

変化する現在^{いま}に適應し，持続する未来^{あす}を切り拓く コミュニケーション科学

——人・社会・環境との調和と共生をもたらす技術の創出

NTT コミュニケーション科学基礎研究所（CS研）は，昨年創立30周年を迎えました。設立当初より，人と人および，人とコンピュータとの「ここまで伝わる」コミュニケーションの実現をめざして，人間と情報の本質に迫る基礎理論の構築と，社会に変革をもたらす革新技術の創出に取り組んできました。本稿では，CS研における最近の研究成果のいくつかと，今後の展望に向けた取り組みについて紹介します。

なや ふとし
納谷 太

NTTコミュニケーション科学基礎研究所 所長

はじめに

NTTコミュニケーション科学基礎研究所（CS研）は，1991年7月4日に京阪奈に設立されました。当初，国際電気通信基礎技術研究所（ATR）内に間借りし，機械学習および情報理論に関する2つの研究グループから開始したCS研は，現在，京阪奈と厚木の2拠点において約150名の所員が集う組織となりました。また，CS研の30年間のあゆみを振り返り，昨年のオープンハウスの開催に合わせて，設立当初からの主な研究成果などをまとめた創立30周年記念ウェブサイトを開設しています⁽¹⁾。

人と人とのコミュニケーションを理解することをミッションとして始まったCS研の研究は，「見る」「聞く」「話す」などに代表されるような，人と同等のコミュニケーション能力をコンピュータに持たせるメディア処理や，量子情報理論や機械学習・データ科学などを含む「人の能力に迫り

凌駕する研究」と，人のさまざまな感覚運動能力のメカニズムを追求する人間科学および，その中でも特にトップアスリートなどの優れた認知能力の解明をめざした多様脳科学などの「人を深く理解する研究」として発展してきました^(図)。これらの研究は，この30年間にわたり，それぞれの年代における技術的な進歩や社会的なニーズを反映して変化し発展しつつも，「コミュニケーションの本質を理解する」というCS研設立当初からの研究理念が貫かれ，人や社会に寄り添う技術の創出に向けて脈々と継続されてきました。以降では，それぞれにおける最近の研究事例のいくつかを紹介します。

人の能力に迫り凌駕する

深層学習技術の登場により，「見る」「聞く」「話す」などのメディア処理に関するAI（人工知能）技術は格段に進歩し続けており，例えば画像認識のコンペティションなどにおいてはす

で人間の識別性能を超えたといわれています。しかし，一般的に深層学習で高い性能を得るためには，入力データとその正解ラベルをペアにした膨大な学習データが必要です。

一方で，人間は少ない情報しか得られない場合でも，これまでの経験からの類推や，欠けている情報を他の情報から補完することなどにより，柔軟で高度な推論を行えます。例えば，写真は三次元の空間情報を二次元画像としてとらえたものですが，人間はこれまでの経験に基づき，写真を見ただけで被写体の持つ形状や奥行きなどの三次元情報のある程度推測できます。

従来，このような能力をコンピュータで学習するには，写真として撮影された二次元画像に加えて，深度センサやステレオカメラなどの特殊な装置で計測した三次元情報を含むデータをペアとして大量に取得し用意する必要があります。これには膨大なコストを要します。このようなデータ取得上の課題を解決するため，CS研では，力



図 NTTコミュニケーション科学基礎研究所の研究領域

メラの持つ光学的な制約として、絞り
とボケの関係を考慮した新たな深層学
習技術「Aperture Rendering
GAN」⁽²⁾を提案しています。本技術
は、インターネット上にある公開画像
などの一般的な写真群（二次元情報）
のみから、奥行きやボケ効果といった
三次元情報を学習することができ、さ
らに、被写界深度を変えたボケ感の異
なる画像を新たに生成することを可能
にしています。

一方、人間は、パーティ会場など複
数の人が会話していたり、雑音や残響
のある環境の中で、聞きたい人の声に
集中して聞き分ける「選択的聴取」と
いう優れた「聞く」能力を持っていま
す。CS研では、これまで音声のみを
用いて聞きたい人の声を聞き分ける技
術SpeakerBeamを提案してきまし

たが、昨年、音声に加えて映像情報を
組み合わせることにより、人間のよう
に複数の手掛かりを活用して選択的聴
取を実現する「マルチモーダル
SpeakerBeam」⁽³⁾を提案しました。
本技術により、声質が似通った複数話
者が存在する場合には唇の動きを主な
手掛かりとし、逆に唇の映像が得られ
ない場合には音声を主な手掛かりとす
ることにより、頑健で高精度な話者音
声の分離を実現しています。本研究
は、音声に限らず、消防車のサイレン
や犬の鳴き声など、注目すべき音だけ
を聞き分けるユニバーサル音抽出技術
へと発展しつつあります。本特集にお
いては、離れた位置にあるマイクで収
録した音声から、話者の近くのマイク
で収録したような高品質の音声を抽出
する最新の音声強調技術⁽⁴⁾を紹介して

います。

CS研では、1990年代から人と自然
に会話する対話システムの研究を続け
てきました。当初は予約や検索などの
特定の目的に特化した対話システムの
研究が中心でしたが、最近では目的に
よらず広い話題を扱いながら、自然な
応答ができる雑談対話システムの研究
を進めています⁽⁵⁾。昨年は、京阪奈
CS研の所在する京都府精華町役場の
協力の下、窓口案内や観光案内などの
業務を行いつつ、雑談対話も楽しめる
AIの実証実験⁽⁶⁾を開始したほか、
Web等で収集した超大規模対話デー
タと深層学習を組み合わせた日本語最
大規模のTransformer対話モデルを
無償公開しました⁽⁷⁾。最近では、車を
運転中の車窓から見た画像や周辺情報
など、その場のコンテキストが時々

刻々変わるような状況を話題とし、ユーザとより一体感が感じられる対話システムを構築しています。さらに、対話相手の嗜好などを記憶し、対話内容の一貫性を保ちながら、より人に寄り添った対話を継続できるシステムの構築に向けた研究に取り組んでいます。

人に寄り添うという点では、機械学習の分野においても進展があります。例えば、融資承認や人材採用など、人を対象とした意思決定を機械学習によって行う場合、単純に予測精度のみを優先する従来の機械学習技術では、性別・人種・障がいなど、人間が持つ微細な特徴に関して不公平な予測になってしまう可能性があります。因果関係に基づく公平・高精度な機械学習⁽⁸⁾は、不公平さに関する事前知識を、特徴・予測結果間の因果関係（因果グラフ）としてモデル化することにより、個々人に対して公平かつ高精度な予測を実現しています。

また、本特集においては、機械学習の最先端の成果として、将来の自動運転社会を見据え、デジタルツインを用いたシグナルフリーモビリティの実現に向けた分散深層学習技術を紹介しています⁽⁹⁾。信号のない道路網において、個々の車どうしが通信して協調することにより、すべての車がぶつかることなく、かつ、全体の車の平均移動速度を向上させる新たな最適化技術です。

人を深く理解する

CS研では、人の感覚知覚運動メカニズムを探るうえで、脳の潜在機能が引き起こすさまざまな錯覚現象を手掛かりに研究を進めてきました。視覚や聴覚に関する錯覚を体験できるWebサイト、イリュージョンフォーラム⁽¹⁰⁾も公開しています。最新では、VR（Virtual Reality）を用いた錯覚提示実験により、壁が動く歩行速度が変化してしまう現象や、投手の投球動作が早くなると打者はボールの伸びを強く感じる現象など、視覚と運動の相互作用が引き起こす興味深い錯覚から脳の多様な潜在機能を明らかにした研究成果が得られています。

また、2021年10月には、静岡県立総合病院と人工内耳装用者の音声・言語認知の共同研究を開始しています⁽¹¹⁾。難聴児であっても、早期に人工内耳を装用することで健聴児と同程度の音声言語を獲得できることが実証されていますが、脳における音声知覚や言語発達のメカニズムは未解明のままです。医学と脳科学の両面からのアプローチにより、高齢者などの難聴者における聴覚機構の態様を明らかにし、音声知覚・言語発達の個人差の背後にあるメカニズムの解明とこれに基づく支援などの研究に取り組んでいます。

言語獲得に関する研究では、CS研は1999年に日本語約8万語の単語のなじみ度合いを7段階で評価した単

語親密度データベースを公開してきましたが、2021年には新たに出現した単語を加えた16万語以上について再調査した「令和版単語親密度データベース」をNTT印刷から提供を開始しています⁽¹²⁾。この単語親密度データベースを用いることにより、50個の単語を知っているか否かを答えることで、おおよその語彙数を推定する技術も構築しており、「令和版語彙数推定」としてインターネットから利用できるWebサイト⁽¹³⁾を公開しています。

人・社会・環境との調和と共生をもたらし技術の創出に向けて

CS研の研究は各分野の専門性を究めることで発展してきましたが、近年では相互の専門分野をまたいだ研究が増えつつあります。例えば、「見る」や「聞く」などの単一のモダリティに特化したメディア処理研究は、前述した音と映像を用いた話者特定などのように、複数のメディアを同時に扱うクロスモーダル情報処理へと発展してきています。人間科学の分野においても多感覚統合のメカニズム解明などに研究がシフトしつつあります。また、昨年、CS研における脳科学・人間科学・メディア処理の知見と技術を総動員した研究成果である「投手シミュレータ」が、世界最大のスポーツイベントにおけるソフトボール日本代表の金メダル獲得に貢献し、新聞やテレビなどで「秘密兵器」として取り上げられまし

た⁽¹⁴⁾。本特集においては、モバイルセンシングとメディア処理、機械学習を組み合わせたパーソナル心臓モデリングの取り組みについて紹介しています⁽¹⁵⁾。

昨今の科学技術の進歩とコモディティ化はますます加速しています。さらに、新型コロナウイルス感染症のパンデミックや、地球規模の気候変動や自然災害の増大、国際紛争など社会情勢の変化により、日々の生活様式や価値観が激変しています。ますます複雑化・多様化する社会課題の解決には、分野内での研究を追求するだけでは限界があり、周辺分野の研究成果とのシナジーにより新たな価値を創造する研究や、異分野と融合した新たな学際分野を切り拓くことの重要性はますます高まっています。

一方、基礎研究においては、これまでとは抜本的に異なる、より長期的視野に立った研究アプローチも必要です。このような問題意識の下、現代数学の基礎理論を研究する組織「基礎数学研究センタ」を2021年10月にCS研内に設立しました⁽¹⁶⁾。基礎数学研究センタにおける取り組みの詳細については、本特集記事『対称性に基づく解析学と幾何学による数論と量子相互作用』⁽¹⁷⁾を参照ください。今後、IOWN (Innovative Optical and Wireless Network) 構想実現に向けて、さまざまな分野で生まれる課題について、最新の現代数学手法を駆使しながらアプローチして解

決していくとともに、多様な価値観を持つ人と、社会、環境との調和と共生をもたらし、未来へと持続させていくコミュニケーション科学研究に取り組んでいきます。

■参考文献

- (1) <https://www.kecl.ntt.co.jp/30th/>
- (2) <https://group.ntt.jp/newsrelease/2021/06/25/210625a.html>
- (3) Delcroix・落合・佐藤・大石・木下・中谷・荒木：“聞きたい人の声に耳を傾けるAI——深層学習に基づく音声の選択的聴取技術 SpeakerBeam,” NTT技術ジャーナル, Vol.33, No.7, pp.21-25, 2021.
- (4) 中谷・池下・加茂・木下・荒木・澤田：“あなたの声を「すぐそば」品質で聴くAI——遠くからでも近接マイク品質で混ざった音を聞き分ける革新的音響処理技術,” NTT技術ジャーナル, Vol.34, No.8, pp.36-40, 2022.
- (5) 杉山・水上・有本・成松・千葉・中嶋：“機械が会話のパートナーになる日——大規模深層学習で拓く雑談対話システムの新境地,” NTT技術ジャーナル, Vol.33, No.7, pp.14-17, 2021.
- (6) <https://group.ntt.jp/topics/2021/11/12/ai-seika.html>
- (7) <https://group.ntt.jp/topics/2021/09/30/transformer.html>
- (8) <https://group.ntt.jp/newsrelease/2021/06/01/210601b.html>
- (9) 丹羽：“デジタルツインでモビリティ群を賢く制御する——分散深層学習がもたらす未来の可能性,” NTT技術ジャーナル, Vol.34, No.8, pp.45-49, 2022.
- (10) <https://illusion-forum.ilab.ntt.co.jp/>
- (11) <https://group.ntt.jp/newsrelease/2021/10/19/211019a.html>
- (12) <https://group.ntt.jp/newsrelease/2020/06/03/200603a.html>
- (13) https://www.kecl.ntt.co.jp/icl/lirg/resources/goitokusei/vocabulary_test/php/login.php
- (14) https://www.rd.ntt/research/JN202112_16471.html
- (15) 柏野・渋谷・塚田：“モバイルセンシングを活用したパーソナル心臓モデリング,” NTT技術ジャーナル, Vol.34, No.8, pp.41-44, 2022.
- (16) <https://group.ntt.jp/newsrelease/2021/10/01/211001a.html>
- (17) 若山：“対称性に基づく解析学と幾何学による数論と量子相互作用,” NTT技術ジャーナル, Vol.34, No.8, pp.50-56, 2022.



納谷 太

CS研は、人・社会・地球環境に自然な調和をもたらす「ここまで伝わる」コミュニケーションの実現に向けて、広く学際的な基礎研究に取り組むとともに、パートナーの皆様とのコラボレーションにより新たな価値創造に貢献していきます。

◆問い合わせ先

NTTコミュニケーション科学基礎研究所
企画担当
TEL 0774-93-5020
FAX 0774-93-5260
E-mail cs-liason-ml@hco.ntt.co.jp