

## 1 インタビュー

# 「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（以下、CS研）では、ICT社会に変革をもたらすコミュニケーション科学にかかわる基礎研究を推進している。「人に迫る」、「人を究める」、「人に寄り添う」に挑戦し続けることで革新技術を創出し、「ここまで伝わる」Heart Touching なコミュニケーションの実現をミッションとするCS研の山田武士所長にお話を伺った。

### コンピュータと人間の関係を追求するコミュニケーション科学領域の基礎研究を推進

—NTT民営化、CS研の成り立ちについてお聞かせください。

山田 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1985年に発足しました。その当時、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、むしろ使用する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とぬくもりがありました。

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。



NTTコミュニケーション科学基礎研究所  
所長 山田 武士氏

### 「あなたを・もっと・知りたくて」コミュニケーションの本質とは？

—CS研の主要ミッションについてお聞かせください。

山田 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、人と人のみならず、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。まさに「あなたを・もっと・知りたくて」を究



図1 CS研のミッションと研究領域

める基礎研究です。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。

主な研究領域は、「人間科学」「メディア処理」「多様脳科学」「データと機械学習」です(図1)。所内組織は、3つの研究部(「メディア情報研究部」「協創情報研究部」「人間情報研究部」)、守谷特別研究室、機械学習・データ科学センタ、上田特別研究室、多様脳特別研究室からなっています。

——最近の代表的な研究開発の概要をご紹介します。

**山田** 今年のオープンハウス2020は、コロナウイルス感染症対策のためにWeb公開での対応となりました。これに沿って、CS研の研究の一部を紹介いたします。

CS研が運営する「イリュージョンフォーラム」(<http://www.kecl.ntt.co.jp/illusionforum/>)で紹介している錯覚の一つに「モザイク音声」というものがあり、体験していただくと人間は非常に荒い情報だけでも音声を聞き取ることができることを実感できます。実際このような音を聞き取る能力というのは、人間を含む動物が長い進化の過程を経て獲得したものと考えられます。そこで我々は音を認識するようにDNN(深層ニューラルネットワーク)を訓練すると脳における音の表現と類似した表現をDNNが獲得することを発見しました。これは動物の脳の進化の過程でも同様のことが起こった結果、音認識に適した性質を獲得してきた可能性を示唆するものです。

我々は非常に騒がしいパーティーで複数の人が会話している状況でも、特定の人の声を聞き分けることができます。これは人間の優れた能力の

一つで、選択的聴取と言われていません。従来はこの選択的聴取はコンピュータにとっては苦手だと考えられていました。我々は、独自のDNNに基づいてこの音声の選択的聴取を実現するSpeakerBeamという技術を考案しています。これはDNNに目的話者の音声の特徴を教えるだけで、目的話者の音声を取り出すことができる技術です。しかし音声情報のみでは声の性質が似た話者同士を聞き分けることは困難です。そこで最近では、これに唇の動きなどの手がかりを加えてマルチモーダルに、すなわち画像情報も使って聞き分けるようにしています。

——音声変換技術やその他の技術の研究もご紹介します。

**山田** 例えばAさんが話した音声を話の内容は保持したまま、あたかもBさんのような声に変換するという音声変換にも取り組んでいます。DNNを駆使したこの技術によって声帯障害を持つ方の電気式人工喉頭の声をより自然な音声に変換できる可能性があります。また、クロスモーダルな音声変換を使うと、効果音のようなものを入力してその音の説明文を生成したり、顔画像を入力して、この顔の人が話すであろう声質を再現したりできるようになります。これらを進めると発声や聴覚機能を拡張した新たなコミュニケーション形態を創出できると考えています。

また我々は、幼児の語彙発達の研究に取り組んでいます。具体的には、3,000名規模の語彙チェックリスト調査から、子どもがいつどんな言葉を習得するかというデータベースを構築しました。赤ちゃんの脳の健全な成長には絵本の読み聞かせなど、保護者の働きかけが重要と言われて

います。我々はこれら語彙獲得研究の知見を活かし、子供の成長に合わせた絵本検索システム「ぴたりえ」や子ども一人ひとりの語彙発達に合わせたパーソナル知育絵本の作成などを行ってきました。

「ロボットは東大に入れるか?」という国立情報学研究所のプロジェクトにも関わっており、英語の試験を担当しています。2019年大学入試センタ試験にて英語科目でAIが185点を獲得するまでに到達しています。

他には、なりきりAI「京町セイカ」、混雑ゲームの均衡計算、情動反応の同期なども研究しています。これらについてはこの後の特集記事で詳細にご紹介します。

多様脳科学ではeスポーツにおける心拍数の同期測定やラグビースタムのハーモニー分析とトレーニングへの活用なども行っています。また、人間科学では錯視についての研究も進めており、これまでも「変幻灯」という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技術、「Hidden Stereo」「浮像(うくぞう)」「Danswing Paper」などの技術を創出しています。

## これからのコミュニケーション技術は どうなるのか?

——最後に今後の抱負をお聞かせください。

**山田** 今後の技術発展を見据えて、コミュニケーションはどうあるべきか?に興味があります。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はないでしょうか?

——本日はありがとうございました。