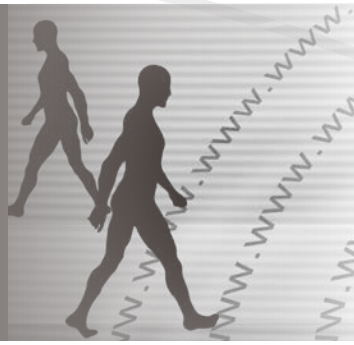
6 社会知の実現に向けたインタラクション理解



角　康之（公立はこだて未来大学）

ヒトと共存する社会知の実現に向けて

ヒトは社会的な生き物である．他人と力を合わせ，1人では解けない問題を解決する．他人と話すことで知識を伝達し，新たな問題に気づく．ヒトは成長とともに，会話や共同作業を行う上での暗黙のルールを身につける．たとえば，集団の会話の中で自分が何か言いたいと思えば，ほかのメンバの発話を遮ることなく，何らかの手段で注目を獲得し，発話にこぎつける．1人では持てない大きな荷物を他人と協力して運ぶときには，掛け声や視線を使いながらタイミングを調整する．このような社会的な能力を社会知（social intelligence）と呼ぶ[1]．

人工知能研究においては，その初期においては，個体としての記号処理能力，すなわち，探索，推論，言語処理，画像処理，歩行などに力が注がれてきた．しかし，ヒトと協力・共存する人工知能の実現を考えたとき，そこに大きく欠けているのは社会知である．コンピュータや人工物を自分自身の身体拡張の道具として直接操作しているうちは，そこに社会知を求める必要性は感じない．しかし，コンピュータ的なものが我々の生活のあらゆるところに浸透し，インタフェースが透明化し，自律的に動作するようになりつつある今，それらには我々の社会的なパートナーとなり得る能力，すなわち社会知を持ってほしい．

ヒトの社会知をコンピュータが扱えるようにするには，我々が無意識のうちにこなしている社会的インタラクションの「辞書と文法」を外在化し，コンピュータが扱える形に表現する必要がある．自然言語については長い年月をかけて辞書と文法が構築され，機械可読化されたことが，情報検索，機械翻訳，質疑応答システム等の発展を支えている．その一方で，身ぶり手ぶり，視線，表情といった非言語的な情報については，まだ辞書や文法と呼べるものは構築されていない．本稿は，ヒトのインタラクション，特にその非言語的な側面の理解を進めるための方法論を概観する．

なぜインタラクション理解なのか？

ヒトのインタラクションが機械的に理解可能になるとなぜ嬉しいのか．

たとえば，ロボットや会話エージェントが社会知を身につけることで，より適切なタイミングで適切な相手にサービスを提供することが可能になるであろう．現状のロボットは，極端に言えば，ヒトを見てもそれを障害物としか認識していない．したがって，会話グループの有無に関係なく，物理的な隙間があれば，グループの中を通り抜けてしまうかもしれないし，話し手に背を向けて聞き手に対して話しかけてしまうかもしれない．現状で最も完成度の高いデモシステムの1つとしては，Bohus らの受付エージェント[2]が挙げられる．インターネットでデモビデオが公開されている☆1ので見ていただきたいが，先に来た2人の来客への対応中に，新しく来た来客にも気を使いながら「ちょっと待ってくださいね」と声をかけるシーンは感動的である．

インタラクションが機械的に理解可能になることで，ライフログやアンビエントサービスの自動化も可能になるだろう．偶発的な立ち話等も，いつ，どこで，誰と，何を，といったインデックスが付与されるであろうし，ミーティングの重要シーンの特定も容易になるであろう．また，行動予測や状況に応じたサービス提供も発展するであろう．

☆1 http://www.aaaivideos.org/2009/situated_interaction/

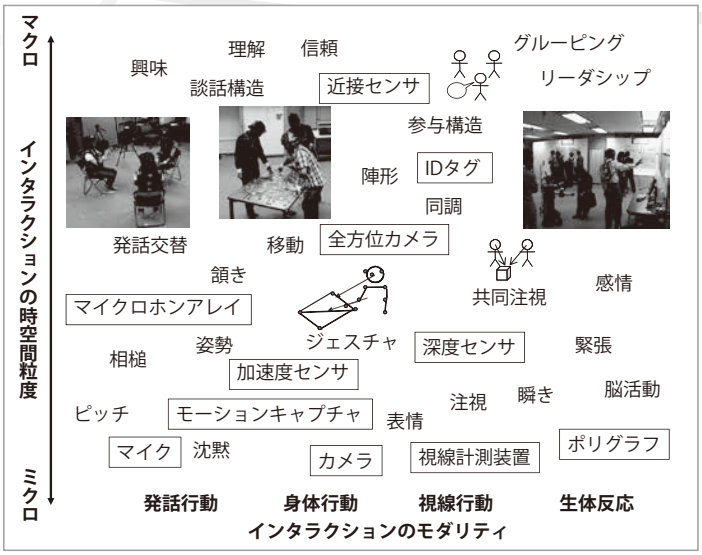


図 -1　インタラクションを構成する要素の関係

インタラクションの何に注目するのか？

ヒトのインタラクションを計測・理解するための方法論を概観する☆2．本稿では，少人数（2, 3 人）の間で数秒単位の短い時間内で起きるやりとり（ミクロなインタラクション）から，集団内で数分単位で起きる社会的ダイナミクス（マクロなインタラクション）までを考える．

ミクロなインタラクションとは，たとえば，話し手の発話に対する聞き手の頷きや，誰かが指差した方向への視線移動といった，無意識の反射に相当するようなインタラクションである．一方，マクロなインタラクションとは，多人数が行き来する公共空間における会話グループの発生・散逸や，グループの中で会話のリーダシップを持つメンバが推移していく様子を指す．

マクロなインタラクションは，ミクロなインタラクションを要素とした組合せで構成されると考えられる．以下，**図 -1** を見ながらインタラクションを構成する要素の関係を説明する．

図 -1 では，インタラクションの時空間粒度を縦軸とし，ミクロなインタラクション要素からボトムアップ的にマクロなインタラクションが構成されていく様子を表現した．また，インタラクションのモダリティを横軸とし，インタラクションの構成要素を配置した．インタラクションのモダリティとして，発話行動，身体行動，視線行動，生体反応に着目して観察可能なインタラクション要素の分類を試みた．なお，図中には，それらのインタラクション要素を計測するためのセンサ類を矩形で囲んだ文字で表記した．

発話行動が発する情報の中心は言語情報である．しかし，言語情報の周辺には，発話のピッチ，沈黙，強弱等の変化があり，それらの非言語的な情報から，発話内容に対する自信や重大さを暗黙的に伝達している．また，指差しなどのジェスチャを用いることで，言語情報の曖昧さを軽減したり，聞き手の注目を獲得する．ジェスチャには，何か具体的なものの形状を表現しようとする図像的なジェスチャや，発話のテンポをとるための拍子的なジェスチャもある．

身体行動としてはほかにも，姿勢（体の向き，前かがみ等），頭部ジェスチャ（頷きや首振り），表情などがある．聞き手の姿勢からは，目の前で起きているインタラクション要素（発話や指差し）への興味を読み取ることができるし，頭部ジェスチャは発話への同意／反対の意思を簡潔に伝えたり，相手の発話行動を促す効果もある．表情は，会話内容に対する感情を表すとともに，発話の聞き取りに対する困難さを表すメタ的な手段としても利用される．身体行動は，発話の言語情報と同時並行して発することができ，かつ，瞬時に発することができる．

このように，ヒトは無意識のうちに，言語情報の周辺で多くの非言語行動を駆使して，言外の情報を伝えたり，インタラクションの制御を行っている．そして，ヒトはそれを読み取り，会話や共同作業に活用している．これだけ見ても，現状のコンピュータがいかにヒト同士の自然な情報伝達のごく一部しか利用していないかが理解できよう．

「目は口ほどにものを言う」と言われるだけあって，視線行動からも多くのことが読み取れる．注視はその対象物への興味を示す．また，複数人で会話している場合は，会話参加者の視線を集めている人が次話者になる場合が多い．瞬きは，集中や緊張等の心理的状態を表す．

視線行動の分析においては，共同注視，すなわち，

☆2　詳細は，たとえば文献 3）を参照されたい．

複数人が同時に同じ対象物を注視するという現象に注目することが多い．共同注視に注目することで，会話や共同作業における重要なシーンの特定や，その内容の推定が可能になる．また，ジェスチャなどの身体行動がどの範囲のメンバに「認定」されているのかを判別するのにも共同注視の検出は役立つ．

さらにマクロなインタラクションとして分析されるものとしては，発話交替，（姿勢や発話の）同調，（立ち話時の）陣形などがある．これらのマクロなインタラクションの解釈は，これまでに挙げてきたミクロなインタラクション要素の組合せによって成り立つ．そしてさらに，これらのマクロな解釈が，さらに上位のマクロな解釈（たとえば，会話の談話的な構造理解や，会話参加者の立場の変化）への構成要素となる．

それらのマクロなインタラクション解釈から，個々人の興味や理解，そして，互いの信頼関係や，誰がリーダシップをとっているかといったことを推定する試みもある．これらの推定は，我々ヒト自身にも難しいことであるが，こういった外部から観測可能な非言語情報から推定していることは間違いない．したがって，社会的パートナーとなる人工的な社会知が，これらのマクロなインタラクションの解釈や利用に挑戦するのは妥当であろう．

インタラクションをどう計測するのか？

これらのインタラクションはどのように計測されるのであろうか．大まかに言うと，発話行動の音声要素はマイクで計測され，身体行動や視線行動などの映像要素はカメラで計測される．さらに個々のインタラクション要素を計測するために機器が進化しており，発話者や音源方向推定のためにマイクロホンアレイが開発され，姿勢，ジェスチャ，表情を計測するためにモーションキャプチャシステムが開発された．視線行動を計測するには，視線計測装置が利用される．視線計測装置には，大きく分けて，頭部装着型のものと据え置き型のものがあり，計測状況の多様性は前者の方が勝るが，注視対象を機械的に特定するにはモーションキャプチャ等との併用が必要となる．

昨今は，会話グループの正面顔の記録と頭部方向の近似推定を可能にする全方位カメラの利用や，Microsoft の Kinect や Intel の RealSense 等の深度センサを用いた人物，姿勢，ジェスチャの推定などの技術が進んだ．これらの技術の発展は，インタラクション理解の大衆化を進め，専門家による分析だけでなく，新しいインタラクション体験を実現するのにも役立っている．

一般的に，ミクロなインタラクションの計測は，データ量が多く，そのためノイズが多い．そして，それらから解釈をボトムアップに高めることは，多くのヒューリスティクスを必要とする．したがって，マクロなインタラクションの理解は一般的には難しい．しかし，マクロなインタラクションを理解したいときに，必ずしも，ミクロなインタラクション計測からボトムアップ的に解釈する必要はない．たとえば，単純な身体行動（たとえば，頷き等）は加速度センサを使って計測すれば良い．また，集団の陣形やグルーピングの現象は，近接センサや無線等の ID タグから近似的に計測すれば用が足りる場合もある．

身体装着型のセンサを利用することで生体反応（脈拍，体温，筋電，脳波等）を計測することができる．これらの計測にはポリグラフと呼ばれる，生体から漏れ出る微弱な電気信号を計測する装置が用いられる．これまで見てきたモダリティの多くが，ヒト自身が眼や耳から観測可能な情報に基づいていたのに対して，生体反応データは，ヒトが直接は観測できない内面的な心理的現象を垣間見ることができるのではないかと期待されている．となれば，ヒトの社会知に追いつこうとしてきた人工知能が，ヒトの社会知を超える可能性もある．

参考文献

1） 西田豊明，角 康之，松村真宏：社会知デザイン，オーム社（2009）．
2） Bohus, D. and Horvitz, E.：Facilitating Multiparty Dialog with Gaze, Gesture, and Speech. In *12th International Conference on Multimodal Interfaces and 7th Workshop on Machine Learning for Multimodal Interaction（ICMI-MLMI 2010）*, 8pages, ACM (Nov. 2010).
3） 坊農真弓，高梨克也（編）：多人数インタラクションの分析手法，オーム社（2009）．

（2015 年 7 月 13 日受付）

角　康之（正会員） sumi@fun.ac.jp

1995 年東京大学情報工学専攻博士課程修了．（株）国際電気通信基礎技術研究所（ATR），京都大学情報学研究科を経て，2011 年より公立はこだて未来大学教授．専門は，知識や体験の共有を促す知的システムや，人のインタラクションの理解と支援にかかわるメディア技術．