人に迫り，人を究め，人に寄り添う ——デジタルとナチュラルの共生・共創に向けて

昨今，AI（人工知能）は特定の機能では人間の性能に迫るほどめざましく進歩していますが，まだ限定的です．一方人間は高度に複雑ですが，それゆえにバイアス（偏り）や錯覚に支配されるなど，不完全で誤りを犯します．本稿では，人間に迫るべくAI技術を研ぎ澄ませていくのと同時に，人間をさらに深く知ることで両者のギャップを埋め，人間に寄り添うAIを実現するためのコミュニケーション科学の取り組みを紹介します．

山田 武士（やまだ たけし）

NTTコミュニケーション科学基礎研究所 所長

はじめに

最近のAI（人工知能）技術の発展にはめざましいものがあります．もともとコンピュータは人間が処理できない大量のデータを一度に処理し，人間が苦手な処理を人間に代わって高速に処理するのが得意です．しかし特に深層学習の発展のおかげで，本来人間が得意で，なかなかコンピュータが追いつけなかった音声や画像の認識や自然言語処理などにおいても，人間の能力に迫り，場合によっては凌駕する性能を実現しつつあります．このようなメディア処理を中心に，今後さらにAIの進歩は加速すると期待されます．とはいえ脳の処理は複雑であり未解明の部分も多く残されています．AIの性能が複雑な人間の脳を超えるほどに進歩するのはまだ先といえます．

一方で人間は認知上のバイアス（偏り）にとらわれ間違いを犯したり，実際にはありもしない錯覚をリアルに感じてしまったりなど，複雑であるがゆえに一見すると不完全な存在でもあります．このように，限定された範囲で急速に発展を続けるコンピュータ（AI）と，複雑であるがゆえに不完全でもある人間とをつなぎ，両者のギャップを埋めることが「コミュニケーション科学」を研究所名に掲げるNTTコミュニケーション科学基礎研究所（CS研）の使命です（**図**）．これをふまえてCS研は人と人，あるいはコンピュータと人の間の「こころまで伝わる」コミュニケーションの実現をめざし，基礎理論の構築と革新技術の創出に取り組んでいます[(1)]．地道な基礎理論の構築の例としては，符号化効率の限界（シャノン限界）まで効率良くメッセージを送受信する符号化法の提案が挙げられます．こちらについては本特集記事『限界まで効率良くメッセージを送れます——シャノン限界を達成する通信路符号』で詳しく説明します[(2)]．今後さらに「こころまで伝わる」をめざすためには，メディア処理を中心とした人間の能力に迫る技術を追究するのはもちろんのこと，人間の機能，特性を解明し，人間のことをよく理解すること，そのうえで人間に寄り添う技術の実現をめざすことが一層重要であると考えています．

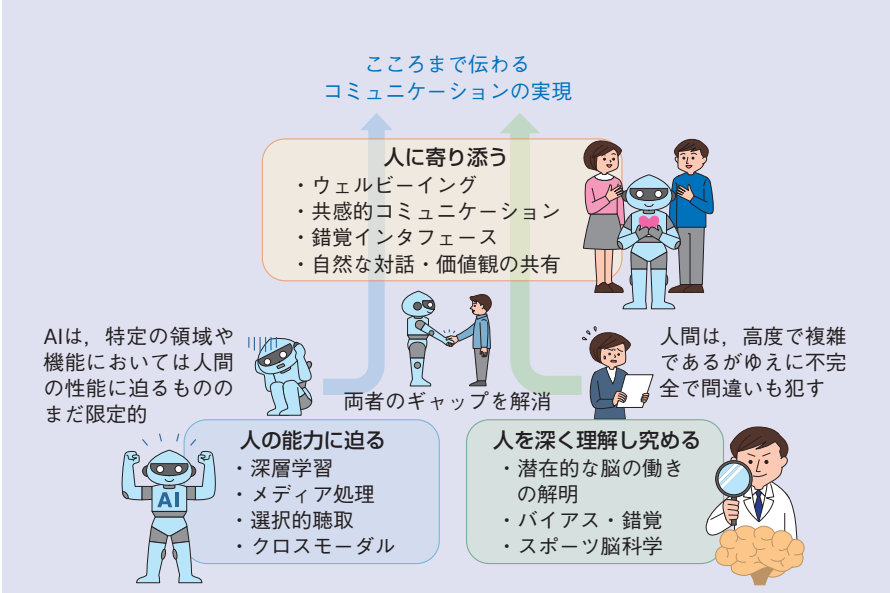

図 コミュニケーション科学のミッション

人間の能力に迫る技術

世の中にはまだまだ，人間は得意でも，コンピュータには苦手な処理が多数存在します．確かに機械翻訳の精度は飛躍的に向上し，大学入試の英語穴埋め問題をある程度正解できるようにはなりましたが[3]，文章の意味を深く理解したり，常識を身につけたり，というレベルにはまだ到達していません．

一方で，深層学習技術を駆使することで，画像認識や音声認識など，特定の面では人間の能力に迫ってきたことも事実です．例えば，会議やパーティでの歓談などにおいて，複数の人が同時に話したり，背景に音楽が流れていたりするとします．人間はこのような状況においても「聞きたい」人の声の特徴を選り分けて，話す内容を聞き取ることができます．これは人間の聴覚の優れた能力の１つで，選択的聴取と呼ばれます．選択的聴取はより広い概念である選択的注意の代表例です．従来，このような選択的聴取を，コンピュータは苦手でしたが，CS研では独自の深層学習技術により，人間同様，コンピュータが目的話者の声の特徴に基づき，その声だけを聞き取る技術を実現し，さらにそれを発展させています[4]．

これらのメディア処理技術が今後さらに進歩し，人間に近づくための鍵となるのがクロスモーダル処理です．クロスモーダル処理とは，「音声」「映像」「テキスト」など単一の「モダリティ」の垣根を越えた処理，という意味です．従来，これら「音声」「映像」「テキスト」などはそれぞれ解析手法も異なり，別々に研究されてきました．しかしここに来て，深層学習といういわば「共通言語」が整備されたおかげで，モダリティの垣根を越えた「認識」「生成」「変換」が可能になりつつあります．

一方，人間は常に複数の感覚（五感）を駆使して外界を知覚し，例えば，音声を聞いただけでその場の情景をある程度頭の中に思い浮かべることができるなど，このようなクロスモーダル処理を日常生活の中で当たり前のように行っています．また，目の見えない人が指先を使って点字を読むといった，障害などで損なわれた感覚の機能を残された感覚で代行する「感覚代行」もよく知られています．確かに人間なら，顔写真を見てその顔に合った声をある程度は想像できるかもしれませんが，そんなことがコンピュータに可能でしょうか？　CS研では実際に，これらのクロスモーダル処理をコンピュータで実現することに取り組んでいます．例えば「音から画像認識」するクロスメディア情景分析技術では，カメラでは死角になってしまうような個所の情報も音を使って「認識」できることをめざしています．CS研が取り組む最新のクロスモーダル処理技術については本特集記事『画像や音を見聞きするだけで賢くなるAI ――クロスモーダル情報処理の展開』で詳しく説明します[5]．

人間を深く理解し究める技術

このように特定の場面ではAIの能力は人間に近づき，凌駕しつつあります．しかし，AIの性能が複雑な人間の脳を超えるほどに進歩するのはまだ先でしょう．一方で人間は，「振り込め詐欺」にも簡単に騙されるなど，時として認知上のバイアスに支配されたり，錯覚にとらわれたりして，自分でも思いがけない誤りを犯します．CS研が運営するWebサイト「イリュージョンフォーラム」には自分の目や耳が信じられなくなるような，さまざまな錯覚の情報が掲載されています[6]．

クリストファー・チャブリスとダニエル・シモンズによる有名な実験[7]では，実験参加者は白シャツと黒シャツの合計６名の選手がバスケットボールをパスする映像を見せられ，白シャツチームの間でボールがパスされる回数を数えるように指示されます．このとき，映像の途中で９秒間かけて，舞台袖からゴリラが現れ，正面でカメラに向かって胸を叩き堂々と去っていくのですが，回数を数えるのに夢中の半数の実験参加者はそのことに気付きませんでした．このように人間はあることに注意を向けると，周囲で起こっている別のことに注意が向かなくなります．すなわち，人間の優れた特性である選択的注意は，裏を返せば選択的不注意であるわけです．しかもそうなっていることに本人は気が付きません．振り込め詐欺に騙されるのは高齢者だけとは限らないのです．

このように複雑であるがゆえに「バイアス」や「錯覚」にとらわれがちで不完全な人間と進歩しつつも今のところ限定的なAI，この両者のギャップを埋めて共生・共創していくためには，安易に「AIが人間の脳を超える」などと信じ込む前に，複雑な人間をもっと深く知ることが必要です．そのために，CS研では「視覚」「聴覚」「運動感覚」といった人間の基本的な感覚に関する「潜在的な脳の働き」の解明に注力しています．錯覚も「潜在的な脳の働き」の解明の重要な手掛かりです．

一口に脳の働きといっても人それぞれ多様です．CS研では，優れた運動能力を持つ一流アスリートに着目し，脳科学の視点から人間の「心・技・体」の関係の本質に迫る，スポーツ脳科学にも取り組んでいます．例えば，優れた打者がわずか0.1秒という短い時間で，いかに遅い球と早い球を見極めて球種に応じたタイミングで動いているか，その仕組みの解明などに挑んでいます．スポーツ脳科学は，ICTを駆使して主に体を鍛える従来のスポーツ科学や，パフォーマンスのみを評価するスポーツ分析手法とは一線を画した，野心的な取り組みです．

ちなみに，前述のクロスモーダル処理は脳内でもさまざまなレベルで行われています．例えば，通常の映像を見るとき，脳は映像中の「動き」「色」「形」（モダリティ）の情報を個別に処理し，後にそれらを統合します．したがって，これらの情報間に不整合があったとしても，統合する過程でそれは補正されます．この脳の処理の仕組みを利用してCS研で考案されたのが変幻灯®です[8]．変幻灯を体験するとき，ユーザは色や形は止まった対象から取得し，動きは投影されたモノクロの映像から取得します．色や形は止まっているので，動きと空間的に「ずれ」が生じます．しかし，辻褄が合ったようにものを見ようとする脳は，「動き」「色」「形」を統合する際に，その「ずれ」を補正します．そのため，変幻灯を体験する際には，ユーザは「動き」「色」「形」のずれに気付かずに，あたかも止まった対象の色や形が動いているように「錯覚」して感じるのです．

人間に寄り添う技術

スポーツ脳科学で得られる成果はスポーツに限らず，人間が普段の生活の中で心身の潜在能力を最大限に発揮する，すなわち，ウェルビーイングのための知見として活かすことができます．この人間のウェルビーイングという，一見，定性的でとらえどころのない課題を人間科学の立場から定量的に扱い，向上させるための設計指針の確立にもチャレンジしています．例えば，複数の人間が，場を一緒に共有することで生じる共感的コミュニケーションの効果測定などがその例です[9]．また，TVやスマートフォンなどのディスプレイ機器に日々囲まれて，ともすると目を酷使する現代人のために，汎用的なタブレット機器を用いてゲーム形式で日常的に目の状態をセルフチェックできる方法も提案しています．こちらは本特集記事『あなたの目の機能を気軽に楽しく測ります』で詳しく説明します[10]．

一方，錯覚は「潜在的な脳の働き」解明の手掛かりとして重要なのはもちろんのこと，人間とAIとのギャップを埋め，人間に寄り添うインタフェースやフィードバックのための鍵でもあります．CS研ではこれまで人間の錯覚を利用したインタフェースとして，引っ張られる錯覚を生じさせるデバイス「ぶるなび®」を考案しました．さらには，座っているのにあたかも歩いているような感覚の生成にも取り組んでいます．こちらは本特集記事『座っていても歩いているような疑似感覚の生成技術』で詳しく説明します[11]．また，前述の，印刷した絵や写真に光を当てるだけで動き出して見える「変幻灯」，3Dメガネを掛けると3D映像に，メガネを外すと鮮明な2D映像を楽しめる「Hidden Stereo」，印刷物などの2次元平面上の対象に対して影に見えるパターンを投影することで，その対象があたかも3次元的に浮き上がって見える光投影技術「浮像®（うくぞう）」[12]などを次々と生み出してきました．これからも，錯覚を利用した新たなインタフェースの提案と同時に，錯覚を通して物理的には生じ得ない体験を生み出すことによる，斬新な知覚表現の可能性も追究していきます．

ロボットやAIと人間との自然な対話を実現する，対話処理技術においては，重要なのは音声認識や自然言語処

理であって，一見，人間のバイアスや錯覚とは無関係にも思えます．しかしAIは人間のように文章の意味を深く理解したり，常識を身につけたりまではできないため，人間とAIの対話は現状ほぼ「一問一答式」に限られます．したがって，話していると少し前に言ったことと矛盾することを言うなどすぐボロが出て，対話は長続きしません．そこで，その限られた能力を効果的に活用しつつ人間のバイアスや錯覚を利用し，人間にとっていかに「賢く見せる」かが重要となります．CS研では2台のロボットでうまく役割分担をすることで，たとえ一問一答式であっても，自然な対話が長続きする対話処理を実現してきました．さらに一問一答式から脱却するために，ユーザの発話の多くがイベントに関する内容であることに着目し，イベント単位に構造化して把握する手法を提案しました．こちらは本特集記事『文脈を理解して話す雑談対話システム』で詳しく説明します(13)．こうすることで文脈理解度が向上するとともに，イベントにマッチするシステムの擬似経験も共有させることができ，その結果，ロボットに対する共感を誘発するなど，まさに人に寄り添う対話が可能になります．

おわりに

以上見てきたように，人間は高度で複雑であり，一方AIは特定の領域，機能においては人間の性能に迫るもののまだ限定的です．「人間を超える知能」はそう簡単には実現しないでしょう．しかし人間は複雑であるがゆえに不完全で，振り込め詐欺に騙されたり，因果関係を錯誤したり，選択的「不注意」とでも言うべきバイアスにとらわれ間違いを犯したりします．また，人間はありのままの物理量を見ているわけではないことが錯視例などからも分かります．以上のことからも，人間に迫るべくAI技術を研ぎ澄ませていくのと同時に，人間をさらに深く知ることによって，両者のギャップを埋め，デジタルとしてのAIが人間にナチュラルに寄り添い，両者が共生・共創する社会を実現することが重要であり，それが「こころまで伝わる」につながるCS研のミッションであると考えています．

■参考文献

(1) 山田：“新たな次元へとシフトする――さらに深化するコミュニケーション科学の取り組み，”NTT技術ジャーナル，Vol.30, No.9, pp.8-11, 2018.
(2) 村松：“限界まで効率良くメッセージを送れます――シャノン限界を達成する通進路符号，”NTT技術ジャーナル，Vol.31, No.9, pp.26-30, 2019.
(3) 東中・杉山・磯崎・菊井・堂坂・平・南：“「ロボットは東大に入れるか」における英語問題の回答手法，”NTT技術ジャーナル，Vol.27, No.4, pp.63-66, 2015.
(4) Delcroix・Zmolikova・木下・荒川・小川・中谷：“SpeakerBeam：聞きたい人の声に耳を傾けるコンピュータ――深層学習に基づく音声の選択的聴取，”NTT技術ジャーナル，Vol.30, No.9, pp.12-15, 2018.
(5) 柏野：“画像や音を見聞きするだけで賢くなるAI ――クロスモーダル情報処理の展開，”NTT技術ジャーナル，Vol.31, No.9, pp.10-13, 2019.
(6) http://www.kecl.ntt.co.jp/IllusionForum/
(7) http://www.theinvisiblegorilla.com/videos.html
(8) 河邉・吹上・澤山・西田：“変幻灯――止まっている対象を錯覚的に動かす光投影技術，”NTT技術ジャーナル，Vol.27, No.9, pp.87-90, 2015.
(9) 渡邊・大石・熊野・Hernández・佐藤・村田・麦谷：“ウェルビーイングを測る，知る，育む，”NTT技術ジャーナル，Vol.30, No.9, pp.29-32, 2018.
(10) 丸谷・細川・西田：“あなたの目の機能を気軽に楽しく測ります，”NTT技術ジャーナル，Vol.31, No.9, pp.14-17, 2019.
(11) 雨宮：“座っていても歩いているような疑似感覚の生成技術，”NTT技術ジャーナル，Vol.31, No.9, pp.18-21, 2019.
(12) 河邉：“「浮像（うくぞう）」――影を利用して印刷物に見かけの奥行きを与える光投影技術，”NTT技術ジャーナル，Vol.30, No.9, pp.20-23, 2018.
(13) 成松・杉山・水上・有本・宮崎：“文脈を理解して話す雑談対話システム，”NTT技術ジャーナル，Vol.31, No.9, pp.22-25, 2019.

山田 武士

今後ますます技術の進歩のスピードが速くなり，競争が厳しくなる中で，CS研は，人に迫り，人を究め，人に寄り添う技術を中心に，これからも新たなチャレンジに大胆かつ粘り強く取り組んでいきます．

◆問い合わせ先
NTTコミュニケーション科学基礎研究所
企画部
TEL 0774-93-5020
FAX 0774-93-5026
E-mail cs-liaison-ml@hco.ntt.co.jp