

あなたを・もっと・知りたくて —— AIで人に迫り脳科学で人を究める

NTTコミュニケーション科学基礎研究所（CS研）は、「ここまで伝わる」コミュニケーションの実現をめざし、「メディア処理」「データ・機械学習」など、「人間の能力に迫り凌駕する」ための革新技術の創出と、「人間科学」「多様脳科学」など、「人間を深く理解する」ことにつながる基本原理の発見に力を入れています。基礎研究の立場から、成果を世の中に具体的に届けることも試行しています。本稿では、このような取り組みのいくつかを紹介します。

やまだ たけし
山田 武士

NTTコミュニケーション科学基礎研究所 所長

はじめに

今からちょうど35年前、日本電信電話公社（電電公社）が民営化され、NTTが発足した1985年当時、電話は当時の流行歌の歌詞*にあるように、親しい人が「いま何してるの」とか「いま何処にいるの」などを知るためのコミュニケーションツールの主役でした。現在はスマートフォンが普及し、ソーシャルメディアが新たな主役として発達した結果、あまり親しくない人であっても、それがある程度分かってしまいます。そもそも、個人が日々利用するスマートフォンは、これらの情報をすべて把握しており、むしろ使用者本人よりも詳しいかもしれません。一方、電電公社時代の黒電話には不思議な存在感やぬくもりがあったと記憶しています。今後さらに技術が発達すると、コミュニケーションはどう変化

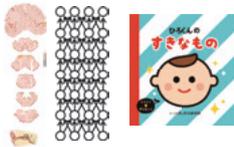
するのでしょうか。人と人との物理的な距離が遠くならざるを得ない時代だからこそ、コミュニケーションの本質が何かを究めることはなおさら重要です。

NTTコミュニケーション科学基礎研究所（CS研）は今から約30年前、1991年の設立当初から、コミュニケーションの本質は、「情報を正確かつ効率良く伝達すること」のみならず、「お互いに理解を深め、感動を共有し、心のふれあいを実現すること」であるとの理念のもと、時代を先取りした基礎研究に取り組んできました。当初は人と人とのコミュニケーションが主題でしたが、現在は、人とのみならず、人とコンピュータとの間の「ここまで伝わる」コミュニケーションの実現をめざし、「メディア処理」「データ・機械学習」など、「人間の能力に迫り凌駕する」ための革新技術の創出と、「人間科学」「多様脳科学」など、「人間を深く理解する」ことにつながる基本原理の発見に力を入れています⁽¹⁾。

*「あなたを・もっと・知りたくて」は、1985年、NTT民営化直後の「TALK ON THE PHONE」イメージソング。

人間科学

五感や知覚、運動にかかわる、人間の脳や身体の情報処理機構の解明と斬新な知覚体験の提供



メディア処理

クロスモーダルなメディア情報処理に基づく柔軟な状況理解や、感覚に訴える高度なメディア表現の実現



「あなたを・もっと・知りたくて」を究める基礎研究



多様脳科学

トップアスリートなどの優れた認知機能を支える脳機能の多様性解明を通じた、人間の「心・技・体」の関係の本質的理解



データと機械学習

世の中に溢れる大量の言語情報と刻々と変化する実世界情報とを結び付けた広汎な知能処理の実現



- 銀行担当
- …
- 親疎
- 改鑄

図 CS研の研究領域

まさに「あなたを・もっと・知りたくて」を究める基礎研究です(図)。また、それらのための地道な基礎理論の構築にも継続して取り組んでいます。後者の例については本特集記事『量子情報処理における量子的間接制御の可能性』で詳しく説明します⁽²⁾。そして、成果を世の中に具体的に届けることにも、基礎研究の立場からパートナーの皆様とのコラボレーションを通じることで工夫して取り組んでいます。本稿ではそのような取り組みの一部を紹介します。

人間の能力に迫り凌駕する

コミュニケーションの基本はまず話し言葉

を認識することです。CS研が運営する「錯覚」解説サイト「イリュージョンフォーラム」で紹介されている「錯聴」の1つに、「モザイク音声(劣化雑音音声)」があります⁽³⁾。これは、雑音のない通常の話言葉の音声を加工しスペクトルの細かい特徴を破壊した、いわば粗いモザイクをかけた映像のような音声のことです。実際聴いてみると、歪みの大きい不自然で聞き取りにくい音声ですが、何と言っているのかなんとか分かります。このように人間は非常に粗い情報だけでも、ある程度声の内容を聞き取ることができます。また、興味深いことに、仮に最初は分からなくても、いったん加工前の音声を聞いた後でま

た聞くなどすると聞き取ることができて、いったん聞き取れると、もはやそのようにしか聞こえません。まさに人間の聴覚機能、音を認識する機能の奥深さを体験できます。

人間を含む動物（哺乳類）は、その長い進化の過程でこのような音を認識するための脳の機能を獲得してきたと考えられます。一般に哺乳類の脳では、耳にある末端の内耳により空気振動としての音信号が神経活動に変換され、聴神経に伝達されます。そこから多数の脳部位を経て階層的に処理され、中枢の大脳皮質にある聴覚皮質に伝達されます。その際、処理できる音の上限周波数や最適変調周波数が末端から中枢へと階層を経るに従って徐々に低下する傾向があります。CS研では、音を直接入力として、自然音の分類課題を行うように深層ニューラルネットワーク（DNN: Deep Neural Network）を学習させると、獲得したDNNの階層構造が全体にわたって脳の階層構造と類似することを発見しました⁽⁴⁾。この類似性はDNNの訓練が進む過程で徐々に獲得され、音の認識精度が高いDNNほど類似性が高くなります。この結果は、動物の脳が持つ性質が進化の過程で音認識に適応した結果、獲得されたものである可能性を示すものといえます。

一方人間は、複数人が同時に話す状況でも、聞きたい人の声に集中し、その声を聞き取ることができます。CS研では、独自に考案したDNNに基づくSpeakerBeamという技術を考案し、人間が持つこの選択的聴取と呼ばれる能力を、コンピュータ上で実現す

ることに成功しました⁽⁵⁾。SpeakerBeamを用いると、システムに目的話者の声の特徴を教えるだけで、複数人の会話音声から、目的話者の音声を取り出すことができます。しかし、音声情報のみでは声の性質が似た話者どうしの音声が含まれる場合など、取り出すことが困難です。そこで声の特徴に加え、唇の動きを手掛かりとすることで、似通った声の人でも聞き分けることにも取り組んでいます。また、DNNを駆使した音声変換技術によって、人が話した音声の内容は保持したまま、声質や抑揚などの特徴を自由に変えることにも取り組んでいます。こちらについては本特集記事『あなたの声はどんな声？どんな声でしゃべりたい？』で詳しく説明します⁽⁶⁾。これらの技術をさらに発展させると、発声・聴覚機能の障がいや加齢による衰えを克服した自然なコミュニケーションや、不慣れな外国語での会話サポートなどが実現できます。

幼児の言語獲得過程の解明も進めています。人間の幼児は親とのコミュニケーションを通じて言語を習得します。人類は長い時間をかけて言語とそれを用いたコミュニケーションを進化させてきました。しかし、人類が文字を使うようになったのは比較的最近のことで、「読む」能力はもともと脳に備わった機能ではありません。「視覚」「聴覚」「言語」「認知」など脳の基本機能を柔軟に組み合わせて実現しています⁽⁷⁾。CS研では、言語獲得の仕組みを理解するために、子どもがいつどんな語を理解し、発話できるかを大規模調査し、モデル化した「幼児語彙発達データベース」

を構築しています。CS研の語彙研究については本特集記事『言葉の難しさを測る ― テキストの難易度と人の語彙数の推定』で詳しく説明します⁽⁸⁾。このデータベースは、子どもの「読む」能力の育成にも役立つと考えています。そこでNTT印刷を通じてサービス化されたのが、お子さん1人ひとりの言葉の成長に合わせて内容をカスタマイズする、「パーソナル知育絵本」です。CS研とNTT印刷は、沖縄県恩納村や徳島市との協力のもと、お子さんに「パーソナル知育絵本」を配布し、成長の早い段階から絵本を読むことに興味を持ってもらい、それを習慣化させる取り組みを始めています⁽⁹⁾、⁽¹⁰⁾。

さらに人間の持つ優れた言語処理、知識処理に迫る取り組みとして、CS研では数年前から国立情報学研究所の人工知能プロジェクト「ロボットは東大に入れるか」（東ロボ）に参画し、人間が実際に解く問題をAI（人工知能）がどこまで解けるのかを明らかにする研究を進めています。特にCS研は英語を担当し、大学入試センター試験の英語筆記科目に挑戦してきました。その結果、最新のDNN技術に基づき、CS研を中心とした東ロボ英語チームの独自技術を適用することで、2019年センター試験の英語筆記本試験において、185点（偏差値64.1）という極めて高い成績を達成しました⁽¹¹⁾。英語問題は、自然言語処理、知識処理の統合的な問題を多く含んでいます。NTTでは得られた知見を、人間との雑談（雑談対話）や、受付や案内での情報提供など（タスク対話）を行う対話型

AIの研究に活かしています。

対話型AI研究の一例として、NTTでは「なりきりAI」と呼ばれるキャラクタ型の雑談対話技術に取り組んでいます。なりきりAIとは、さまざまなユーザに、あるキャラクタ（有名人や小説、ゲームのキャラクタなど）に「なりきって」質問やその回答を入力してもらい、蓄積した対話データに基づき、そのキャラクタの振る舞いを再現した対話型AIを構築するものです。これまでもNTTドコモ、株式会社ドワンゴとの連携を通じて、複数のキャラクタを対象としてなりきりAIを構築してきました。このたびCS研の所在地である京都府相楽郡精華町との連携による「なりきりAI京町セイカ」プロジェクトが始まりました⁽¹²⁾。「京町セイカ」は精華町の公式広報キャラクタで、SNSなどを通じたファンも多く、そのコンセプトはなりきりAIにぴったりです。CS研ではこれまでも、雑談対話とタスク対話を自由に行き来し、対話の話題を自然に制御する技術の基礎研究を行ってきました。「なりきりAI京町セイカ」は、「人の心に寄り添う対話」実現に向けたこれら技術の実証の場となります。まずは精華町民の皆様とご協力のもと、精華町に関する多数の知識や経験を含む対話データを収集します。そして個性的なキャラクタを持つ京町セイカが、適切な情報提供や相談をタスク対話として行う合間でもユーザからの雑談や質問に柔軟に答えるなど、ユーザの気持ちや要望に寄り添うタスク対話を実現します。

人間を深く理解し究める

一方で、AIの発展により、人間のことを深く知り、理解することは、ますます重要になってきました。最近のインターネット検索サイトは各ユーザの検索履歴等を把握しており、例えば、ユーザがインターネットで調べものをしている最中に、検索ワードにマッチした商品広告がさりげなく提示されると、いつのまにかクリックして購入してしまう、ということが起こり得ます。その際ユーザは、あくまでも自分の意志でその商品を購入したのであって、第三者に操られて商品を買わされたとは、あまり自覚しません。今後AI技術が発展すると、より巧みな、いわばAI版「サブプリミナル効果」のリスクが高まると考えられます。

このようなリスクを防ぐためにも、人間がいつ、どのような思い込みを持つのか、そしてその思い込みがどう行動につながるかなどを、人間科学、特に脳情報処理の観点から深く理解することが重要です。CS研では卓越したスキルを持つアスリートに着目し、彼らが情報をどうとらえて判断しているのかを、身体に現れるさまざまな生体情報の解析を通じて解明する研究に取り組んでいます。例えば野球選手において、打てる打者と打てない打者は何が違うのか、「ボールを良く見て打つ」というのは本当か、ストレートは本当に「まっすぐ」か、などです。これらの知見をアスリートに効果的にフィードバックすれば、脳機能を鍛えるトレーニング手法としても活

用できます。

ウィズコロナの時代、ともすると「心のふれあい」はますます希薄になりがちです。「神の見えざる手」で有名なアダム・スミスは、その著書『道徳感情論』で「共感」(Sympathy)の重要性について論じる中で、「我々はしばしば他人の悲しみから(自らの)悲しみを引き出す」と述べています⁽¹³⁾。感情(Sentiment)や情動(Emotion)の共有、他者の経験を自分のものと感じること、それすなわち共感であり、心のふれあいに通じます。CS研では、東京大学 亀田達也研究室との共同研究により、「痛み体験の共有」について研究しています。初対面の2人が同時に痛み刺激(熱刺激)を受ける実験において、2人の痛みに対する不随意、すなわち意識しない、自律神経反応(ここでは指尖容積脈波という、指先の脈に応じた血液の変化量を測定)が対面状況では同期するが非対面では同期しないことが分かりました。また、この刺激を繰り返し経験すると、反応の弱い人が強い人に合わせるかたちで変化することを明らかにしました⁽¹⁴⁾。すなわち、対面のインタラクションにおいて、痛みのようなネガティブな情動が、意識しない、自律神経反応のレベルで、個人間で共有され、増幅することが示唆されたのです。さらにeスポーツでの1対1の対戦を対象とした実験においても、上級者どうしの接戦では、プレイヤーどうしの心拍数が同期して変動すること、しかし一方的な試合展開では心拍数は大きく乖離することが分かりました。その他、CS研で取り組んでいる

人間科学研究については本特集記事『知覚心理学で探る触覚の仕組み』⁽¹⁵⁾、『巧みで素早い運動を支える脳内情報処理——視覚的な身体情報による伸張反射の調整』⁽¹⁶⁾でさらに紹介します。

おわりに

日本語の「幸せ」の語源は、「為合わせ」や「仕合わせ」、対応する動詞は「仕合わせる」であって、これはまさに、他者とのインタラクションやコミュニケーションが「うまく行った」状況をさしているという説があります⁽¹⁷⁾。すなわち、心のふれあいの相乗効果が人々の幸福、最近の言葉でいえばウェルビーイングの向上につながることとなります。eスポーツでの心拍の同期は、まさに「仕合せている」状態、「仕合」そのものといえます。CS研はこれからも、人の能力に迫る研究、人を究める研究に取り組み、たとえ「離れていても心は君のそばにある」ような「ここまで伝わる」理想のコミュニケーションとは何かを探求していきます。

■参考文献

- (1) 山田：“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,” NTT技術ジャーナル, Vol. 31, No. 9, pp.6-9, 2019.
- (2) 加藤：“量子情報処理における量子的間接制御の可能性,” NTT技術ジャーナル, Vol. 32, No. 9, pp. 34-38, 2020.
- (3) http://www.kecl.ntt.co.jp/IllusionForum/a/noise_vocodedSpeech/ja/index.html
- (4) <https://www.ntt.co.jp/news2019/1907/190710a.html>
- (5) <https://www.ntt.co.jp/news2018/1805/180528c.html>
- (6) 田中：“あなたの声はどんな声？どんな声でしゃべりたい?,” NTT技術ジャーナル, Vol. 32, No. 9, pp. 29-33, 2020.
- (7) S. Dehaene: “Reading in the Brain: The New Science of How We Read,” Penguin Putnam Inc, 2010.
- (8) 藤田：“言葉の難しさを測る——テキストの難易度と人の語彙数の推定,” NTT技術ジャーナル, Vol. 32, No. 9, pp. 39-44, 2020.
- (9) <http://www.nttprint.com/company/itemid419-000048.html>
- (10) <http://www.nttprint.com/company/itemid419-000053.html>
- (11) <https://www.ntt.co.jp/news2019/1911/191118a.html>
- (12) <https://www.ntt.co.jp/news2020/2007/200703a.html>
- (13) https://en.wikipedia.org/wiki/The_Theory_of_Moral_Sentiments
- (14) A. Murata, H. Nishida, K. Watanabe, and T. Kameda: “Convergence of physiological responses to pain during face-to-face interaction,” Scientific Reports, Vol. 10, No. 1, p. 450, 2020.
- (15) 黒木：“知覚心理学で探る触覚の仕組み,” NTT技術ジャーナル, Vol. 32, No. 9, pp. 18-22, 2020.
- (16) 伊藤：“巧みで素早い運動を支える脳内情報処理——視覚的な身体情報による伸張反射の調整,” NTT技術ジャーナル, Vol. 32, No. 9, pp. 23-28, 2020.
- (17) 玄侑：“しあわせる力 禅的幸福論,” 角川SSC新書, 2010.



山田 武士

CS研はこれからも「ここまで伝わる」コミュニケーション探求の取り組みを続けるとともに、パートナーの皆様とのコラボレーションを進めることで、ウィズコロナであっても、離れていても心ふれあう、心が豊かになる社会の実現へとつなげていきます。

◆問い合わせ先

NTTコミュニケーション科学基礎研究所
TEL 0774-93-5000
FAX 0774-93-5015
E-mail takeshi.yamada.bc@hco.ntt.co.jp