変化する現在（いま）に適応し、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～

Communication science that adapts to the changing present and creates a sustainable future
– Aiming to create technology that brings harmony and symbiosis with people, society, and the environment –

NTT コミュニケーション科学基礎研究所 所長
納谷 太
Futoshi Naya

はじめに

CS研は1991年7月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、2つの研究グループからスタートしたCS研は、現在、京阪奈と厚木の2拠点において総勢約150名の所員が集う組織となりました。設立から現在に至るまで、さまざまな形で支えていただきました皆さまにこの場を借りて御礼申し上げます。また、この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、創立30周年記念ウェブサイトを開設しました[1]。設立当初からの主な研究成果をまとめていますので、是非ご覧いただければ幸いです。

この30年間の研究テーマの変遷を振り返りますと、人と人とのコミュニケーションを理解することを中心として始まったCS研の研究は、見る・聞く・話すといった人と同等の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・データ科学などの「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、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。これらの研究は、それぞれの年代における技術的・社会的な背景を反映して変化し発展しながらも、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継続されてきました。以下では、それぞれにおける最近の研究事例のいくつかをご紹介します。

人の能力に迫り凌駕する

深層学習技術の登場で、見る・聞く・話すといったいわゆるメディア処理に関するAI技術は格段に進歩しました。しかし、一般的に深層学習では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においても、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などにより、柔軟かつ高度な推論を行うことができます。

例えば、写真は三次元の空間情報を二次元画像に変換したものですが、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をある程度推測できます。通常、このような能力をコンピュータで学習するには、三次元情報を深度センサーやステレオカメラなどの特殊な装置で計測した入力データと、写真として撮影された二次元画像としての出力データをペアとして大量に取得し用意せねばなりませんが、これには膨大なコストを要します。CS研では、このようなデータ取得上の課題を解決するため、カ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しい深層学習技術「Aperture Rendering GAN」[2]を提案しました。本技術により、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可能です。

人間は、複数の人が同時に話している場合においても、聞きたい人の声に集中して聞き分ける能力（選択的聴取）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、最近では、音声に加えて話者の映像を組み合わせることにより、人間のように複数の手が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がかりとし、逆に唇の映像が得られない場合には音声を主な手がかりとすることにより、頑健かつ高精度な話者の特定が可能になります。本研究は、音声だけでなく、消防車の音や犬の鳴き声など、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。

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[4]。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[5]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[6]。今年のオープンハウスでは、車を運転中の車窓から見た画像や周辺情報を話題とした対話システムを展示しています。今後は、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、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、人間が持つ機微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[7]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ました。このように、人の置かれている状況を考慮し、人の求める性能により近い出力を行うAIのニーズは高まっています。今年の展示においても、被災者の順次帰宅を考慮した避難所計画の最適化や、指定した語句を必ず使うニューラル機械翻訳などの研究を紹介しています。

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上で、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[8]も公開しています。今年の展示においても、「指先で感じる“こっち”」「ノビのある速球は錯覚？」「壁が動くと速く歩く？」などは、引っ張られる感覚や、視覚および運動との相互作用が引き起こす興味深い錯覚現象から脳の多様な潜在機能を探る研究です。

また、2021年10月には、静岡県立総合病院と人工内耳使用者の音声・言語認知の共同研究を開始しました[9]。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獲得ができることが実証されていますが、このような脳における音声知覚や言語発達のメカニズムは未だ明らかになっていません。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着手しています。

言語獲得に関する研究では、CS研では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年には新しく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ました[10]。単語親密度データベースを用いることにより、数十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[11]も公開しています。

人・社会・環境との調和と共生をもたらす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継続・発展してきました。これまで、「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、複数のメディアにまたがったクロスモーダル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。さらに昨年では、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「投手シミュレータマシン」が、東京オリンピック2020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[12]。

昨今の技術の進歩と普及スピードは凄まじく、分野内での研究を研ぎ澄ますだけではなく、周辺分野の研究成果を組み合わせて新たな価値を創造する研究や、異分野と融合させた新たな学際分野を切り拓く気運は高まっています。一方で、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、国際紛争の勃発などの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る時代です。ますます複雑化・多様化する社会課題の解決には、これ

までとは抜本的に異なる長期的視野に立った研究アプローチが必要です。このような問題意識のもと、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[13]。今後、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まいりたいと思います。

納谷　太：1992年慶応義塾大学理工学部電気工学科卒、1994年同大学院理工学研究科計算機科学専攻修士課程修了。同年、日本電信電話株式会社入社。2003年より2009年までATRメディア情報科学研究所、同知識科学研究所出向。コミュニケーションロボット、センサネットワークによる実世界センシング、人流の時空間予測・最適誘導技術、テーラーメイド学習支援技術等の研究に従事。博士（工学）。2022年4月より現職。

●参考文献

[1] NTT コミュニケーション科学基礎研究所 創立30周年記念ウェブサイト　https://www.kecl.ntt.co.jp/30th/
[2] 世界初、一般的な写真群から未知の奥行きとボケ効果の学習を実現～限られたデータから三次元世界の理解が可能なAIの実現に向けて前進～
https://group.ntt/jp/newsrelease/2021/06/25/210625a.html
[3] 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, Marc Delcroix, 落合翼, 佐藤宏, 大石康智, 木下慶介, 中谷智広, 荒木章子, NTT技術ジャーナル 2021.7
[4] 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, 杉山弘晃, 水上雅博, 有本庸浩, 成松宏美, 千葉祐弥, 中嶋秀治, NTT技術ジャーナル 2021.7
[5]「なりきりAI京町セイカ」を用いた実証実験を精華町役場で開始　https://group.ntt/jp/topics/2021/11/12/ai_seika.html
[6] 日本語最大規模のTransformer対話モデルを無償公開　https://group.ntt/jp/topics/2021/09/30/transformer.html
[7] 因果関係に基づく公平・高精度な機械学習予測を実現～どんな予測が差別的かを指定しながら、人を対象とした効果的な予測が可能に～
https://group.ntt/jp/newsrelease/2021/06/01/210601b.html
[8] 錯視と錯聴を体験 イリュージョンフォーラム　https://illusion-forum.ilab.ntt.co.jp/
[9] 難聴者・人工内耳装用者の音声・言語認知の共同研究に関する協定の締結について
～医療と脳科学の両面で難聴のメカニズム解明に貢献することを期待～
https://group.ntt/jp/newsrelease/2021/10/19/211019a.html
[10] 令和版単語親密度データベースの構築と語彙数推定テストの作成～語彙数推定から学習支援へ～
https://group.ntt/jp/newsrelease/2020/06/03/200603a.html
[11] 令和版語彙数推定テスト　https://www.kecl.ntt.co.jp/icl/lirg/resources/goitokusei/vocabulary_test/php/login.php
[12] 女子ソフトボール × スポーツ脳科学　https://www.rd.ntt/research/JN202112_16471.html
[13] NTT基礎数学研究センタを設立～未知なる量子コンピューティングの原理解明などに挑戦する研究を加速～
https://group.ntt/jp/newsrelease/2021/10/01/211001a.html

MEMO