OS-35

社会的信号処理と AI

Social Signal Processing and AI

岡田　将吾　北陸先端科学技術大学院大学
Shogo Okada　Japan Advanced Institute of Science and Technology.
okada-s@jaist.ac.jp

石井　亮　NTT メディアインテリジェンス研究所
Ryo Ishii　NTT Media Intelligence Laboratory.
ishii.ryo@lab.ntt.co.jp

Keywords: social signal processing, multimodal interaction, machine learning, pattern recognition.

1. は　じ　め　に

社会的信号処理（Social Signal Processing）とは，言語・音声・視線・姿勢・ジェスチャ・生体情報などの複数のチャネルより得られる情報を統合し，人間の情動・態度・個性・スキル・リーダシップや，人間同士のコミュニケーションのメカニズムといった，人間が行動・コミュニケーションを通じて形成する社会性の側面を理解・計算するための技術である [Burgoon 17, Vinciarelli 09]．この技術は，対話システム，マルチメディア解析，コミュニケーション支援，インタラクティブシステムを構築するうえでの基盤となるため，近年，大きな注目を集めている．社会的信号処理研究を推進するには，言語・音声・画像・生体信号処理や，機械学習・データマイニングといった，人工知能に関連の深い技術に加え，人間の行動・コミュニケーションの理論を構築する社会学・言語学・心理学などの知見が重要となり，学際的な取組みが必要である．社会的信号処理に関する研究は，近年，複数のトップレベルの国際会議で発表されている*1．一方，日本国内においては，社会的信号処理に関して発表する場が少なく，そのため研究者間の交流を促進するためのコミュニティも十分に形成されていない．このような状況を踏まえ，社会的信号処理の研究に携わる国内の研究者を一堂に集めて，研究発表・議論を行う場を構築することを目的として，我々は 2017 年度人工知能学会全国大会（第 31 回）においてオーガナイズドセッション（OS-35）：「社会的信号処理と AI」を開催した．今回初めての開催にもかかわらず，11 件の多様かつ質の高い研究発表が集まり，当日の参加者は約 70 名と盛況のうちに終えた．本解説では，社会的信号処理の研究概要と本 OS で発表された研究内容を紹介する．

2. 社会的信号処理の基礎

社会的信号処理は社会心理学・社会言語学の知見（例えば [Knapp 09]）をもとに発展した．[Vinciarelli 09] によると，社会的信号処理で対象となる対人コミュニケーションチャネルは下記のように分類されている．

（1）発話言語
（2）音響，音声的特徴（韻律など）
（3）体形・身長，服装などを含む容姿
（4）ジェスチャや姿勢
（5）表情，視線方向，注視状態
（6）対人距離，座席の位置関係，コミュニケーション環境

また，社会的信号処理で扱われる変数は，例えば以下のように分類できる．

（a）感情（Emotion）
（b）個性，スキル（Personality, Skill）
（c）社会的な地位，役割（Status, Role）
（d）優位性（Dominance）
（e）説得性（Persuasion）
（f）調和的，親密な関係，態度（Rapport, Attitude）
（g）その他の内面状態（Others）
（h）コミュニケーションにおける調整（Regulation）

近年，カメラ，マイク，モーションセンサが安価になると共に，音声・画像処理，計測工学が発展し，発話言語や音声特徴の認識（(1), (2)），ジェスチャ，姿勢，視線，表情，対人距離といった視覚情報（(3) 〜 (6)）の処理基盤が整備され，多様な言語・非言語情報を計算機モデルの入力データ（X）として扱えるようになった．コミュニケーション中に表出する言語・非言語情報（X）を入力として，上記（a）〜（h）に関連する目的変数（Y）を推定する機械学習の問題として定式化することが，社

*1 ACM ICMI, ACM Multimedia, IEEE FG, ACII, IEEE ICME, ACM IUI, Interspeech, IEEE ICASSP, ACM Ubicomp.

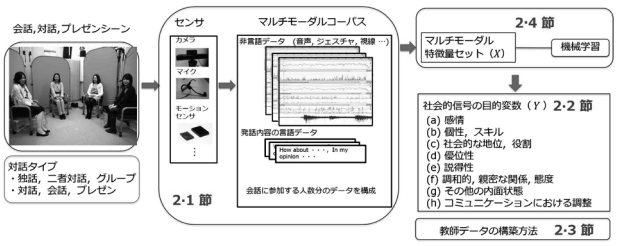

図 1 基本的な社会的信号処理の流れ

会的信号処理の代表的な取組みである．社会的信号処理で観測される複数種類の言語・非言語情報は多変量時系列データとして取得され，さらにコミュニケーションに参加する人数分の多変量時系列データを対象として，社会的信号の目的変数を推定するために適したデータ構造，特徴表現を探索する必要がある．そのため，社会的信号処理においてデータマイニングの技術も必要不可欠である．このように，社会的信号処理は，センシング，信号処理，機械学習，パターン認識，データマイニング技術を統合的に扱い，これまで社会学で得られてきた，社会的側面・人間の特性と言語・非言語情報の関係性をモデル化する技術である．

次節から，社会的信号処理の基本的な流れを紹介する．2·1 節では，社会的信号処理の入力データを得るためのセンシング手段とマルチモーダルコーパスの構築方法，2·2 節では，社会的信号処理の目的変数に関して述べる．2·3 節では機械学習のための教師付きデータの作成方法，2·4 節では一般的な社会的信号処理のための機械学習手法について説明をする．

2·1 センシングとマルチモーダルコーパス

社会的信号処理の入力対象となる言語・非言語情報のセンシングには多様な計測技術が必要となる．(1) 発話言語とその (2) 音声特徴は，指向性の携帯型マイクやマイクロフォンアレーから得られる音声情報から抽出される．マイクロフォンアレーを利用した場合は，会話者個人の音声情報を取得するために音源定位・分離が別途必要となる．(1) 発話言語は音声認識を通じて自動的にテキスト情報に変換されることが望ましいが，自由会話における発話を完全に自動認識することは難しいことから，書き起こしテキストが研究に用いられることが多い．

ジェスチャ，姿勢，視線，表情，対人距離といった視覚情報 ((3)〜(6)) は，カメラ映像に対して画像処理を行って抽出するか，カラー画像のみの情報から精度良く抽出するのが困難な場合，深度センサやモーションキャプチャ装置を用いたり，装着型の視線計測装置・対人距離計測装置などを用いる．その他，心拍，発汗，呼吸，脳波といった生体信号を抽出する場合は，専用の生体信号計測用センサを用いる必要がある．

計測されたデータは，そのまま信号波形あるいは数値データとして扱われるほか，発話言語や非言語行動については，おのおの，意味のある行動要素を時間軸に沿って区切り，イベントとして記述する．会話における言語·非言語の分析単位は，例えば [坊農 09] に詳しく示されている．

映像，音声および多様なデータを統合的に管理，視聴，アノテーションを行えるツールはデータ作成や分析を行ううえで有用であり，ツールの開発も多くされている．広く利用されているツールとして，例えば，Anvil [Kipp 01]，ELAN [Brugman 09] がある．一連のデータは，時間的に同期した時系列データとして統合され，マルチモーダルコーパスが構築される．

これまで構築された代表的な会話のマルチモーダルコーパスとして，国外では，AMI（Augmented Multi-party Interaction）[Carletta 06]，VACE（Video Analysis and Content Extraction）[Kasturi 09]，CHIL（Computers in Human Interaction Loop）[Mostefa 07] のプロジェクトで大規模コーパスが構築されている．国内では，ATR インタラクションコーパス [角 03]，IMADE コーパス [角 08], NTT コーパス [Otsuka 08] などがあげられる．

2·2 社会的信号の目的変数

(a) 感情（Emotion）の推定に関する研究は音声·画像·マルチモーダルモデリングの分野で盛んに行われてきた [東中 16]．(b) 個性（Personality）に関しては，心理学で提案された BigFive の指標 [Costa 92]（5 種類の性格特性のレベル）を目的変数とした研究が多く行われている [Gatica-Perez 09]．また，関連する変数として，ビジネス·教育·医療·面接·グループディスカッションといったさまざまな場面でのコミュニケーション能力の推定についても盛んに取り組まれている [Nguyen 14，Okada 16b]．(c) 社会的な地位 (Status)，(d) 優位性 (Dominance) に関しては，会話への参加者がどれくらい影響を与えているかといった優位性の推定 [Jayagopi 09] や，リーダシップをもつメンバの推定 [Sanchez-Cortes 12] に関する研究が行われている．(e) 説得性（Persuasion）に関しては，social media におけるユーザの商品レビュービデオの説得力の推定 [Park 14] や，ストーリーテリングの上手さの推定 [Okada 16a] の研究が行われている．(f) 調和的，親密な関係，態度に関しては，会話参加態度（Engagement）の推定 [Ishii 13] が行われている．(g) その他の内面状態に関しては，例えば視線や発話状態から人の興味対象 [Misu15] を推定したり，疲労の状態·飽きている状態を推定したり，(a) 〜 (f) に含まれない内部状態の変数を扱う研究も多く行われている．(h) コミュニケーションにおける調整（Regulation）に関連して，コミュニケーションにおけるさまざまな対象の自動推定が試みられてきた．会話における代表的な推定対象として，参加者が誰に注意を向けているのか，誰が誰に対して発話を行っているのかといった視覚的な注意（Visual attention）の向け先，円滑な会話を行ううえでの適切な話者交替（Turn-taking）のタイミングがある．上記の

会話行動の計算機モデルは，音声対話システムが人間と自然かつ円滑な会話を実現するために重要である．

2·3 教師・評価データセットの構築方法

（a）～（g）は人間の内面状態であり，直接観測することができない．内面状態の推定のために，機械学習のアプローチを適用する際，教師データセットの構築方法が問題となる．現在大きく分けて二つのアプローチが存在する．1 番目に，心理学の知見から構成された質問紙が利用できる場合，実験参加者の同意のもと，質問紙調査を行い，この結果を教師データとする方法がある（BigFive もその一つ）．2 番目に，複数の第三者コーダに対話場面を閲覧させて，アノテーションする方法がある．複数のコーダの評定の一致度を計算し，十分なレベルの同意が得られた場合，複数コーダの評定値平均を教師データとする場合が多い．一方で，二者の「共感度合い」のように，必ずしもコーダ間の評定値が一致しない場合も存在し，コーダの個人差をモデル化する研究も進められている [Kumano 15]．（h）に関連した目的変数のラベルに関して，会話イベントは直接観測可能であるため，教師ラベルを作成（アノテーション）することが可能である．例えば，話者交代を予測するモデルを機械学習で構築する場合，ある時刻 T までの発話者から観測される特徴量を入力として，時刻 $T+1$ 以降に話者交代が起きるか否かのラベルを付与することで，教師付き学習の枠組みに適用できる．

2·4 マルチモーダル特徴量と機械学習

社会的信号処理のためには，一般的なパターン認識のための機械学習アルゴリズムを適用することが可能であるが，会話中に表出する非言語行動は無意味な動作も含むため，目的変数の推定のために観測された時系列データを直接入力してもうまく学習できないことが多い．そのため，タスクに適した特徴量の抽出が重要となる．

多くの研究では，社会心理学の知見をもとに，音声・言語・画像処理分野で提案されている基本的な特徴量が組み合わせて用いられている．発話言語データからは，形態素解析などの自然言語処理技術を通じて，語彙，品詞，LIWC（Linguistic Inquiry and Word Count）などの特徴量（素性）が抽出される．議論の流れを分析する場合には，別途アノテーションした談話行為タグも特徴量として使われる．各参加者の音声データから発話区間検出（VAD）を通じて取得した発話区間情報より，その話者の発話時間，発話回数，他者の発話とオーバラップして発言した回数などの特徴量が抽出される．また韻律特徴（パワー，ピッチなど）が音声処理技術により抽出される．画像，各種センサから取得されるジェスチャ，姿勢，視線情報からは，姿勢変化量，手の動作量，頭部動作量，視線変化パターンなど目的に応じた特徴量が抽出される．また表情認識の結果から出力された感情ラベル・強度を特徴量とする場合もある．

個性やスキルの推定タスクでは，多くの場合，会話参加者単位にラベルが付与される．個性やスキルの指標の大・小など離散的なレベルを推定する際には分類モデルが，連続値を推定する際には回帰モデルが用いられる．

会話場面のイベント（話者交代，次話者，相づち）を推定する場合，イベントが生起した時間内に観測された系列データと出力ラベルが対応し，シーケンスラベリングの問題として扱えるため，HMM（Hidden Markov Model），CRF（Conditional Random Fields）などのモデルを利用できるほか，時系列の特性を一つの特徴量として扱うことで，SVM（Support Vector Machine），Random Forest などの分類モデルも利用可能である．会話場面のイベント推定には，社会言語学で得られている知見を利用して特徴量を設計することが有用であることが知られている．例えば，話者交代の際には，視線方向の遷移パターンが重要であることが古くから知られていたが [Kendon 67]，これを利用することで次話者の予測精度を向上できることが報告されている [Ishii 16]．

複数の非言語情報から定義されるマルチモーダル情報を用いた社会的信号処理では，異種のモダリティー（言語，音声，画像など）から得られるデータの統合方法が重要となる．従来研究では，複数モダリティーで得られる特徴量を一つのベクトルに統合し，機械学習への入力として用いる方法や，各モダリティーの特徴量セットごとに識別器を用意・訓練し，識別時には各モダリティーに対応するモデルからの出力値を統合する方法が用いられてきた．近年，深層学習アルゴリズムの発展により，各モダリティーごとに DNN（Deep Neural Network）や CNN（Convolutional Neural Network）を用意し，さらに各ネットワークのアウトプットを統合するネットワークを用意して訓練する手法も提案されており [Ngiam 11]，社会的信号処理における応用も提案されている [Nojavanasghari 16]．

社会学の知見や，コミュニケーションの構造を加味したうえで，ニューラルネットワークの構造を規定し，社会的信号処理に特化した深層学習の枠組みを開発することは今後の重要な課題の一つである．また実際の会話データを収集するためには，実験参加者の募集を含め，多くのコストがかかるため容易に大規模データを収集することは難しい．このため，収集済みの異種の会話データを利用したマルチタスク学習，転移学習，半教師付き学習の方法論についても今後適用可能性を探求する必要がある．

2·5 社会的信号処理の応用

社会的信号処理により得られた情報を用いた，人のコミュニケーションスキル向上支援，システムとの対話に基づくユーザのコミュニケーション支援などに応用されている．例えば，人のコミュニケーションスキル向上支援として，プレゼンターの説得力やコミュニケーションス

キルの高低を可視化したり，聴衆エージェントの振舞いをコントロールしてリアルタイムにフィードバックを行うシステムが考案されている [Chollet 15]．また，マルチモーダルデータから議論における重要シーンを認識・検出し，会議映像を要約する試みも行われている [二瓶 17]．

社会的信号の認識・理解・生成技術を統合することで，より人間らしい対話が実現できる可能性があるため，対話システムにとって社会的信号処理は重要な技術である．これまで，ユーザの感情 [東中 16]，エンゲージメント [Ishii 13] の推定結果に応じてシステムの振舞いを制御することにも適用されている．また，多人数会話を対象とした研究においては，円滑な話者交替の実現 [Bohus 10] や，受話者推定の結果を踏まえて振舞いを変化させる機構 [中野 14] も検討されている．

3. 発表論文概要

本章では本OSで発表された11件の研究を紹介するとともに，各研究が2章で述べた社会的信号処理の基礎のどの部分に位置付けられるかを述べる．紹介に用いる語句は各研究の原稿に従った．

(1) 視線情報からの未知語検出における個人適応と単語親密度の影響の調査

非母語者がチャットやメールを利用するうえで，円滑な情報伝達を阻害する要因として未知語が考えられる．本研究では，読み手の注視特徴から自動で未知語を検出し，注釈を付与することを目的とした．(1) 個人に適応した識別器の作成，(2) 単語難易度や単語親密度の使用により，評価実験で有意に検出精度が向上することを確認した［未知語を見た場合の認知状態（(h) その他の内面状態）を反映する視線情報を利用した未知語検出］．

(2) 自閉スペクトラム症者支援に向けた自動ソーシャルスキルトレーニング手法

著書らは対話中の言語・非言語情報を自動的に抽出し，対面コミュニケーションスキル向上のためのアドバイスを生成・可視化する機能を有する対話エージェントシステムを実装している．本研究では，システムを自閉症者の訓練に利用し，スキル評定値は訓練前後で改善することを示した［(b) コミュニケーションスキルの評定値の推定機能とアドバイス機能を備えた，コミュニケーション支援システムの構築と評価］．

(3) インタビュー対話における重要シーン推定のための言語・非言語特徴量

インタビュー対話ロボットと人のインタビュー履歴を要約することを目標とし，重要な回答シーンを人のマルチモーダルデータから推定する方法を検討した．インタビュー回答時の音声データ，回答の様子を捉えた深度センサデータを入力として韻律・姿勢・ジェスチャなどの非言語情報を取得したほか，音声認識に基づいて取得した回答時の発話内容の言語特徴を用いて推定を行った［インタビュー要約のための発話時の (f) 態度の推定］．

(4) 深層学習を用いた会話中の人物頭部ジェスチャ認識

対面会話における頭部ジェスチャの自動認識のために，頭部姿勢の時系列信号の特徴抽出と教師あり機械学習を組み合わせた方法が提案されているが，会話中に自発的に生じる頭部ジェスチャはその強度や周期性，個人性など多様であり，高精度な自動認識は依然として困難であった．本研究では，画像分野での認識タスクにおいて着目されているCNNを用いた時系列認識手法を，会話中の頭部ジェスチャ認識に適用してその有効性を示した［センサ情報を用いた頭部ジェスチャの推定は2·1節に関連］．

(5) 保育の質の定量化のための人間行動センシングと解析ツールの開発

近年，待機児童問題や保育士の離職率の高さの社会問題が深刻化しているため，著者らは保育業務を支援可能なAI技術の開発を通して保育士の負担軽減を目指している．これまで，熟練保育士による子供の「関心」のアノテーションにより，子供の個性を推定したが，長時間の作業を必要とした．本研究では，アノテーションをAI技術により半自動化するツールの設計とプロトタイプシステムを構築した［(h) 感心推定，アノテーションツールの提案は2·1節に関連］．

(6) ユーザの態度推定に基づき適応的なインタビューを行うロボット対話システムの開発

(3) の発表に関連して，インタビュー対話ロボットの開発に関する発表であった．本研究では，インタビュー時のユーザの回答意欲に応じて，インタビュー戦略を適応させることを目指し，簡易なインタビュー対話機能と，対話中のインタビュー対象者の表情・動作・音声情報をリアルタイムに取得するためのマルチモーダルセンシング環境をもつインタビューロボットシステムを開発した．実験参加者とのインタビュー対話実験を実施し，コーダにより回答意欲があると判断される箇所を，韻律・姿勢情報から約70％で自動推定できることを示した［(f) 態度推定に基づく対話ロボットへの応用］．

(7) グループディスカッション参加者の役割に基づいた会話状況とコミュニケーション能力の分析

グループディスカッション時の参加者のコミュニケーションスキル（能力）を推定し，能力向上を支援するシステムの開発に向けて，参加者の会話における役割を推定する方法を提案した．参加者の役割および役割により規定される会話状況を定義し，参加者のコミュニケーション能力の印象評価との関係を分析した．その結果，コミュニケーション能力が高く評価された参加者は，低く評価された参加者と比べて積極的にディスカッションに参加していたことが示唆された

［(b) コミュニケーション能力と (c) 役割の推定に関連］.

(8) 評定者個人に特化した他者感情理解モデル

個人が他者の感情状態をどう理解・評定するかを予測するための数理モデルを提案した．評定者の評定傾向を説明する評定者モデルと，ある行動が不特定の評定者からどう評定されそうかを説明する行動モデルの統合により，評定者の属性・特性と対象場面の社会的信号からその評定者が与える評定値を予測する．独自に収集した共感認知データを用いて提案法の有効性を確認した［(a) 二者間の感情状態の推定に関連］.

(9) 障害物を含むオフィス空間でのインタラクション対象の推定

近年，複数人物間の対面インタラクションの参与者グループを認識するさまざまな手法が提案されている．これまで著者らは，F 陣形という円形の身体配置をレベルセット法で抽出する手法を提案してきた．これらの手法は，主に障害物のない自由空間を対象としてきたが，実際のオフィス空間ではパーティションなどの障害物が存在する．そこで，このような空間を対象に，障害物を含む空間における人物の対面インタラクションの対象の推定法を提案した［インタラクション対象の推定は 2·1 節に関連］.

(10) 議論中の言語・非言語情報に基づく発散的・収束的発話の識別

グループディスカッションをファシリテートするシステムの実現を目指し，各発話が，議論を発散させるものであるのか，収束させようとするものであるのかを識別するモデルを提案した．モデルの特徴として，発話中の新規名詞数や既存名詞数などの言語情報，韻律・発話長などの非言語情報を用いた．発散・収束・その他の 3 値に分類するモデルを決定木学習により作成した結果，分類精度は約 57 %であった［議論の発散·収束は（h）コミュニケーションの調整に関連］.

(11) 映画からのマルチモーダル対話コーパスの作成

本研究はライセンスフリーの大規模な映画リソースを用いて，さまざまなシチュエーションで起こる会話·対話シーンを抽出し，マルチモーダル対話コーパスを構築することを試みた．音声情報を用いて，対話シーンのスポッティングを行い，対話シーンを収集した．本研究は，社会的信号処理研究のためのデータセット構築に焦点を当てた［マルチモーダルコーパスの自動構築（2·1 節）に関連］.

4. 展望と今後の課題

近年，大規模データの収集とそのデータを用いた深層学習の適用が，音声，画像，言語などの各種メディア処理に適用され，その有効性が顕著に示されている．社会的信号処理のためのマルチモーダル情報を大規模かつ高品質に収集することは容易ではなく，現状，比較的小規模なデータセットを扱うことが多い．より大規模かつ高品質なコーパスを構築するスキームづくりが課題の一つとしてあげられる．また社会的信号を構成するマルチモーダル特徴量は，対話・会話の種別，タスク，対象者のジェンダ・年齢・国籍などの個人差に依存して異なる．現状，個々の研究者·グループが，特定の個人属性に偏ったデータを収集し，研究を進めている段階である．個別に研究された成果の知見を統合することや，多様な個人属性をもつ大規模データの構築も必要である．

マルチモーダル処理の特性に関して理論面の研究を強化することも今後の課題である．多くのノイズを含むマルチモーダルインタラクションデータから本質的な社会的信号を抽出するための方法論を構築する必要がある．

また，基盤となる視聴覚情報処理，コミュニケーション科学，社会学の理論との接点を再考し，分野を横断した研究コミュニティを形成することも重要である．

また，人間のさまざまな社会的振舞いを理解したうえで，どのようなアプリケーションへの応用を行い，問題を解決できるかはこれからの大きな課題である．また，アプリケーション応用のためには，社会的信号処理をリアルタイムかつ安定的に行えるシステム開発も重要である．

研究の性質上，顔画像を含む個人データを扱ううえに，個人の内面状態を推定する技術に焦点を当てているため，実験段階でも倫理的な配慮が欠かせない．本技術をシステム実装する際には注意が必要であり，対象者のプライバシーを保護する方法や，社会実装する上で問題となる倫理面の議論が必要不可欠である．

5. ま と め

社会的信号処理研究の基礎的な流れについての概要と，JSAI 2017 の OS「社会的信号処理と AI」で発表された研究の紹介を行った．誌面の都合上，本記事で紹介した文献はごく一部のものであり，すべての研究・応用システムを網羅できていないことにはご容赦いただきたい．社会的信号処理は研究対象が多岐にわたり，またさまざまな研究分野を横断した萌芽的なテーマである．今後，多様な専門性をもつ研究者が社会的信号処理研究に参画し，ますます研究分野が盛り上がっていくことを期待している．本記事を通じて，一人でも多くの方が本研究分野に興味をもっていただければうれしい限りである．また，来年度 JSAI 2018 においても，本オーガナイズドセッションの企画を予定しており，多くの読者の皆様にご投稿，ご参加をいただけることを期待したい．

謝　辞

OS「社会的信号処理と AI」の発表者，参加者の皆様，JSAI 2017 の運営および本誌の編集にご尽力いただいた皆様に，感謝の意を表します．

◇ 参 考 文 献 ◇

[Bohus 10] Bohus, D. and Horvitz, E.: Facilitating multiparty dialog with gaze, gesture, and speech, *Proc of ACM ICMI-MLMI,* pp.5:1-5:8（2010）

[坊農 09] 坊農真弓，高梨克也：知の科学：多人数インタラクションの分析手法，オーム社（2009）

[Brugman 09] Brugman, H. and Russel, A.: *Annotating Multimedia /Multi-modal Resources with ELAN*（2009）

[Burgoon 17] Burgoon, J. K., Magnenat-Thalmann, N., Pantic, M. and Vinciarelli, A.: *Social Signal Processing,* Cambridge University Press（2017）

[Carletta 06] Carletta, J., Ashby, S. and Bourban, S., et al.: *The AMI Meeting Corpus: A Pre-announcement,* Springer Berlin Heidelberg（2006）

[Chollet 15] Chollet, M., Wörtwein, T. and Morency, L.-P., et al.: Exploring feedback strategies to improve public speaking: An interactive virtual audience framework, *Proc. ACM Ubicomp, UbiComp'15,* pp. 1143-1154, New York, NY, USA, ACM（2015）

[Costa 92] Costa, P. T. and MacCrae, R. R.: *Neo Personality Inventory-Revised*（*NEO PI-R*）, Psychological Assessment Resources Odessa, FL（1992）

[Gatica-Perez 09] Gatica-Perez, D.: Automatic nonverbal analysis of social interaction in small groups: A review, *Image Vision Computing,* Vol. 27, No. 12, pp. 1775-1787（2009）

[東中 16] 東中竜一郎，岡田将吾，藤江真也 ほか：対話システムと感情，人工知能，Vol. 31, No. 5, pp. 664-670（2016）

[Ishii 13] Ishii, R., Nakano, Y. I. and Nishida, T.: Gaze awareness in conversational agents: Estimating a user's conversational engagement from eye gaze, *ACM Trans. on Interactive Intelligent Systems,* Vol. 3, No. 2, pp. 11:1-11:25（2013）

[Ishii 16] Ishii, R., Otsuka, K.and Kumano, S., et al.: Predicting of who will be the next speaker and when using gaze behavior in multiparty meetings, *ACM Trans. on Interactive Intelligent Systems,* Vol. 6, No. 1, p. 4（2016）

[Jayagopi 09] Jayagopi, D., Hung, H. and Yeo, C., et al.: Modeling dominance in group conversations using nonverbal activity cues, *IEEE Trans. on Audio, Speech, and Language Processing,* Vol. 17, No. 3, pp. 501-513（2009）

[Kasturi 09] Kasturi, R., Goldgof, D. and Soundararajan, P., et al.: Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol, *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 31, No. 2, pp. 319-336（2009）

[Kendon 67] Kendon, A.: Some functions of gaze direction in social interaction, *Acta Psychologica,* Vol. 26, pp. 22-63（1967）

[Kipp 01] Kipp, M.: ANVIL - A generic annotation tool for multi-modal dialogue, *Proc. 7th European Conf. on Speech Communication and Technology,* pp. 1367-1370（2001）

[Knapp 09] Knapp, M. and Hall, J.: *Nonverbal Communication in Human Interaction,* Cengage Learning（2009）

[Kumano 15] Kumano, S., Otsuka, K. and Mikami, D., et al.: Analyzing interpersonal empathy via collective impressions, *IEEE Trans. on Affective Computing,* Vol. 6, No. 4, pp. 324-336（2015）

[Misu 15] Misu, T.: Visual saliency and crowdsourcing-based priors for an in-car situated dialog system, *Proc. ACM ICMI,* pp. 75-82（2015）

[Mostefa 07] Mostefa, D., Moreau, N. and Choukri, K., et al.: The CHIL audiovisual corpus for lecture and meeting analysis inside smart rooms, *Proc. LREC,* Vol. 41, No. 3, pp. 389-407（2007）

[中野 14] 中野有紀子，馬場直哉，黄 宏軒 ほか：非言語情報に基づく受話者推定機構を用いた多人数会話システム，人工知能学会論文誌，Vol. 29, No. 1, pp. 69-79（2014）

[Ngiam 11] Ngiam, J., Khosla, A. and Kim, M., et al.: Multimodal deep learning, *Proc. ICML,* pp. 689-696（2011）

[Nguyen 14] Nguyen, L., Frauendorfer, D., Mast, M. and Gatica-Perez, D.: Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior, *IEEE Trans. on Multimedia,* Vol. 16, No. 4, pp. 1018-1031（2014）

[二瓶 17] 二瓶芙巳雄，高瀬 裕，中野有紀子：非言語情報に基づくグループ議論における重要発言の推定―グループ議論の要約生成に向けて―, 信学論（A）, Vol. J100-A, No. 1, pp. 34-44（2017）

[Nojavanasghari 16] Nojavanasghari, B., Gopinath, D. and Koushik, J., et al.: Deep multimodal fusion for persuasiveness prediction, *Proc. ACM ICMI,* pp. 284-288（2016）

[Okada 16a] Okada, S., Hang, M. and Nitta, K.: Predicting performance of collaborative storytelling using multimodal analysis, *IEICE Trans.,* Vol. 99-D, No. 6, pp. 1462-1473（2016）

[Okada 16b] Okada, S., Ohtake, Y. and Nakano, Y. I., et al.: Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets, *Proc. ACM ICMI,* pp. 169-176（2016）

[Otsuka 08] Otsuka, K., Araki, S. and Ishizuka, K., et al.: A realtime multimodal system for analyzing group meetings by combining face pose tracking and speaker diarization, *Proc. ACM ICMI,* pp. 257-264（2008）

[Park 14] Park, S., Shim, H. S. and Chatterjee, M., et al.: Computational analysis of persuasiveness in social multimedia: A novel dataset and multimodal prediction approach, *Proc. ACM ICMI,* pp. 50-57（2014）

[Sanchez-Cortes 12] Sanchez-Cortes, D., Aran, O., Mast, M. S. and Gatica-Perez, D.: A nonverbal behavior approach to identify emergent leaders in small groups, *IEEE Trans. on Multimedia,* Vol. 14, No. 3, pp. 816-832（2012）

[角 03] 角 康之，伊藤禎宣，松口哲也 ほか：協調的なインタラクションの記録と解釈, 情処学論, Vol. 44, No. 11, pp. 2628-2637（2003）

[角 08] 角 康之，西田豊明，坊農真弓 ほか：情報爆発時代におけるわくわくするITの創出を目指して：パートⅡ：情報分野研究者のためのオンリーワン共有イノベーションプラットフォーム：5.IMADE：会話の構造理解とコンテンツ化のための実世界インタラクション研究基盤，情報処理，Vol. 49, No. 8, pp. 945-949（2008）

[Vinciarelli 09] Vinciarelli, A., Pantic, M. and Bourlard, H.: Social signal processing: Survey of an emerging domain, *Image and Vision Computing,* Vol. 27, No. 12, pp. 1743-1759（2009）

著 者 紹 介

岡田 将吾（正会員）

2008年東京工業大学大学院知能システム科学専攻博士課程修了．同年，京都大学大学院情報学研究科知能情報専攻特定助教，2011～16年まで東京工業大学大学院助教．2014年IDIAP Research Institute滞在研究員．2017年より北陸先端科学技術大学院大学情報科学系准教授．本学会創立30周年記念論文賞（最優秀論文），ほか受賞．ACM，電子情報通信学会各会員．博士（工学）．

石井 亮（正会員）

2008年東京農工大学大学院工学府情報工学専攻修士課程修了．同年，日本電信電話株式会社入社．現在，NTTメディアインテリジェンス研究所研究主任．2013年京都大学大学院情報学研究科博士後期課程修了．2011～13年成蹊大学客員研究員．ACM 15th Int. Conf. Multimodal Interaction（ICMI 2014）Outstanding Paper Awardほか受賞．電子情報通信学会会員．博士（情報学）．