

特集

AIと脳科学で あなたをもっと知る ——人に迫り人を究める コミュニケーション科学

人工知能



触覚



深層学習

量子コンピュータ



言語彙テスト

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」ための革新技术の創出と、「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る。本特集ではCS研のさまざまな取り組みについて紹介する。

Communicat

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

CS研の「あなたを・もっと・知りたくて」を究める基礎研究について、その取り組みの一部を紹介する。

12

知覚心理学で探る触覚の仕組み

人間の触覚系が安定した知覚を生み出すために情報を取捨選択する仕組みを、「錯覚」に注目して解明した最近の研究成果を紹介する。

18

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

人間の身体動作を支える無意識の感覚・運動プロセスの1つ「伸張反射」の概要と、その脳内での情報処理に関する最近の研究成果を紹介する。

23

あなたの声はどんな声？どんな声でしゃべりたい？

近年著しく発展し続ける深層学習を活用して音声を変換する際の従来技術の課題と、それを解決する研究所での取り組みを紹介する。

29

量子情報処理における量子的間接制御の可能性

量子情報処理を実現する手段として量子的間接制御が有用な性質を持っていることを明らかにした最近の研究成果を紹介する。

34

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

テキストの難しさを自動的に推定するための難易度の推定方法と、読み手側がもつ知識量（語彙数）の推定方法についての取り組みを紹介する。

39

主役登場

藤田 早苗（NTTコミュニケーション科学基礎研究所）
私が欲しいと思うものは他の人も欲しいはず

45

ion Science

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

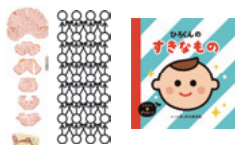
するのでしょうか。人と人との物理的な距離が遠くならざるを得ない時代だからこそ、コミュニケーションの本質が何かを究めることはなおさら重要です。

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



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



心まで伝わる
コミュニケーションの実現

「あなたを・もっと・知りたくて」
を究める基礎研究



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し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、中枢の大脑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、不慣れた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、株式会社ダウ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にぴったります。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であります。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「こころまで伝わる」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「こころまで伝わる」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

知覚心理学で探る触覚の仕組み

指先は、机の上に落ちた髪の毛や、金属の磨き度合いなど、かなり細かな違いを区別できる素晴らしいセンサです。私たちのグループは知覚心理学の手法を使って触覚の仕組みを研究しています。本稿では「錯覚」を通して、触覚システムがどのように情報を取捨選択し、安定した知覚を形成しているのかを調べた研究を紹介します。こうした研究は、効率的な情報提示技術につながるものが期待されています。

くろき
黒木

しのぶ
忍

NTTコミュニケーション科学基礎研究所

触覚 × 知覚心理学

手は道具でありセンサです。手指を巧みに使うことで、人間は上手に進化を遂げてきたと思います。道具をうまく使うには速くて正しい計測と処理が有効です。私たちはこの、皮膚を通じた情報処理に興味を持っています。「手に取るように分かる」「肌で感じる」など、触覚にかかわる比喩表現には、ものごとを深く理解する様子をイメージさせるものが多くあります。ところが、触覚が実際に何をやっているのか、その情報処理に関する研究は意外なほど疎で、その処理のメカニズムはあまり良く分かっていません。自己と世界の境界から得られる感覚であるため、その接触面で起きる現象を直接計測したり、人工的に再現したりすることが難しいなど、技術的な課題もありました。近年の技術進歩に伴い、視覚や聴覚の研究を追いかけるかたちで、少しずつ触覚についても解明が進んでいます。

知覚心理学とは、人間をシステムとして扱い、その内部処理を同定する学問です。人間に対して、何か画像や音のような入力を加え

たときに、出力としてどのような知覚が得られるのかを調べます。例えば、明かりをゆっくり点けたり消したりすると明るさの変化を知覚することができますが、切り替え速度を上げていくと明滅は徐々に見えにくくなり、蛍光灯ほど高速になると変化は検出できません。このように、少しずつ入力を変化させながら、私たちの知覚に影響を及ぼす物理量や、その範囲を調べていきます。

本稿では、触覚の仕組みや特徴について考える手掛かりとなる「錯覚」を2つ紹介します。ここで言っている錯覚というのは、例えば、物理的には全く同じ入力に対して、異なるものを知覚してしまう、あるいは物理的に異なるものに対して同じものを知覚してしまう、といった現象です。錯覚を通して、脳がどのような情報に感度を持ち、またどのような情報にはあまり感度を持たないのか、を調べることができます。この試みは、効率的な情報提示方法を考えるうえでも役立ちます。

2つの振動が混ざってしまう 周波数の錯覚

モノの表面に触れると、指先の皮膚に振動が発生します。その振動を通じて、さらさら、ざらざら、といった質感を感じることができます。これらは皮膚の中にある受容器、触覚のセンサで符号化された、振動の周波数の情報に基づいた知覚です。網膜に複数のセンサがあり、それぞれ異なる色に対して感度を持つように、皮膚にあるセンサにもいくつかの種類があります。それぞれのセンサは、ある種類は小粒で皮膚表面に密集し、また別の大きなものは皮膚深くにぽつぽつと点在するなど、異なるかたちと分布を持っています。その結果、ゆっくりした変形に応答するセンサや、着信バイブレーションのような高周波の振動に良く応答するセンサといった具合に、種類ごとに個性が生まれます（図1）。では、周波数というものはこれら異なるセンサの信号をどのように使って符号化されているので

しょうか。

触覚だけでなく、視覚や聴覚においても、時間周波数の異なる入力はその違いを知覚することができます。例えば目の場合、赤色の光（波長が長い）と緑色の光（波長がやや短い）を同時に、近い場所に提示すると、黄色の光（赤と緑のおよそ中間の波長）を感じます。赤と緑、それぞれを感じることはできません。有名な光の三原色です。一方で音の場合、例えばドの音とミの音を聞いたときに、中間のレの音が聞こえるということはなく、和音として両方聞こえる、分解できると思います。手はどちらに似た処理をしているのでしょうか。

私たちは、振動の三原色のような現象が観察できるか、複数の振動に触れた場合に中間の振動周波数を感じるかどうかについて、実験を通して検討しました。ここで、単純に皮膚表面に2つの周波数を合算した合成振動を加えてしまうと、皮膚が弾性体であるため、センサに振動が届くころには物理的に振動が

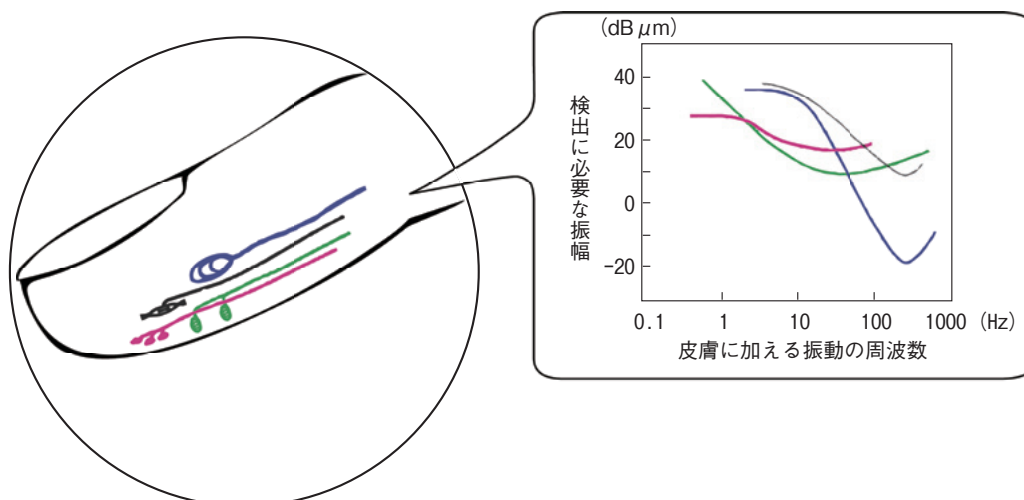


図1 指先皮膚に存在する複数のセンサと、それらにつながる神経の応答特性

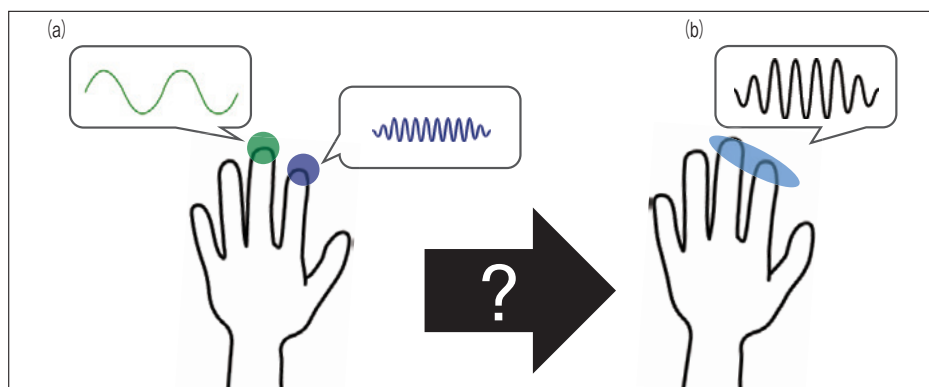


図2 2つの振動の周波数が混ざってしまう錯覚

減衰してしまうおそれがあります。そこで、人差し指と中指、2本の異なる指に対して、異なる周波数を持つ2つの振動を1つずつ、同時に加えることで、赤色と緑色を同時に見ているような状況を疑似的に再現しました。実験では、2つの指に異なる周波数を持つ正弦波振動を加える「合成振動」について、その周波数を、2つの指に同じ正弦波振動を加える「比較振動」の周波数と比較してもらいました（図2）。すると、合成振動で提示した2つの振動の中間の周波数を持つ振動を提示したときに、合成振動と同じくらいの周波数であるように感じられるということが分かりました。ここだけみると、光の三原色と同じような結果になっています。また、この現象は、合成振動を片手の2本の指に加えた場合だけでなく、右手と左手の指に加えた場合にも観察されることが分かりました。右手と左手という明らかに異なる部位に、全く異なる周波数の振動を加えた場合にも、私たちはまるでその2つの振動が同じ周波数を持つ入力であるかのように、混ぜて感じてしまうようです。

ただし、合成振動は完全に混ざってしまうわけではない、ということも実験を通して分かってきました。2つの指で異なる周波数を持つ振動「合成振動」に触れた場合、右の指と左の指に何らかの（例えば強さなどの）違いを感じます。また、周波数では同じくらいの正弦波振動と比べても、合成振動はノイズに感じます。触覚の場合、視覚と違って、合成振動と普通の振動で、知覚的等価性が担保されないのです。

まとめると、触覚では周波数の異なる複数の振動に触れたときに、複数の振動があることが分かるので、目とは異なる符号化を行っていることが分かります。しかし、2つの周波数を別々に感じることはできないので、耳とも異なっています。触覚は、視覚と聴覚の間のような特徴を持っているのではないかと考えられます⁽¹⁾。

違って見えるのに触ると似てしまう テクスチャの錯覚

私たちの身の周りに存在する物体は、多様な表面の凹凸パターン、テクスチャを持って

います。思いどおりのテクスチャを人工的に設計することは技術的に難しかったため、テクスチャの研究は簡単な実験刺激を作成するか、世の中にすでに存在する素材をそのまま使うことで行われてきました。私たちのグループでは、単純な実験刺激と複雑な現実世界の間をつなぐ試みとして、レーザーカッターや3Dプリンタを使い、複雑な空間パターンを持つ表面テクスチャを作成して、研究を進めています。

テクスチャの設計には画像をハイトマップとして利用しました(図2(a))。白い所が凸、黒い所が凹み、彫りが深いところを表します。画像を用いることで、表面凹凸の持つ空間周波数や、その帯域などの統計量を、簡単に操作することができます。実験では、各画像を4 cm × 4 cm、彫りの深さ最大で2 mmの試料片として3D印刷して、表面テクスチャを作成し、それぞれのテクスチャを区別できるかを調べました。図3(a)に示されている5枚の画像は、紙やすりの番手などで表されるような中心周波数、比較的単純な統計量を変

調してあります。これらの画像に基づいて3D印刷した試料片は、右に行くほど中心周波数が高く、ぼこぼこした表面からザラザラした表面になります。実験でこの表面の触り分けを行うと、高い成績が得られます。良く知られている現象ですが、人間は中心周波数の変調には非常に感度が高いのです。では、全く違う画像に基づく試料片の弁別はどうでしょうか。図3(b)の5枚は、画像データベースから取ってきた自然画像です。輝度値の平均と分散をそろえてありますが、白黒の空間的な配置に違いがあるため、目で見ると簡単にパターンの違いを見分けることができます。しかし、手で触るとあまり違いを触り分けることができません。特に、石・サンゴ・葉っぱの3枚(図3(b)の紺枠で囲んだ3枚)については、それぞれ見た目が全く異なるにもかかわらず、区別ができませんでした。テクスチャの弁別という課題においては、目と手で、入ってくる情報の計算の仕方が違うらしい、ということがこの錯覚から分かります。

では、目と手の情報処理はどのように異な

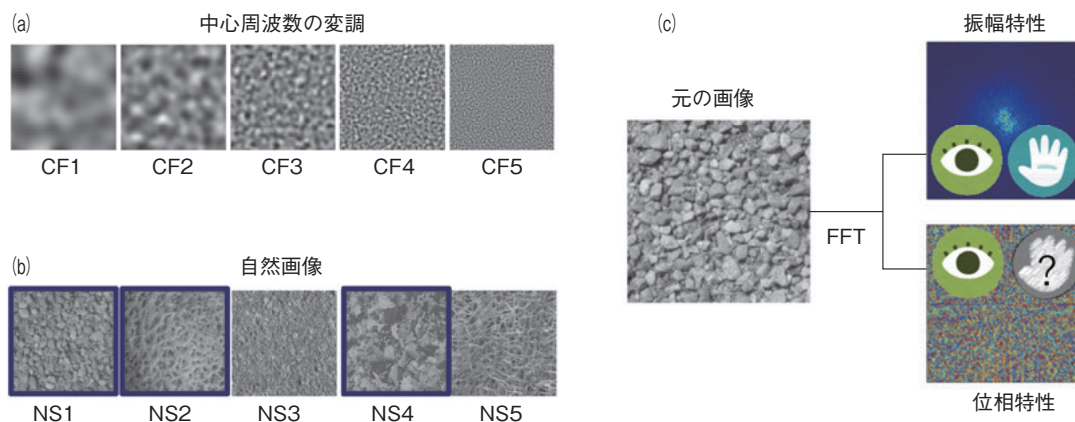


図3 違って見えるのに触ると似てしまうテクスチャの錯覚

るのでしょうか。私たちは、テクスチャ表面の設計図となった画像について解析を行いました。自然画像というのは、空間周波数が上がるにつれて振幅が下がるという特徴的な振幅特性を持っています。図3(b)の5枚の画像に対してそれぞれフーリエ変換をかけて、振幅（濃淡の強さの情報）と位相（濃淡の分布情報）という異なる統計量に分けてから見比べると、5枚の振幅はよく似ている一方で、5枚の位相は全く異なっています。目は、画像をフーリエで分解した先の、振幅と位相（の一部）を両方情報処理しています。一方で、手では主に、振幅の特性のほうを情報処理して触り心地を計算し、位相特性のほうはあまり使っていないのではないかと、ということが考えられます（図3(c)）。この性質を利用することで、違って見えるのに触ると似てしまうテクスチャを自由に作成することができるようになりました⁽²⁾。位相の算出には振幅の算出よりも計算コストがかかります。触覚は、位相の情報処理にあまりリソースを割かず、どちらかというと、振幅特性の細かな違いに感度を持つように進化をしたのではないかと仮説を立て、引き続き研究を行っています。

紹介してきたように、私たちのグループでは知覚心理学の手法を使って触覚の仕組みを研究しています。五感の研究はもともと、視覚・聴覚を対象にしたものが多く、触覚の研究はあまりありませんでした。研究を行うための道具が進歩してきたこともあり、最近やっと手の研究も増えてきています。五感には似たような処理もたくさんありますが、やはり、別々のかたちをして進化をしてきた感覚器なので、それぞれ、異なる情報処理を行っている可能性があります。視覚には視覚の、聴覚

には聴覚の、そして触覚には触覚の処理、知覚のルールがあります。こうした知覚のルールに関する理解を積み重ねていくことが、触覚の理解につながると同時に、将来的な触覚情報提示の可能性を啓くと期待しています。

■参考文献

- (1) S. Kuroki, J. Watanabe, and S. Nishida: "Integration of vibrotactile frequency information beyond the mechanoreceptor channel and somatotopy," *Scientific Reports*, Vol. 7, No. 2758, 2017.
- (2) S. Kuroki, M. Sawayama, and S. Nishida: "Haptic metameric textures," *bioRxiv*, doi: <https://doi.org/10.1101/653550>, 2019.



黒木 忍

私たちは自身のセンサとしての特性も含めたうえで、外界を知覚しています。外界の物理情報と内的な知覚状態の関係性について、機能的なモデルを構築することで、人間理解への科学的な貢献と、産業分野への工学的な貢献をめざしています。

◆問い合わせ先

NTTコミュニケーション科学基礎研究所
人間情報研究部
TEL 0774-93-5020
FAX 0774-93-5026
E-mail cs-liaison-ml@hco.ntt.co.jp



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

観客を惹きつけるアスリートのスーパープレーから、私たちが何気なく行っている日常の動作まで、人間の身体動作は反射に代表される無意識の感覚-運動プロセスに支えられています。その1つである伸張反射が、多感覚統合による身体表象を介して調整されることが、最新の研究成果より分かってきました。このことは、反射による運動制御が、従来考えられてきたより高度な脳内情報処理を経て行われている可能性を示唆します。

いとう しょう
伊藤 翔

ごみ ひろあき
五味 裕章

NTT コミュニケーション科学基礎研究所

人間の運動を支える反射

人間の身体動作は意識に上らないさまざまな脳の仕組みによって支えられています。その1つが「反射」と呼ばれる感覚-運動プロセスで、視覚や体性感覚*などによる外界からの刺激に対して、意識を介せずに運動応答を生成します。反射による運動制御は、随意運動（意識を介して行われる運動）よりも高速な応答を引き起こすため、例えばスポーツにおいて相手の動きに対応した瞬間的な反応を可能にすると考えられます。あるいは“歩く”“立ち上がる”“物に手を伸ばす”といった、日常生活の中で何気なく行っている動作も、反射による姿勢の制御に支えられて初めて実現されるといえます。NTTコミュニケー

ション科学基礎研究所では長年、反射がどのような脳内の情報処理を経て生成され、人間の運動実行に役立っているかを調べています^{(1)~(3)}。本稿では体性感覚情報によって引き起こされる反射の1つ「伸張反射」について概要を述べ、その情報処理に関する最近の研究成果⁽⁴⁾を解説します。

伸張反射の仕組みと機能

伸張反射は筋の受動的な伸展によって生じる反射で、主に姿勢を安定に保つうえで重要な役割を果たすと考えられています（図1）。障害物との接触などによって意図しない姿勢の変化が生じると、筋の長さの変化をとらえる受容器である筋紡錘が反応し、「筋が伸ばされた」という情報を上行性の感覚信号として脳・神経中枢に伝達します。この信号は随意運動とは異なる脳部位や神経経路で処理され、伸ばされた筋を収縮させる運動応答を生成します。伸張反射応答は脊髄レベルの神経

* 体性感覚：人間の感覚機能のうち、触覚および深部感覚を合わせて称する用語。皮膚や筋、腱、関節などの受容器からの入力によって生じる感覚を指します。体性感覚があることで、人間は視覚情報が得られない状況でも自己の姿勢や運動の様子を知覚することができます。

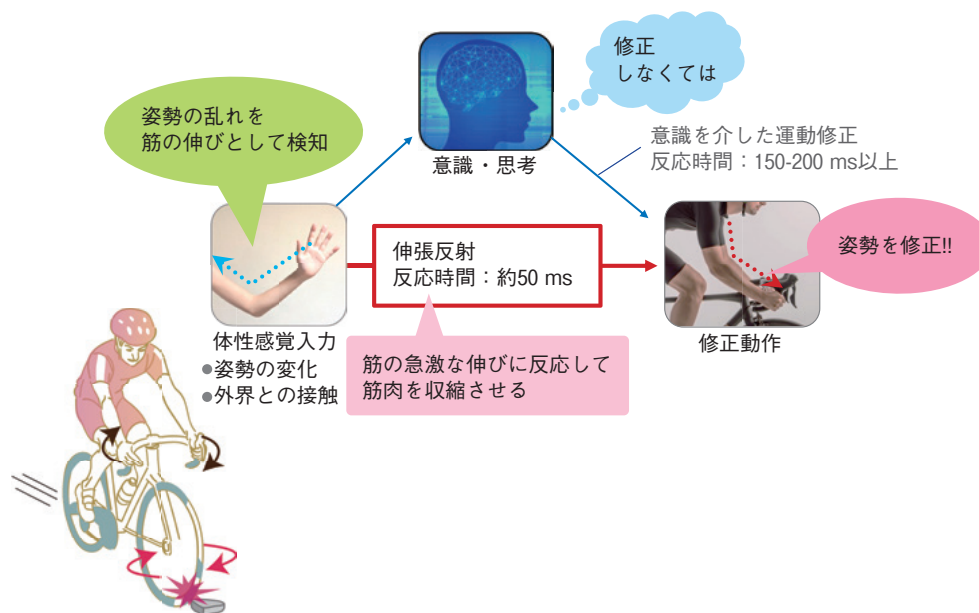


図1 伸張反射のはたらき

経路によって生じる短潜時成分と、大脳皮質運動野まで含む神経経路を経て生成される長潜時成分とを含むことが明らかになっています⁽²⁾。そのうち、比較的応答の遅い長潜時の伸張反射でさえ、刺激の入力から約50 msという極めて短い反応時間で筋活動が発生するため、反応に100～150 ms以上かかる随意運動と比較して素早く姿勢の変化を補償することができると考えられます。

伸張反射は入力刺激に対して常に一定の応答が生じるわけではなく、課題や環境の変化に依存して応答の調整がみられることが種々の先行研究で示されています^{(3), (5)}。脳はこのような反射系の調整を通して、状況に応じた柔軟な運動制御を行っていると考えられます。一方で、その調整計算のために脳内でどのような情報処理が行われているかについて

は、詳細には分かっていません。一例として、伸張反射の調整が体性感覚情報のみに基づくのか、あるいは視覚情報など他のモダリティにおける感覚情報も統合した身体表象を利用して行われているのかについてはこれまで未解明でした。本研究ではこの点を検証するため、運動実行中の自己位置を示す視覚情報を操作し、伸張反射に影響がみられるかを調べました。

反射調整における多感覚統合の関与

身体状態に応じた伸張反射の調整の仕組みについて、2つの仮説が考えられます(図2)。1つは、伸張反射は体性感覚入力によって生じる応答であることから、その調整も体性感覚に依存して行われるという考え方で、これに対してもう1つは、体性感覚情報

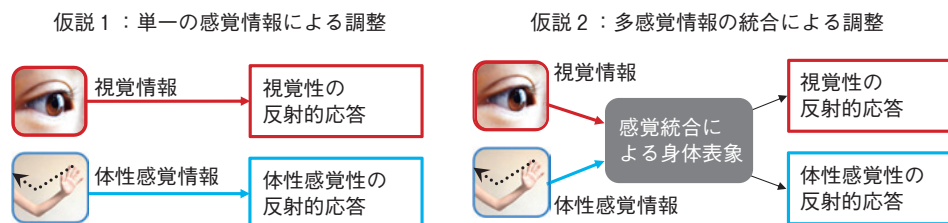
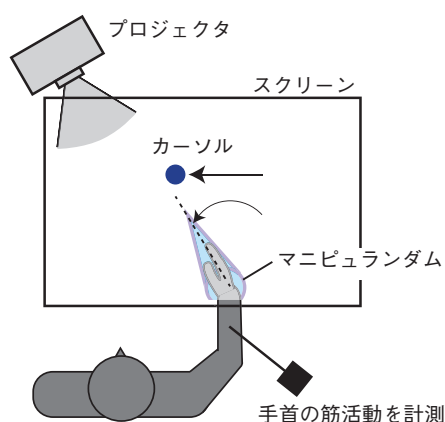
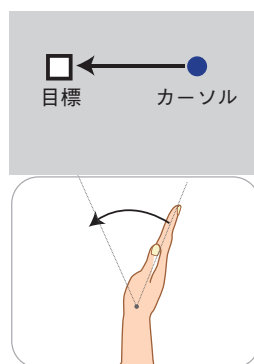


図2 反射調整の情報処理についての仮説

実験装置



視覚目標への到達運動



運動中に手首へ外力を作用

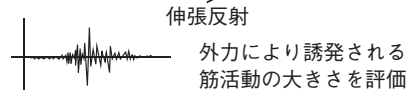
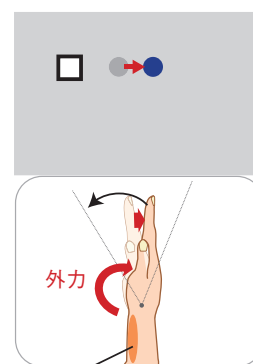


図3 実験設定

に加えて視覚情報など他のモダリティの感覚情報も統合した身体表象を介して、より精度良く身体状態を推定し、反射応答の調整に使用しているという考え方です。本研究では、これらの仮説のどちらがより確からしいかを検証するため、視覚情報を実験的に操作し、それにより伸張反射が影響を受けるかを調べました。

実験として、図3左に示すような実験装置を用いて、手首の屈曲運動による視覚目標への到達課題（図3中央）を行いました。試行

を繰り返し行う中で、一定の確率で到達運動の最中に瞬間的な外力を作用させ、手首の急激な伸展を引き起こしました（図3右）。この操作によって手首の屈筋に伸張反射を誘発することができ、その大きさを皮膚に張った電極によって計測しました。ここで実験条件として、手の位置を表すカーソルの移動方向に回転変換を加えることで、視覚による身体情報と実際の運動の間に不一致を生じさせました（図4左）。ただし、目標位置や運動開始位置も同じ回転変換により移動させること

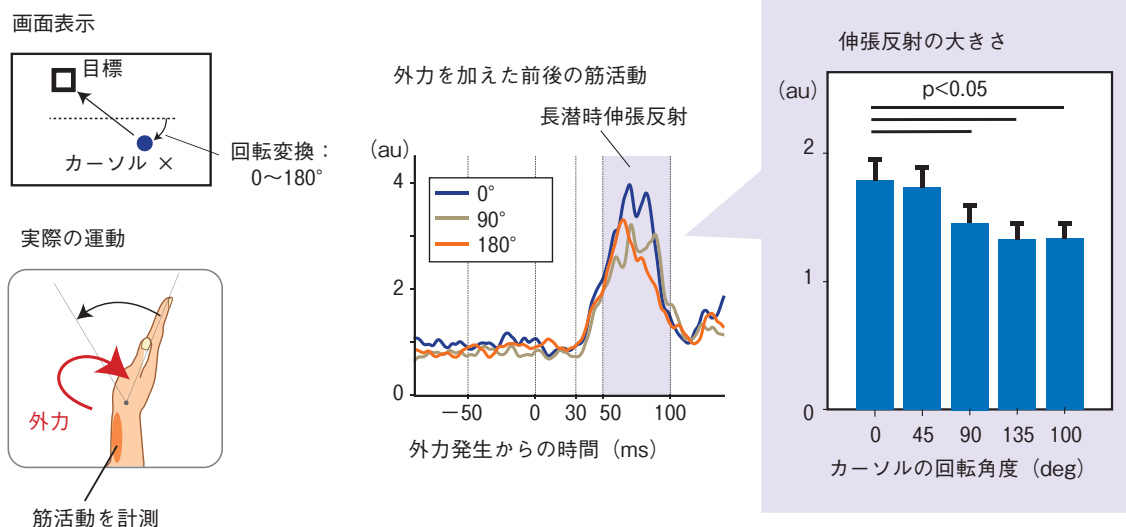


図4 視覚情報の不一致（回転変換）が伸張反射に与える影響

で、行われる運動は常に同一になるようにしました。この条件下で伸張反射応答の大きさを比較することで、視覚情報の伸張反射への影響を調べました。その結果、図4右に示すように、カーソルの回転変換が大きくなるほど、長潜時の伸張反射応答が小さくなることが分かりました。このことは、伸張反射応答の調整に視覚情報を含む多感覚統合による身体表象が関与している、という仮説を支持する結果といえます。反射系の調整の基盤となる身体表象も、随意運動と同様、複数の感覚情報を統合することによって状態推定の精度を高めていると考えられます。

身体状態推定の不確かさの影響

第一の実験では、視覚フィードバックの回転変換に応じて伸張反射応答が小さくなることを示しました。この現象の説明として、回

転変換によって運動中の身体状態推定が不確かになり、そのことが伸張反射応答の減衰につながった可能性が考えられます。この仮説を検証するため、運動中に自分の手の位置を示すカーソルを消去する実験を行いました（図5左上）。このとき身体状態を表す視覚情報に不確かさを与えるため、カーソルの提示時間を4段階に変化させ、運動への影響を調べました。その結果、図5右上に示すように、カーソルの提示時間が短いほど運動到達位置のばらつきは大きくなり、運動中の身体状態推定がより不確かになることが確認されました。一方、伸張反射応答を計測したところ、カーソルの提示時間が短いほど長潜時の伸張反射応答が小さくなることが分かりました（図5下）。この実験の結果から、運動中の身体状態の不確かさに応じて伸張反射が調整されていることが示唆されました。身体状態が曖

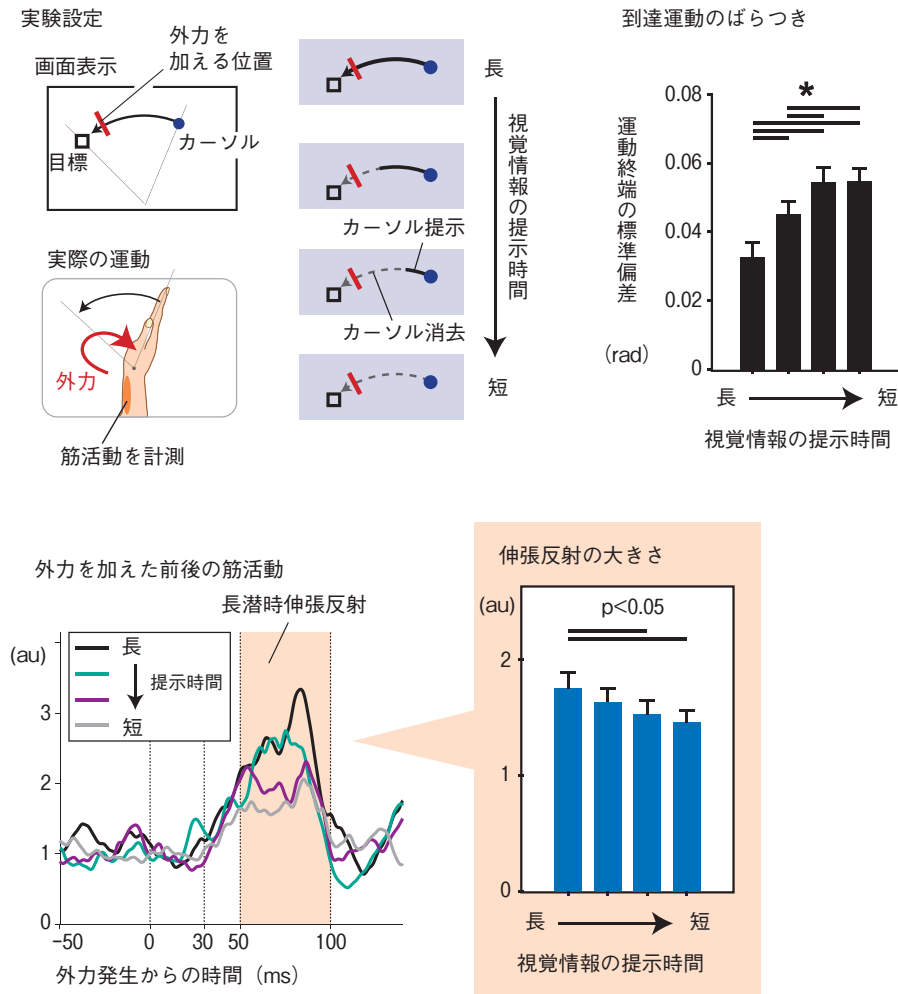


図5 身体情報の不確かさが伸張反射に与える影響

昧な状況では誤った方向に大きな運動修正を行ってしまう危険性があるため、脳は反射応答を小さくすることでリスクを回避する調整を行っていると考えられます。

まとめと今後の展望

本研究では、伸張反射が視覚による身体情報に依存して調整されることを明らかにしました。実験により、視覚目標への到達運動中

の伸張反射が①身体運動の視覚フィードバックと実際の運動の不一致、②身体運動の視覚フィードバックの消去、という条件下で通常時より小さくなることを示しました。これらの結果から、伸張反射の調整には体性感覚情報だけでなく、「視覚情報も統合した身体表象」が利用されていると考えられます。さらに、視覚情報の提示時間に応じて変化する反射応答と運動到達位置のばらつきが逆相関す

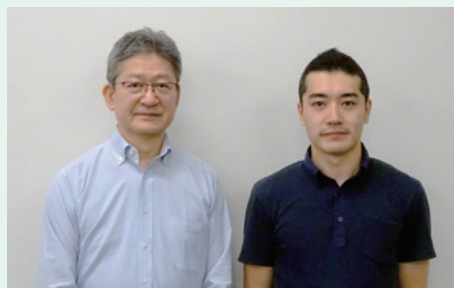
ることを見出しました。このことから、運動中の身体状態推定がどのくらい不確かかということによって伸張反射の調整が行われている、ということが示唆されます。これらの発見は、反射応答の調整計算が、従来考えられていたより複雑な脳内情報処理を経て行われていることを示唆しています。

今後、本研究で得られた知見を基に、反射系調節のために行われている脳内情報処理のさらなる解明をめざします。特に、調整計算にかかわる神経基盤を明らかにすることや、情報処理の理論的側面を詳細に調べることがこれからの課題です。このような検討を通して人間の身体制御の仕組みをより深く理解していくことで、将来的には、アスリートの能力の解析や、科学的なトレーニング方法の開発などに発展していくことが期待されます。また、本成果によって、人間の運動実行において多感覚情報が統合・利用される脳内情報処理過程の一端が明らかになりました。運動時の感覚統合メカニズムは、身体動作を伴うインタフェースのデザインを検討するうえで重要であり、将来的にはVR（Virtual Reality）システムの設計開発などに応用できるよう、知見を積み重ねていきます。

本研究の一部は、平成28～令和2年度文部科学省科学研究費補助金新学術領域研究「人工知能と脳科学の対照と融合—潜在的運動における学習適応メカニズムの解明と計算モデル構築（JP16H06566）」の助成を受けて行われました。

■参考文献

- (1) N. Saijo, I. Murakami, S. Nishida, and H. Gomi: "Large-field visual motion directly induces an involuntary rapid manual following response," *Journal of Neuroscience*, Vol. 25, No. 20, pp. 4941–4951, 2005.
- (2) T. Kimura, P. Haggard, and H. Gomi: "Transcranial magnetic stimulation over sensorimotor cortex disrupts anticipatory reflex gain modulation for skilled action," *Journal of Neuroscience*, Vol. 26, No. 36, pp. 9272–9281, 2006.
- (3) T. Kimura and H. Gomi: "Temporal development of anticipatory reflex modulation to dynamical interactions during arm movement," *Journal of Neurophysiology*, Vol. 102, No. 4, pp. 2220–2231, 2009.
- (4) S. Ito and H. Gomi: "Visually-updated hand state estimates modulate the proprioceptive reflex independently of motor task requirements," *Elife*, Vol. 9, e52380, 2020.
- (5) J.A. Pruszynski and S.H. Scott: "Optimal feedback control and the long-latency stretch response," *Exp. Brain Res.*, Vol. 218, No. 3, pp. 341–359, 2012.



(左から) 五味 裕章 / 伊藤 翔

無意識の運動応答は一見、単純にみえますが、調べるといろいろと興味深い点が見えてきます。反射をはじめとする運動の仕組みを通して、私たちの脳が行っている巧みな情報処理を探ってみたいと思います。

◆問い合わせ先

NTTコミュニケーション科学基礎研究所
人間情報研究部
TEL 0774-93-5020
FAX 0774-93-5026
E-mail cs-liaison-ml@hco.ntt.co.jp

あなたの声はどんな声？ どんな声でしゃべりたい？

人と人とのコミュニケーションには、物理的・能力的・心理的な状態に起因するさまざまなかたちの制約が存在します。私たちは、ある音声を異なる音声や望まれる音声へと変換する技術の研究を通じてこのような制約を取り除き、あらゆる人が不自由なく快適にコミュニケーションを行える環境を実現することをめざします。本稿では、「近年著しく発展し続けている深層学習を活用することによって音声をどのように変換できるのか？」を題材に、従来技術の課題と私たちの取り組みを紹介します。

たなか	こう	かねこ	たくひろ
田中	宏	金子	卓弘
ほうじょう	のぶかつ	かめおか	ひろかず
北条	伸克	亀岡	弘和

NTTコミュニケーション科学基礎研究所

非言語情報を「変換」する

音声は、言語情報だけでなく話者性などの非言語情報も伝達できるという大きな特徴を有し、利便性に特に優れたコミュニケーション媒体であるため、人々がお互いにコミュニケーションを取るうえで基本的なツールの1つとなっています。発話することで、自分・相手の意図や感情を、伝える・理解することができるため、音声の特徴（例えば、抑揚や声質・リズム）をその時々で変化させることで、相手へ与える印象を変えることもできます。

しかしながら、一個人の生成できる音声の表現力は身体的・能力的・心理的制約により制限されてしまいます。この制約を超え、発話者が所望の音声で思いのままに表現できるよう能力の拡張を行う技術が音声変換です。その適応先は、話者性の変換や発声障がい者補助、感情などの発話スタイル変換、語学学習のための発音・アクセント変換など、多岐にわたります。これらの利用シーンに応じて、変換したい音声特徴・学習データ・リアルタイム性に関する要件など、さまざまな前提条

件が想定されます。私たちは、高品質であること、少量データ・非パラレルデータ*で学習可能であり効率的であること、リアルタイムに音声変換が動作すること、声質だけでなく抑揚やリズムといった超分節の特徴などの柔軟な変換が可能であること、上記の4点が音声変換において重要な要件であると考えています。以降、これら4点に着目した私たちの具体的な取り組みについて紹介していきます（図1）。

音声 × 深層生成モデルの取り組み

従来技術において代表的なものは、混合ガウス分布に基づく統計的声質変換⁽¹⁾です。入力音声の特徴量から目標音声の特徴量への変換関数を得るために、事前に時間整合をとった入力音声と目標音声の同一発話文（パラレルデータ）を用意することで、両音声の特徴量の同時確率を記述したモデルです。また、近年では、前述のパラレルデータを必要とする枠組みにおいて、性能改善のため、ニュー

* 非パラレルデータ：入力音声と目標音声とで発話内容が異なるデータ（非同発話文）を示します。

ラルネットワークを用いた手法や非負値行列因子分解などを用いた事例ベースの手法の検討も進められています。しかしながら、これら従来技術には、①学習データとして同一発話内容の音声ペアが必要であったり、②変換可能な音声特徴が声質に限られていたり、③音声の特徴量から波形を合成する際に古典的なボコーダを用いているため、合成される音声と実音声は容易に聞き分け可能であるなど、技術的制約があります。

一方、画像認識や自然言語処理の分野において、2014年以降、変分自己符号化器 (VAE: Variational Auto-Encoder)、敵対的生成モデル (GANs: Generative Adversarial Networks)、系列変換モデル (Seq2Seq: Sequence-to-Sequence model) など、非常に興味深い深層学習モデルが台頭してきました。Seq2Seqの1モジュールである自己再帰型モデル (AR: Auto-Regressive model) とVAE、GANsを総して三大深層生成モ

デルと呼ばれることもあり、画像処理や自然言語処理、音声信号処理などさまざまな分野・タスクでその有効性が確認されています。また、2015年中期の機械翻訳タスクにて、注意機構 (Attention mechanism)⁽²⁾がニューラルネットワークに導入され、その有効性とともに瞬く間に注目を浴びました。

NTTコミュニケーション科学基礎研究所では、前述の従来音声変換技術の課題を克服しさまざまな利用シーンに柔軟に対応可能な多用途音声変換システムを実現すべく、敵対的学習や系列変換学習などの方法論をベースに、①声質だけでなく長期依存特徴である韻律やアクセントの変換を行える「音声系列変換機能」、②発話内容に制約のない非パラレルデータを用いて声質変換を行える「非パラレル声質変換機能」、③合成音声波形から実音声波形へ波形空間上で変換を行い、出力音声の高音質化を実現する「波形ポストフィルター機能」などの新機能を創出しました。結

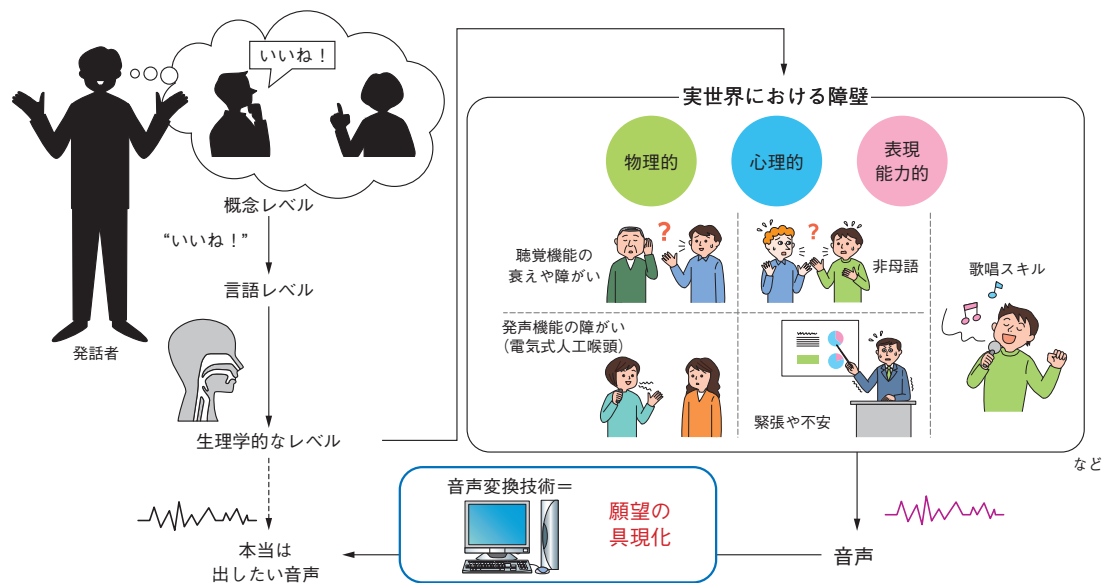


図1 音声 × 深層生成モデルの取り組み

果として、人の声と聴き分けられない品質の音声変換や任意の目標話者へ変換、非パラレルデータを用いることができるという高効率化、リアルタイム化を実現しています。また、近年盛んに研究されている分野横断型の研究として、目標話者の顔画像を用いて条件付けを行い、音声を変換する「クロスモーダル音声変換機能」も実現しています。

安定学習可能・高精度な音声系列変換

ここでは、前述の「音声系列変換機能」⁽³⁾を掘り下げようと思います。自然言語処理において離散値として扱われる単語（シンボル）と違い、音声は連続値の系列として観測されます。系列変換を用いた音声変換アプローチでは、大幅な高品質化が見込める一方で、連続値の系列を扱う際の学習の難しさが課題となります。この課題を克服するため、従来の音声系列変換では、音声認識とテキスト音声合成を組み合わせるアプローチが主流でした。音声認識を用いて、入力音声から単語のようなシンボルを認識し、離散値の系列であるシンボル系列上で変換を行い、変換されたシンボル系列から所望の音声を合成する、というアプローチです。離散値であるシンボルを用いて変換するので比較的安定した学習が可能な一方で、モデルを学習するために音声だけでなくテキストラベルも必要となります。また、テキスト化という処理を挟むことにより、テキスト化が困難な笑い声などの変換も容易ではありません。

NTTコミュニケーション科学基礎研究所では、テキストラベルを用いずに安定学習可能な音声系列変換を実現しました。図2に示されるとおり、後述の文脈保持機構を新たに導入し、音声データのみからすべてのモジュール

を学習します。テキストラベルは用いないものの、文脈保持機構が存在するため、音声の文脈構造が崩れる変換を許さないという制約が働き、適切な学習を導くことができるようになります。文脈保持機構において、入力音声の復元をタスクとするモジュール（図2左）は、入力音声を復元できるように、入力音声の情報を変換の過程で保持するよう働きかけます。いわゆる自己符号化と呼ばれる技術となります。一方で、目標音声を予測するよう新たに導入されたモジュール（図2右）は、変換の過程で得られる中間特徴量を、入力音声も目標音声も予測可能な特徴量にするよう働きかけます。いわば、入力音声を“話者共通の音声空間”へ写像させるような制約となります。これにより、文脈情報を保持するため音声認識に近いことを行いますが、音声認識により得られるシンボルよりも、“リッチな”中間特徴量を得ることができ、高精度な変換を手助けします。

合成音声波形から実音声波形への波形変換

音声変換のもう1つの課題は、音声の特徴量から波形を合成した際に、波形合成器の精度の影響を受けるという点にあります。世の中にあふれる音声を聞いて、「これは機械により合成された音声だな」と判断できることもあるかと思います。そういった“合成音声っぽさ”をなくし、より高音質で肉声感のある自然な音声波形への波形の直接補正をめざします。

波形を深層学習で扱う際の難しさは2つあります。1つは、例えば、16 kHzサンプリング音声には1秒間に1万6000個の点があります。合成音声と自然音声のペアデータを用

いて対応点を求めようとする、非常に困難なタスクとなることは想像が容易かと思えます。もう1つは、「位相」という扱いが非常に難しい特徴量の存在です。そのため、音声変換では位相情報を捨てて振幅情報のみを扱うのが定石となっています。

NTTコミュニケーション科学基礎研究所では、前述の課題を克服し、深層学習を用いた音声波形の直接補正を実現しました⁽⁴⁾。図3に示されるとおり、補正法のコアモジュールである循環型敵対学習モデルにおける変換器は2つの指標に沿って学習されます。1つは、敵対学習の学習指標である「識別誤差の最小化」です。合成音声と自然音声との定量化しづらい違いを識別モデルに認識させ、その差をできるだけなくすように合成器を学習します。また、1つの識別器ではなく、複数の識別器を用いてさまざまな角度から違いを認識させることにより、高精度な変換を実現しています。もう1つは、循環モデルの学習

指標である「再構成誤差の最小化」です。合成音声を自然音声化し、再度、合成音声化することで、入力である合成音声を復元します。この際に、完全に復元するように誤差を最小化します。ここがポイントで、位相まで含めて波形を再構成するよう制約付けることで、位相を適切に考慮した学習が可能となります。これらにより高音質化を実現していますが、さらに嬉しいことに、循環型敵対的モデルは、パラレルデータを必要としない枠組みであるため、合成音声と自然音声を適当にかき集めることで、所望のモデルを学習することができます。

今後の展開

NTTでは、より多様な音声への変換に向けて、技術改良と実証を進めていきます。現在までに、具体的な目標話者が存在する音声への高品質な変換には成功していますが、「もっと可愛い声で喋りたい」や「もっと渋

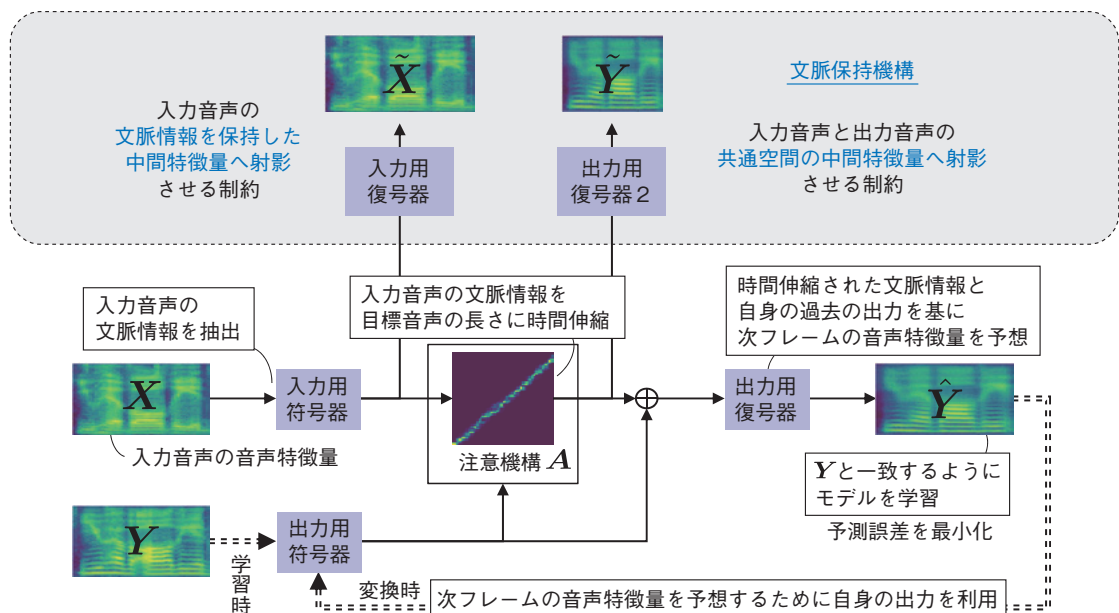


図2 安定学習可能・高精度な音声系列変換

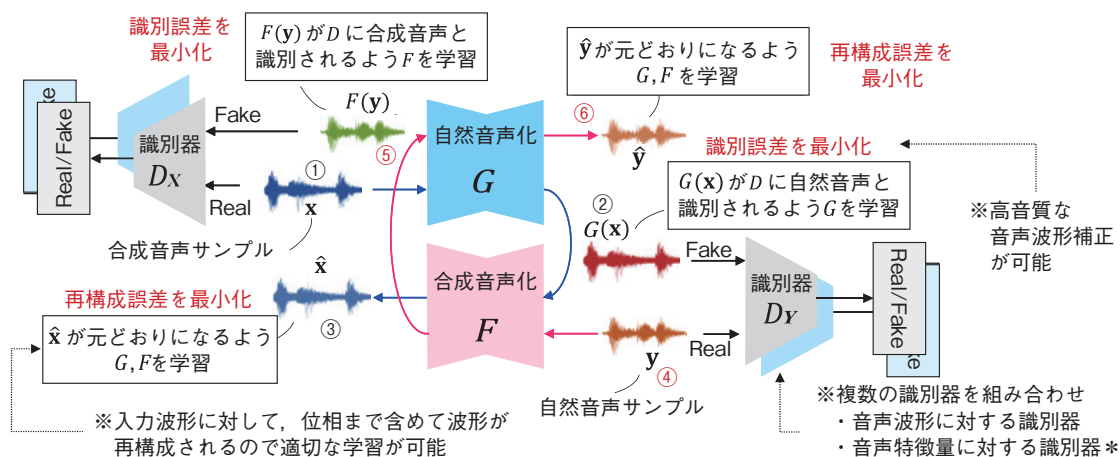


図3 合成音声波形から実音声波形への波形変換

い声で話したい」といった、目標話者ではなく音声のイメージを伴う変換を実現するには、音声の知覚的な空間を掌握し、空間上の潜在変数でうまく補間できるように、技術自体の性能を高めていく必要があります。ユーザのあらゆる願望にこたえ、さまざまな利用シーンに柔軟に対応可能な多用途音声変換システムの実現に向けて、研究開発を推進していきます。

参考文献

- (1) T. Toda, A. W. Black, and K. Tokuda: "Voice conversion based on maximum likelihood estimation of spectral parameter trajectory," IEEE Trans TASLP, Vol. 15, No. 8, pp. 2222-2235, 2007.
- (2) D. Bahdanau, K. Cho, and Y. Bengio: "Neural Machine Translation by Jointly Learning to Align and Translate," in Proc. ICLR 2015, San Diego, U.S.A., May 2015.
- (3) K. Tanaka, H. Kameoka, T. Kaneko, and N. Hojo: "AttS2S-VC: Sequence-to-sequence voice conversion with attention and context preservation mechanisms," in Proc. IEEE ICASSP 2019, Brighton, U.K., May 2019.
- (4) K. Tanaka, T. Kaneko, N. Hojo, and H. Kameoka: "Synthetic-to-natural speech waveform conversion using cycle-consistent adversarial networks," in Proc. IEEE SLT 2018, Athens, Greece, Dec. 2018.



(上段左から) 田中 宏 / 金子 卓弘

(下段左から) 北条 伸克 / 亀岡 弘和

深層学習の目覚ましい発展により、メディア生成・認識においてさまざまなことが高精度に実現可能となりました。私たちは、あらゆる人が所望の音声で思いのままに表現でき、不自由なく快適にコミュニケーションを行える環境を実現することをめざしています。

◆問い合わせ先

NTTコミュニケーション科学基礎研究所
メディア情報研究部
TEL 0774-93-5020
FAX 0774-93-5026
E-mail cs-liaison-ml@hco.ntt.co.jp

量子情報処理における量子的間接制御の可能性

近年、量子コンピュータや量子暗号など量子と名の付く話題を目にする機会が増えています。これら量子情報処理を実現できれば、通常のコンピュータで解けない問題が解けたり、絶対に秘密が守られる暗号が実現できたりします。しかし、その実現にはまだ越えなければならない山も存在します。本稿では、その山を越えるための手段の1つとして、量子的間接制御という手段を紹介するとともに、量子情報処理を実現する手段として量子的間接制御が有用な性質を持っていることが発見されたことを示します。

かとう ごう
加藤 豪

NTTコミュニケーション科学基礎研究所

量子情報処理の現状

量子情報処理の1つである量子コンピュータは、2つの状態が同時に存在できるという量子力学の性質を利用した計算機であり、並列の処理が自然にできる計算機です。その結果、現在使われているどんな暗号も容易に解読するなど、現在のコンピュータでは不可能とされることができるようになります。それ故に、量子コンピュータの実装が大きく期待されているところです。また、量子情報処理のもう1つとして、量子暗号がありますが、同時に計ることができない量があるという量子力学の性質を利用した暗号であり、原理的に安全な暗号です。量子暗号は、従来使われている現代暗号の欠点を補完する要素として期待されています。

実用化という面において、量子暗号が実装された装置はすでに開発されています。こういった装置は現時点では一般に販売されてい

るわけではないものの、特殊な場合であれば実社会に導入されるのも時間の問題でしょう。実際、中国・EU・米国においては、量子暗号の実機を使ったフィールド実験の研究が国家主導で盛んに行われています。

他方、量子コンピュータの実用化への道りはまだ今少し遠い状況です。Googleがつくった量子コンピュータが従来のコンピュータを凌駕することを示したという記事や、IBMがクラウド上で量子コンピュータを使うサービスを始めたという話題などを目にした方もいらっしゃるかもしれません。実際、数十量子ビットの量子コンピュータはすでに実現されています。しかし、社会のありかたを変えることができるほどの規模・精度の量子コンピュータにまでにはなっていません。例えば、現在インターネット上でまさに使われている暗号を解読するためには、ノイズが全くない理想的な量子コンピュータであっても数千量子ビット、ノイズがわずかに残って

いるものならば、数百万量子ビット必要です。つまり、現在の量子コンピュータの規模は、社会に影響を与えるにはまだ桁違いに小さいことがお分かりいただけるでしょう。

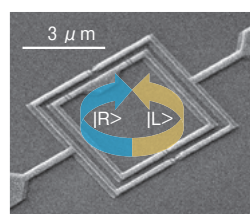
量子コンピュータの規模を大きくするには現在2つの解決すべき問題点が存在します。

1つはノイズの増大です。いかに小規模なシステムであろうと、量子コンピュータにおいてノイズを完全に除去することは現在できていません。そして、量子コンピュータの規模が大きくなればなるほど、ノイズの影響は大きくなっていきます。そのため規模を大きくしてもノイズの影響を増加させない手立てを見つけ出す必要があります。量子誤り訂正符号はこの問題を解決する1つの有力な手段であり、現在もよく研究されていますが、実用できるまでにはなっていません。もう1つの問題は、大規模な量子コンピュータを制御するのに必要な制御機構は現実的には実装困難な点です。通常の量子コンピュータのモデルにおいては、その量子コンピュータを構成する量子ビットの数に比例する数の制御機構が必要です。つまり、量子コンピュータが大規模となり量子ビットの数も増大するに従って、膨大な量の制御機構が必要になるわけですから、量子ビット数が数百万というレベルでは、必要とされる制御機構の設計は現実的には大変難しいのです。通常のコンピュータではバスという概念で同様の問題を解決していますが、現時点では量子コンピュータの実装において量子バスとでも呼べるものは、いまだ大規模には実現していません。

量子コンピュータ実現に対する課題を解決する手段としての量子的間接制御

前述の2つの問題を解決するために、従来からさまざまな研究がされているところですが、新たな側面からの解決手段を探索することも重要です。量子的間接制御という考え方は、まさにその新たな解決手段になり得るものです。

量子的間接制御についてまず説明します。量子的間接制御とは、制御対象となる量子系が、直接操作が可能な量子系（S系）と、間接的にしか操作できない量子系（E系）とからなっていて、S系をうまく操作することで、S系とE系からなる全体を間接制御するという考え方です。例えば、操作が容易で1量子ビットと見なせる量子系、つまりS系として用い得る系として、超伝導体をリング状に配置することで実現できる超伝導量子ビット（図1(a)）があります。また直接操作が困難な1量子ビットと見なせる量子系、つまりE系として用い得る系として、ダイヤモンド中の窒素による格子欠陥によって実現されるNV（Nitrogen Vacancy）センタ（図1(b)）があります。超伝導量子ビットの上にNVセンタの含まれるダイヤモンドプレートを置くだけで、その2つの量子系は相互作用によってつながり、超伝導量子ビットをS系・NVセンタをE系とする量子的間接制御を行う系が実現します。実際に1量子ビットの量子メモリ⁽¹⁾をこの系によって実装する実験はすでに行われています。

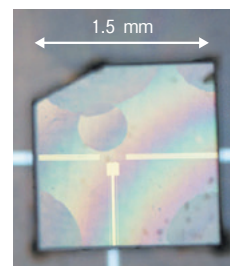


(a) 超伝導量子ビット

直接制御
可能な系



間接的にしか
制御できない系



(b) NVセンタ

図1 量子的間接制御の例

こういった量子的間接制御の構造が一般的に、前述した第一の問題点であるノイズの増大を効果的に解決できる手段であることは、次のようなかたちで理解することができます。量子系の操作をするには物理的な操作の配線を配置する必要があります。しかし、物理的な配線がつながっていれば、それが必然的にノイズの侵入する経路にもなります（図2）。他方、ノイズの侵入を減少させようと思うと、操作配線を減らして、操作性を犠牲にせざるを得ません。つまり、量子系への操作性の高さとノイズ侵入の困難性との間にはトレードオフの関係が存在します。現状の量子コンピュータの実現形態においては、量子系全体への高い操作性を前提としており、そのことがノイズ侵入経路を減らすうえでの限界を与えています。そのため、量子的間接制御のように操作性が低くても、自由に全体制御が可能であることが分かれば、ノイズ侵入経路を減らし、ひいてはノイズ低減自体を高いレベルで実現したうえで、量子情報処理を実現できることが期待されます。

さらに量子的間接制御で量子系全体を自在に制御できることは、前述した第二の問題点に対しても有効に対処可能であることを意味します。というのも、量子的間接制御はE系にある量子情報をS系に取り出して処理してはE系に戻すことの繰り返しであることからえることができ、抽象的には量子バスの構造を持っているといえるためです。つまり大規模な量子的間接制御を行えるようになることで実現する機能は、大規模な量子バスを実現することで得られる機能を含んでいます。

量子コンピュータを量子的間接制御で実装するにあたっての課題

前述のように、量子的間接制御は大規模な量子コンピュータをつくるうえでの主要な2つの問題を解決する手段となり得ます。にもかかわらず、量子メモリのような小規模な利用例はあるものの、大規模な量子情報処理への実装手段としては検討されず、十分な研究もなされてきていませんでした。というのも、S系とE系をつなぐ相互作用のわずかな違い

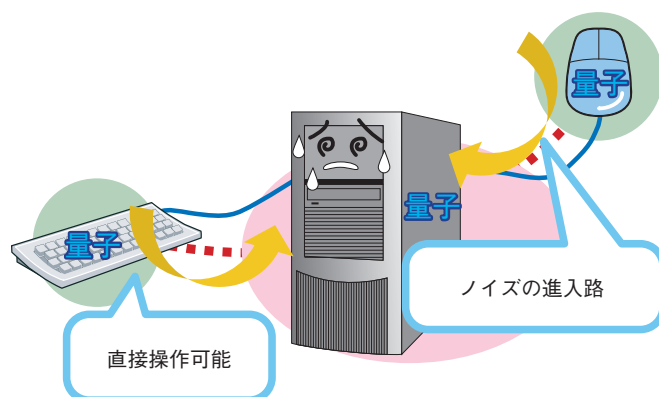


図2 ノイズの侵入経路

で、実行可能な制御の種類さえ大きく変わってしまい、そのような状況では情報処理の能力を議論することさえも困難だからです。実際、量子コンピュータとして機能するためには、どんな制御でも実行可能である必要がありますが、与えられた量子的間接制御がそのような性質を持っているかという基本的な質問にさえ、今までは明確に答えることはできていませんでした。

本稿で示す研究成果⁽²⁾は、この基本的な質問に対して答えるものです。ある量子系が与えられたとき、その量子系でどのような制御が実行できるかは、動的Lie代数^{*1}と呼ばれる歪エルミート行列^{*2}の集合によって判断できることが知られています。そのため、量子系が与えられたときに、その動的Lie代数が何かを知ることこそが、その量子系が実行できる制御の範囲を調べることに相当します。例えば、現在考えられている典型的な量子コンピュータの場合、対応する動的Lie代数にはすべての歪エルミート行列が含まれる

ことがすでに知られており、そのことは当該量子系においてどんな制御でも実行可能であることを意味しています。

ところが、一般の量子系の場合、一般的に実現し得る動的Lie代数には多くの可能性が存在します。それ故に、量子的間接制御において、上記の基本的な質問にさえ答えを与えることが困難でした。しかし、最近行われた研究⁽²⁾によって、操作が量子的間接制御によって行われるという事実さえ仮定すれば、その動的Lie代数にはJordan代数^{*3}と深く関連する構造の存在が要求されることを発見しました。Jordan代数は動的Lie代数に比べて多様性が非常に小さいことが知られてい

*1 動的Lie代数：交換子積に関して閉じていて線形空間を成す元からなる集合。つまり、この集合は、この集合の任意の2つの元をA, Bとし、任意の実数をx, yとすると、 $AB - BA$ や $xA + yB$ もこの集合に含まれます。

*2 歪エルミート行列：複素共役＋転置によって自身に負符号をつけたものになる行列。

*3 Jordan代数：反交換子積に対して閉じていて線形空間を成す元からなる集合。つまり、この集合は、この集合の任意の2つの元をA, Bとし、任意の実数をx, yとすると、 $AB + BA$ や $xA + yB$ もこの集合に含まれます。

ます。この事実を使うことで、量子の間接制御の場合の実現し得る動的Lie代数を完全に分類するとともに、量子の間接制御が量子情報処理を実装する手段として有用な性質を持つことが分かりました。

得られた動的Lie代数の分類によると、直接操作可能な量子系（S系）の次元^{*4}によって、動的Lie代数の構造が全く異なることが分かりました。特に、S系の次元が3以上の場合は、量子の間接制御によって、影響を及ぼせる空間に対する任意の歪エルミート行列が動的Lie代数に必ず含まれることが分かりました。つまり、S系の次元が3以上の場合には、実質的に背後にある量子系に対してどのような制御でも実行可能であると認識して良いことを意味しています。つまり、量子情報処理を実装する装置として十分な性能を量子の間接制御は普遍的に有することを意味しています。

まとめ

このように、量子の間接制御の数学的な性質を研究することで、量子の間接制御が量子情報処理を実現する手段として有効であることが分かりました。ここで示した結果をもって、ノイズを非常に低く抑えた量子コンピュータが明日にでも実現するというわけではありませんが、量子コンピュータ実現への課題の解決手段として今まで検討されてこなかった量子の間接制御という手法に可能性が有るこ

*4 次元：通常の情報処理で言うところの基数。例えば3量子ビットからなる量子系は8次元です。

とを理論的に示したことは、それ自体が大きな一歩です。さらには、この研究で得られた理論体系を、制御の実行時間をも扱える理論へと発展させることなどで、量子コンピュータ実現へのさらなる大きな貢献ができるでしょう。

参考文献

- (1) X. Zhu, S. Saito, A. Kemp, K. Kakuyanagi, S. Karimoto, H. Nakano, W. J. Munro, Y. Tokura, M. S. Everitt, K. Nemoto, M. Kasu, N. Mizuochi, and K. Semba: "Coherent coupling of a superconducting flux qubit to an electron spin ensemble in diamond," *Nature*, Vol. 478, pp. 221-224, 2011.
- (2) G. Kato, M. Owari, and K. Maruyama: "Algebra and Hilbert space structures induced by quantum probes," *Ann. Phys.*, Vol. 412, 168046, 2020.



加藤 豪

量子情報は、実用化がもうすぐそのものから、まだ夢物語のものまでいろいろありますが、より多くのものを実現していけるよう、理論的な探求を行っていきます。

◆問い合わせ先

NTTコミュニケーション科学基礎研究所
メディア情報研究部
TEL 0774-93-5020
FAX 0774-93-5026
E-mail cs-liaison-ml@hco.ntt.co.jp

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

文章を読むときに、難しいと感じるか、易しいと感じるかは、文章の難易度と読み手自身の知識量の両方に依存します。文章（テキスト）の難しさを自動的に推定するための難易度推定と、読み手側の知識量（語彙数）の推定が、簡単に、高精度にできれば、ちょうど良い難しさのテキストを推薦することができ、学習支援につなげることも可能です。本稿では、私たちが取り組んでいるテキストの難易度推定方法と語彙数推定方法を紹介します。

ふじた さなえ
藤田 早苗

NTTコミュニケーション科学基礎研究所

難しさを測る意味

文字を覚えてたての子どもが自分で読むつもりで選んだ絵本が読めず、読んであげることになったことはありませんか。中学1年生のときとても苦労して読んだ英文が、大学生になるころにはとても簡単に感じられたことはありませんか。同じ文を読もうとしても、難しいと感じるか易しいと感じるかは、読み手の知識量に依存します。

もし、読み手にとってちょうど読めるくらいの、あるいは少し頑張れば読めるくらいの絵本や本、英文を薦めることができれば、無理なく読み手の知識を増やしていけるかもしれません。しかし、「ちょうど読めるくらい」や「少し頑張れば読めるくらい」を判断するのは簡単ではありません。人の知識量と文（テキスト）の難易度の両方を適切に推定する必要がありますからです。

本稿では、この両方の推定方法の研究と、

推定を支える言語資源の構築について紹介します。

人の語彙数を測る

人に必要な知識の1つとして語彙の知識が挙げられます。NTTでは20年以上前から、さまざまな年代の人の語彙数の調査や推定に取り組んできました。

幼児を対象とした調査では、語彙数自体は多くないので、理解・発話できるすべての語彙を調査することも不可能ではありません。実際私たちは、1500組以上の親子モニターの皆様にご協力をいただき、子どもがいつごろどのような語を覚えるか、発話できるかというデータを蓄積し、幼児語彙発達データベースを構築してきました。

しかし、小学生以上となると、知っているすべての語彙を調査することは困難です。そこで、提示した語を知っているか回答してもらうことにより、語彙数を推定します。提示

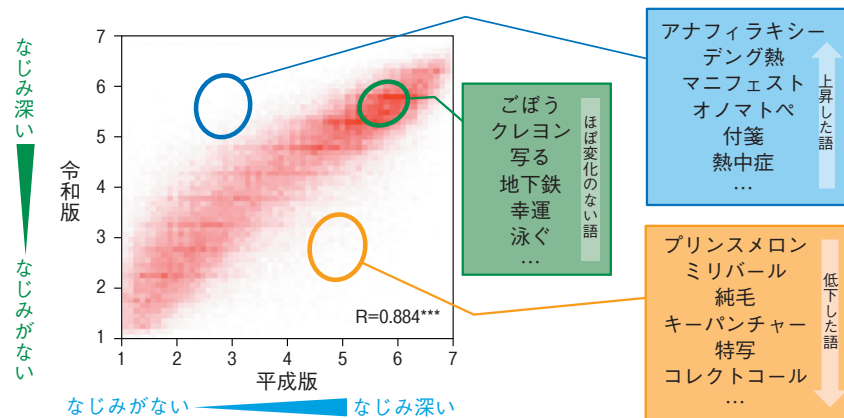


図1 平成版から令和版への単語親密度の経年変化

する語は多ければ多いほど正確な推定ができますが、数十語でも推定可能です。

この推定方法では、ある語を知っていると回答したときに、何語知っていると仮定するかがポイントとなります。例えば「しょっぱい」と「検潮儀」だと、「検潮儀」のほうが知っている人は少ないでしょう。そのため、「しょっぱい」だけを知っている人より「検潮儀」も知っている人のほうが、より多くの語を知っていると仮定します。では、「検潮儀」を知っていれば何語ぐらい知っていると仮定できるのでしょうか。その仮定の根拠となるのが、次に紹介する「単語親密度」です。

単語親密度データベース

「語のなじみ深さ」を評定実験によって数値化したものを「単語親密度」と呼びます。語に付与された数値が大きければ大きいほど、多くの人になじみのある語であり、数値が小さければ小さいほど、多くの人にとってなじみのない、あまり知られていない語であることを示します。

NTTでは20年以上前から単語親密度データベースなどの基盤的言語資源の構築に取り組んできました。過去に構築した約7万7千

語からなる平成版の単語親密度データベースは、NTTデータベースシリーズ「日本語の語彙特性」として公開され、心理学や言語教育、言語聴覚療法分野などの基礎指標として幅広く活用されてきました。しかし、調査から時間が経ち、単語親密度自体が時代とともに変化した可能性もあり、新しく出現した語（「インターネット」や「コンビニ」など）に対応していないといった問題もありました。

そこで、再調査と新しい語の追加調査を実施し、令和版単語親密度データベースとして約16万3千語という過去最大のデータベースを構築しました⁽¹⁾。さらに、平成版単語親密度からの変化を調査し、両者に強い相関があり、多くの語では20年以上経っても親密度に大きな変化がないことを確認しました。一方で、大きく変化した語も一部存在すること、こういった語が大きく変化したかを明らかにしました（図1）。

単語親密度を用いた語彙数推定方法

単語親密度を用いた語彙数推定では、親密度の高い語から低い語まで、何段階かの親密度の語をサンプリングして出題します（図2）。基本的に、知っていると回答された語

以上の親密度の語はすべて知っているとして仮定して、語彙数を推定します。例えば、「しょっぱい」までなら約1500語、「検潮儀」まで知っていれば約13万9千語と仮定します。ただし、知っている語と知らない語の境界付近では、知っているかどうかにはばらつきがあると考えられます。そこで、回答結果にロジスティック回帰曲線を当てはめ、知っている確率がちょうど50%になる語彙数を推定結果とします（図3）。

このようにすると、サンプリングした少数の語をチェックしてもらうだけで、回答者の

語彙数を簡単に推定することができます。また、理論的には同じ親密度の語であれば何を出題しても良いため、出題する語の変更が容易で、さまざまなバリエーションのテストを作成することができます。もちろん、チェックしてもらう語は多いほど推定精度を向上させることができます。

本手法で推定できる語彙数の上限は単語親密度データベースのサイズに依存しますが、令和版単語親密度データベースの構築により、推定できる語彙数の上限が大きく上昇し、テストの汎用性を高めることができました。

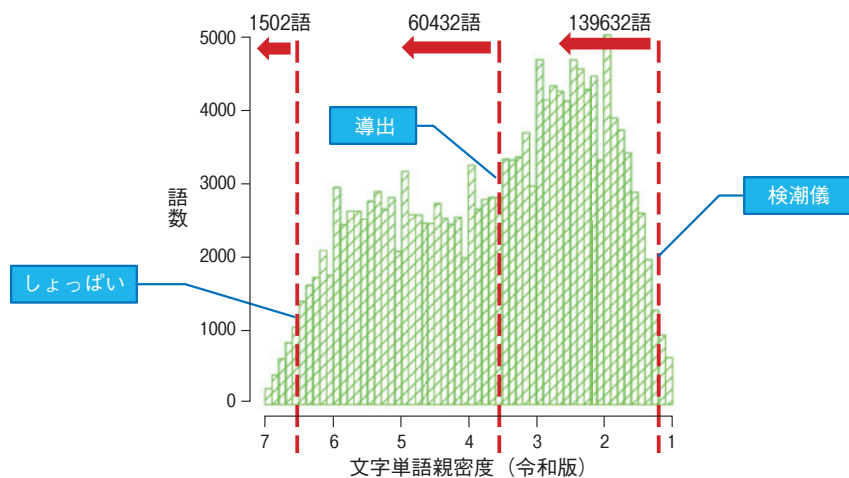


図2 単語親密度に基づくサンプリング

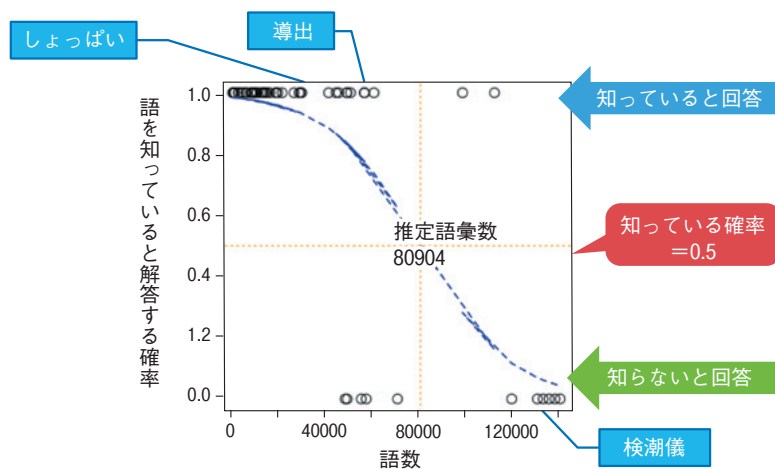


図3 語彙数推定方法

大規模な語彙数調査

これまでは児童・生徒を対象とした大規模な語彙数調査はほとんど実施されてきませんでした。私たちは、本手法を用いて、公立の小学生～高校生2800人以上を含む、約4600人の語彙数調査を行いました。

その結果、特に、小中学生では急激に語彙数が上昇すること、成人でも、年齢とともに語彙数が上昇することを確認しました（図4）。調査結果から、同じ学年でも生徒によって語彙数に大きなばらつきがみられるため、支援が必要な生徒を見つけるのに役立つと考えています⁽²⁾。

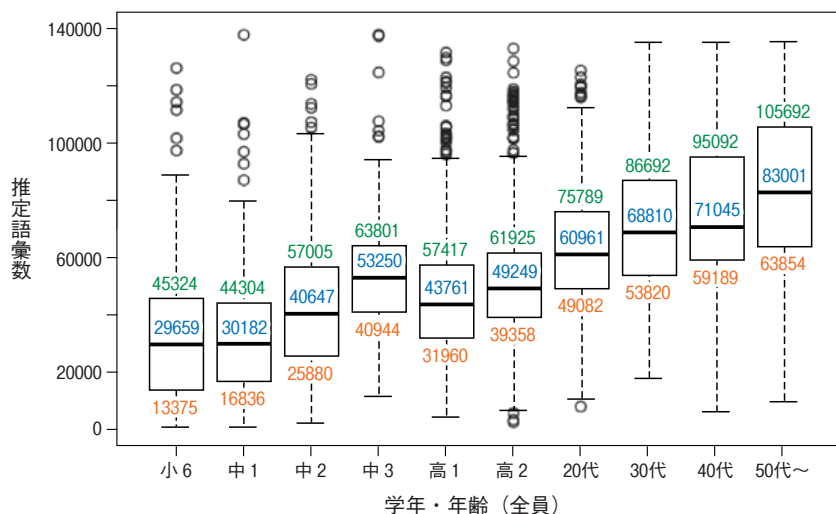
さらに、各学年・年齢における、単語親密度と語彙獲得状況（その語を知っている人の割合）との関係を分析しました（図5）。これは、各単語親密度の語を知っていると回答した人の割合（獲得割合）を各学年・年齢ごとに示したものです。どの学年・年齢でも親密度が高い語ほど、知っている人の割合は高くなる傾向があり、年齢が上がれば上がるほど、この傾向は顕著になります。一方、成人

に比べて、小学生や中学生では、比較的親密度の高い語であっても、知っているかどうかの個人差が大きく、ばらつきがあります。こうした分析から、単語親密度を手掛かりとして、児童・生徒がこれから重点的に獲得するだろう語彙、あるいは獲得したほうが良い語彙を見つけていくことができると考えています。

また、学校での調査用とは別に、一般の方に試していただける令和版語彙数推定テストをオープンハウス2020に合わせて公開しました（<http://www.kecl.ntt.co.jp/icl/lirg/resources/goitokusei/>、2020年6月4日公開）。公開後約10日で3万人以上の方にお試しいただきました。よろしければ、皆様もお試ください。

テキストの難しさを測る

次に、テキスト側の難易度推定方法を紹介します。最初に取り組んだのは、絵本の難易度推定です。NTT研究所でこれまで取り組んできた幼児の語彙発達研究と組み合わせることで、語彙発達の解明や発達支援に寄与で



※青字は中央値

図4 学年・年齢の語彙数推定結果

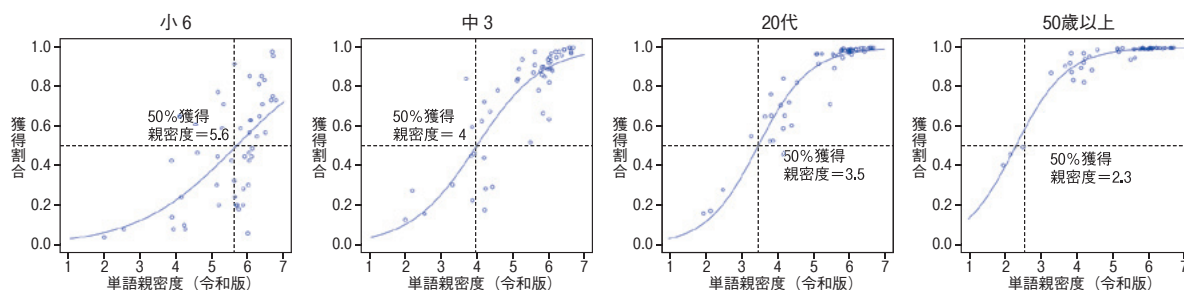


図5 単語親密度と語彙獲得状況の関係

きるのではないかと考えたからです。

テキストの難易度には、使われている語彙の難しさ、文構造の難しさの両方が影響します。絵本の場合、ひらがなを正しく解析することも必要です。例えば、「とうさん」が「父さん」か「倒産」かによって、難易度は大きく異なります。そうしたひらがな解析の精度向上に取り組み、幼児の語彙発達データベースなどの語彙の難しさを反映する特徴量と、文の長さ等の文構造の難しさを反映する特徴量を用いることで、0～2歳、3歳、4歳、5歳という4クラス分類で87.8%という高精度な難易度推定を可能にしました⁽³⁾。また、一口に3歳向けといっても、2歳に近い3歳という場合もあれば、4歳に近い3歳という場合もあります。そのため、前後の年齢向けと推定してもあまり問題にはならないと考えられます。前後の年齢向けと推定しても良いとして評価すると、精度は96.7%となりました。つまり、提案手法は、対象年齢を大きく間違えることがほとんどなく、ロバストで信頼度の高い難易度推定が可能であるといえます。この難易度推定方法は、絵本検索システム「ぴたりえ」⁽⁴⁾で子どもにあった読みやすさの絵本を探すために使われています。

なお、テキストの難易度推定の研究では多くの場合、教科書の学年を推定できるかどうかで評価を行います。そこで、提案手法を適

用して教科書の学年を推定する評価も実施したところ、小学1年生から中学3年生の9クラス分類で、98%以上という高い精度で学年を推定できました。つまり、絵本の難易度推定の精度向上のための工夫は、小学生以上のテキストの難易度推定にも有効であることが分かったのです。

NTT 絵本・児童書コーパス

前述の難易度の推定には、電子データ化された絵本の文章のデータ（コーパス）を利用しています。しかし、絵本のコーパス自体ももとは存在しなかったため、その構築からスタートしました。実はこの部分がもっとも泥臭くて時間がかかる部分です。絵本では絵の中に文字が書かれることが多く、OCRによる文字認識もできません。結局、ほとんど人手で本文を入力することになりました。地道な作業の甲斐あって、NTTの絵本コーパスは、日本語の絵本6000冊以上、英語の絵本2500冊以上からなる世界に類をみない規模になりました。しかも今でも拡張中です。

さらに私たちは、構築したNTT絵本・児童書コーパスを用いて、絵本と幼児の語彙発達や感情発達との関係調査など、多くの新しい研究に取り組んでいます。



図6 語彙力にあった英語コンテンツの推薦

今後の展開

今回紹介した人の語彙数やテキストの難易度の推定は独立に行っています。しかし、両者を組み合わせれば、定期的な語彙数の確認をしつつ、個人ごとに「ちょうど読めるくらい」や「少し頑張れば読めるくらい」のテキストを推薦することが可能になります。実際私たちは、英語の語彙数推定と難易度推定の研究も進めており、語彙力にあった英語絵本を推薦して学校の英語教育に活かす取り組みも始めています⁽⁵⁾(図6)。

私たちは今後も、日本語でも英語でも、幼児でも小中高校生でも、大人に対しても、エビデンスを積み重ねながら、1人ひとりに合った育児・教育支援の実現をめざしていきます。

参考文献

- (1) 藤田・小林：“単語親密度の再調査と過去のデータとの比較,” 言語処理学会第26回年次大会, 2020.
- (2) 藤田・小林・山田・菅原・新井・新井：“小・中・高校生の語彙数調査および単語親密度との関係分析,” 言語処理学会第26回年次大会, 2020.
- (3) 藤田・小林・南・杉山：“幼児を対象としたテキストの対象年齢推定方法,” 認知科学, Vol. 22, No. 4, pp. 604-620, 2015.
- (4) 藤田・服部・小林・奥村・青山：“絵本検索システム「びたりえ」～子どもにぴったりの絵本を見つけます～,” 自然言語処理, Vol. 24, No. 1, pp. 49-73. 2017.
- (5) 藤田・服部・小林・納谷：“日本人初學者の語彙数推定方法の検討,” 2020年度人工知能学会全国大会, 2020.



藤田 早苗

自分の当たり前と、他人の当たり前は違います。忘れがちなことですが、自分や大人の「当たり前」を押し付けるのではなく、1人ひとりに合った育児・教育支援の手助けとなるよう今後も取り組んでいきます（自戒を込めて）。

◆問い合わせ先

NTTコミュニケーション科学基礎研究所
 協創情報研究部
 言語知能研究グループ
 TEL 0774-93-5020
 FAX 0774-93-5026
 E-mail cs-liaison-ml@hco.ntt.co.jp

主役登場

私が欲しいと思うものは
他の人も欲しいはず

藤田 早苗

NTTコミュニケーション科学基礎研究所
主任研究員



私自身は子どものころから本好きで、家族の貸し出しカードまで使って図書館で手あたり次第に本を借りて読んでいたことを覚えています。私には子どもが3人いますが（娘2人と息子1人）、子どもたちにも本好きになってほしいと思い、寝る前の読み聞かせを心掛けていたのですが…。

好きな絵本を持ってこさせると、娘たちは物語の絵本を持ってくるのに、息子は恐竜図鑑を持ってくるのです。正直私は恐竜には全く興味がありませんでした。それなのに、毎晩、延々と、恐竜の名前、体長、生きていた時代や地域を読まされるのです。せめて物語になっていれば私も読む気がする、と思って、図書館で恐竜の出ってくる絵本を片っ端から借りて読みましたが、ほどなく読みつくしてしまいました。しかも、図書館にある恐竜の出ってくる絵本と図鑑を（おそらくすべて）読み切ったら、息子はもう図書館には行きたがらなくなりました。これでは本好きにはならないかもしれません。焦った私は、まだ読んでいない恐竜の絵本があるかもしれない、絵本でなくてもいいかもしれない、恐竜じゃなくても古生物やモンスターも好きかもしれないと、息子が興味を持つかもしれない、読めるかもしれない本を探して図書館をさまよいました。ところが、図書館の何万冊もの本の中から、息子の興味を引きそうでちょうど読めそうな本を探すのは簡単なことではありませんでした。

この経験からできたのが、絵本検索システム「びたりえ」です。1人ひとりの子どもに合った読みやすさで興味のありそうな内容の絵本を簡単に探せるようにとつくりました。私の専門は自然言語処理（日本語や英語といった自然

言語を機械で上手く扱うための研究分野）ですが、絵本の文の難易度推定やひらがなの解析に自然言語処理技術が活かされています。「びたりえ」は2019年にNTTデータ九州から商用化されました。いずれ全国津々浦々の図書館や保育園・幼稚園で使っていただけるようになって、絵本の読み聞かせに悩む親御さんの一助になればと願っています。

このこと自体は大変嬉しいことですが、残念ながら、基礎研究が商用化されるより子どもの成長のほうが早く、子どもたちは商用化前に絵本の読み聞かせ年齢を卒業してしまいました。でもこれであきらめる私ではありません(!?)。

自然言語処理技術は日本語だけでなく、英語にももちろん適用できます。折しも2020年度から小学校で英語が必修化されるなど、英語の学習は重点項目になっています。ということで、次は英語の学習支援です。英語のテキストの難易度推定や英語の語彙数推定など、日本語で培った技術を活かし、英語力にあった英語コンテンツの推薦などの学習支援に取り組んでいます。

自分の研究を自分が必要だと思うことに活かせる、ましてや自分自身の子育てに活かせるなんて、これほど幸せなことはありません。公私混同?いえいえ、私が欲しいと思うものは他の人も欲しいはずを信条に、英語の学習支援は子どもの成長に負けないスピードでの実現をめざして頑張っていきます。