検定の結果、有意差 ($p < 0.05$) を示したペアの平均相関係数とそのペア数を相関の傾向ごとに、図 2 に示す。

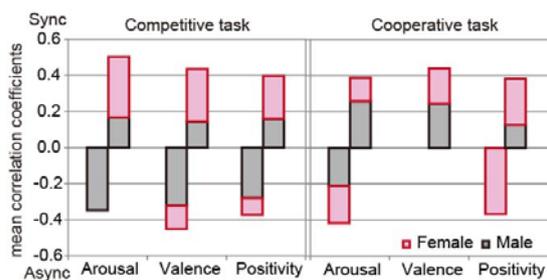


図 2 相関検定の結果

各課題において同調傾向を示すペアと非同調傾向を示すペアが存在することが分かる。Valence で負の相関を示すペアが存在したのは競合課題だけであった。競合課題では話者間で競わせたため、快-不快を示す Valence が非同調傾向となるペアが存在したことが示唆される。一方で、協調課題では Valence が非同調傾向となるペアが存在しなかった。協調課題では二人で協力し合って課題を解決していたために、Valence が相反する傾向となることがなかったと考えられる。

各ペアにおける表出行動および生理反応の同期傾向を調べるため、ペア内の各特徴量の相関係数を対話ごとに求めた。相関係数が正の相関を示せばその特徴量は対話を通して同期していたことになる。逆に負の相関を示せば非同調していたことになる。図 3 に有意差 ($p < 0.05$) を示した特徴量の平均相関係数とそのペア数の比率を相関の傾向ごとに示す。簡潔に記すため、平均相関係数とペア数の比率は表出行動および生理反応の種類ごと (SPEECH, FACE, SCL, HR) に示した。課

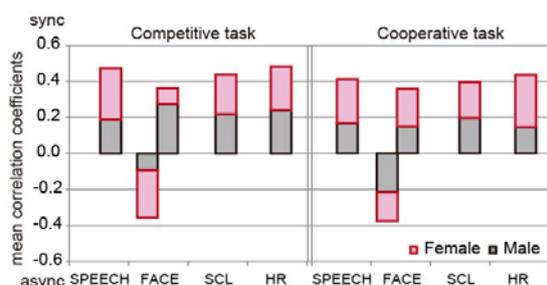


図 3 有意差を示した特徴量の平均相関係数

題の種類にかかわらず、SPEECH・SCL・HR は中程度の正の相関を示すペアが多くみられる一方で、FACE は負の相関を示すペアが

存在した。また、生理反応の特徴量である SCL と HR は、協力的に課題を遂行する協調課題よりも、対話者と競う競合課題の方が有意な正の相関を示すペアが多かった。

情動の同調と各特徴量の同期との関係进行分析するため、ペアごとの自己評価値の相関係数と各特徴量の相関係数との相関検定を行った。自己評価値の相関係数と各特徴量の相関係数との間に強い相関があれば、情動の同調と各特徴量の同期との間の関係を示すことができる。相関検定の結果、競合課題における Arousal の同調傾向と有意な負の相関 ($p < 0.05$) を示したのはふたつ表情の特徴量(右眼, 右内眼角(ともに $r = -0.43$))であった。また、声の高さを表す特徴量は競合課題における Positivity の同調傾向と有意な負の相関 ($r = -0.40$) を示した。協調課題においても 5 個の表情の特徴量(右眼, 右内眼角, 右外眼角, 右下眼瞼, 右上眼瞼)が Arousal の同調傾向と有意な相関を示し(それぞれ, $r = -0.64, -0.61, -0.55, -0.65, -0.57$)、声の強さを表す特徴量と 6 個の表情の特徴量(右眼, 右眉中央, 左内眼角, 右外眼角, 右下眼瞼, 右上眼瞼)が Positivity の同調傾向と有意な相関を示した(それぞれ, $r = 0.45, -0.44, -0.44, -0.40, -0.45, -0.41, -0.50$)。図 3 から、生理反応の特徴量は対話を通してペア間でよく同期していたことが示唆されるが、情動の同調と関連して同期していたという結果は得られなかった。また、競合課題・協調課題ともに Valence の同調傾向と関連して同期している特徴量はなかった。

(2) マン-マシン間対話における表出行動の模倣のタイミングと感情の変化の分析に向けた検討

マン-マシン間対話を収録した。被験者は大学生および大学院生 28 名(女性 13 名, 男性 15 名, 平均年齢 21.05 歳, 標準偏差 1.81)である。簡易な防音環境にある実験用ディスプレイの前に被験者を着座させ、ディスプレイに映されるキャラクターエージェントと共に課題を行わせた。人間同士の対話と同様の課題とするため、マン-マシン間対話の課題でもしりとりを行わせた。課題は、キャラクターエージェントの身体表現有りの試行と身体表現無し試行の 2 回実施した。課題は 1 試行 10 分間行った。被験者には、しりとりでキャラクターエージェントを負かすよう戦略的にしりとりを進めるように指示した。

キャラクターエージェントには MMDAgent を使用した。MMDAgent と課題を遂行させるため、対話は WOZ 方式で行った。課題であるしりとりを円滑に行わせるため、ある程度のキーボード入力を省略し、速やかにキャラクターエージェントのターンを構成するため、しりとりで特化した WOZ システムの構築を行った。本 WOZ システムは MMDAgent の Plug-in としてサーバプログラムを実験用 PC で動作させ、クライアントプログラムがインストールされた PC からネットワークを経由

して、キャラクタエージェントを操作するメッセージを送信する。本システムは、次の発話だけでなく、身体表現のメッセージも任意のタイミングで送信することが可能である。MMDAgent 付属の身体表現のモデルを実験の刺激として利用した。付属の身体表現には、表現のラベルが付いている。そのラベルと実際のエージェントの表現を知覚した被験者の評価が一致するか確認するため、実験前と実験後に表現のみの評価を行わせた。話者の表情およびしぐさを記録するため、話者の様子をビデオカメラ (Panasonic HD ビデオカメラ V360M) で撮影したのち、30fps, 640×480pixel で記録した。キャラクタエージェントの操作画面は実験用 PC (DELL XPS 8900) に搭載した GPU (NVIDIA GeForce GTX 960 2GB DDR5) に付属の画面キャプチャリングソフト (NVIDIA ShadowPlay) を使用した。被験者の音声はヘッドセットマイク (AKG HSC271) を使用して、キャラクタエージェントの音声とともに 48kHz, 16bit で録音した。エージェントの音声を複数の機器に振り分けながら (被験者が装着しているヘッドホンや収録用 PC など) 収録するため、音声キャプチャボード (ROLAND UA-101) を使用した。自律神経系反応の計測には Biopac MP150 system (Biopac Systems, CA, USA) を使用した。皮膚コンダクタンス水準 (SCL) の計測のため、利き手の人差し指と中指に電極を装着させた。さらに、心電図 (HR) は第二誘導で測定した (左右両足の脛付近と右腕に電極を装着)。さらに、被験者の表情への微細な反応を計測するために、利き目側の皺眉筋と大頬骨筋の表情筋活動 (EMG) を計測した。SCL・HR・EMG はともに 1kHz でサンプリングを行っている。映像・音声・生理反応の同期を取るため、被験者の音声を生理反応計測器でも記録した。

<引用文献>

- [1] <https://sites.google.com/site/roddycowie/workresources>
- [2] C. Lee, A. Katsamanis, M. P. Black, B. R. Baucom, P. G. Georgiou, and S. S. Narayanan, "An Analysis of PCA-based Vocal Entrainment Measures in Married Couples' Affective Spoken Interactions," in *Interspeech2011*, pp. 3101-3104, 2011.
- [3] 小森政嗣, 福井正昇, 長岡千賀, "対面場面における表情の同調的表出に関する形態測定学的検討," 対人社会心理学研究, vol. 11, pp. 73-79, 2011.
- [4] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2067-83, Dec. 2008.
- [5] S. Koelstra et al., "DEAP : A Database for Emotion Analysis Using Physiological

Signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18-31, 2012.

- [6] M. F. Valstar and M. Pantic, "Fully automatic recognition of the temporal phases of facial actions.," *IEEE Trans. Syst. Man. Cybern. B. Cybern.*, vol. 42, no. 1, pp. 28-43, Feb. 2012.
5. 主な発表論文等
〔雑誌論文〕(計2件)
- [7] Y. Arimoto and K. Okanoya, "Multimodal features for automatic emotion estimation during face-to-face conversation," *Journal of Phonetic Society of Japan*, vol. 19, no. 1, pp. 53-67, 2015.
- [8] Y. Arimoto and K. Okanoya, "Mutual emotional understanding in a face-to-face communication environment: How speakers understand and react to listeners' emotion in a game task dialog," *Acoustical Science and Technology*, vol. 36, no. 4, pp. 370-373, 2015.

〔学会発表〕(計11件)

- [1] Y. Arimoto and K. Okanoya, "Comparison of Emotional Understanding in Modality-controlled Environments using Multimodal Online Emotional Communication Corpus," in *Proceedings of The International Conference on Language Resources and Evaluation (LREC2016)*, pp. 2162-2167, 2016.
- [2] H. Mori and Y. Arimoto, "Accuracy of Automatic Cross-Corpus Emotion Labeling for Conversational Speech Corpus Commonization," in *Proceedings of The International Conference on Language Resources and Evaluation (LREC2016)*, pp. 4019-4023, 2016.
- [3] Y. Arimoto and K. Okanoya, "Emotional synchrony and covariation of behavioral/physiological reactions between interlocutors," in *Proceedings of the 17th Oriental COCOSDA (International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques) Conference 2014*, pp. 100-105, 2014.
- [4] 横山雅季, 有本泰子, 森大毅, "対話音声の感情認識における音響的特徴と言語的特徴の効果," 日本音響学会 2017 年春季研究発表会講演論文集, pp. 341-342, 2017.
- [5] 有本泰子, 森大毅, "クロスコーパスラベリングによる感情空間への感情カテゴリのマッピング -音声知覚および音響分析による検討-, " *HCG シンポジウム 2016 論文集*, pp. 456-461, 2016.
- [6] 有本泰子, 森大毅, "音声からの知覚による感情カテゴリと感情次元の関係," 日

- 本音響学会 2016 年秋季研究発表会講演
論文集, pp. 219-222, 2016.
- [7] 有本泰子, 森大毅, "クロスコーパス感情
ラベリングによる対話音声の比較," 日
本音響学会 2016 年春季研究発表会講演
論文集, pp.359-360, 2016.
- [8] 【招待講演】有本泰子, "コミュニケーシ
ョン場面におけるリアルな感情表出の
分析," 日本音響学会 2015 年秋季研究発
表会講演論文集, pp. 1317-1320, 2015.
- [9] 有本泰子, 江田康太郎, 菊池英明, 岡ノ
谷一夫, "オンラインコミュニケーション
におけるモダリティ統制下の情動理
解," 日本音響学会 2014 年秋季研究発表
会講演論文集, pp.385-386, 2014.
- [10] 【招待講演】有本泰子, "対話場面におけ
るマルチモーダル感情コミュニケーション," ヒューマンストレス産業技術研
究会第 27 回講演会, 2014.
- [11] 有本泰子, 河津宏美, "音声チャットを利用
したオンライゲーム感情コーパス"
日本音響学会 2013 年秋季研究発表会講
演論文集, pp.385-388, 2013.

[その他]

ホームページ等

<https://sites.google.com/site/ogcorpus/>

6. 研究組織

(1)研究代表者

有本 泰子 (ARIMOTO Yoshiko)

芝浦工業大学工学部・准教授

研究者番号：60586957