

挑戦する 研究者たち CHALLENGERS



中谷 智広

NTTコミュニケーション科学基礎研究所
上席特別研究員

人とコンピュータが同じ音空間を共有して、自由に協力し合える鉄腕アトムのような世界を実現したい

コンピュータによる自動音声認識技術が急速に発展し、スマートフォンやスマートスピーカ等の音声インタフェースで当たり前利用される昨今、人との対話に向けてさらなる技術の向上が待ち望まれています。さまざまな音が聞こえている中から周囲の雑音や残響を除去して聞きたい音だけを高精度に抽出する技術に関して、世界最先端の研究を続けている中谷智広上席特別研究員に研究活動の進捗と世界初に臨む研究者の姿勢について伺いました。



あらゆる環境で人の会話を理解する 音声入力インタフェースを実現

現在手掛けている研究から教えていただけますでしょうか。

私はさまざまな音が聞こえている環境の中から目的の音声を聞き分け、会話調のくだけた音声をも認識できる技術の構築をめざしています。2016年のインタビューの際にも同じ目標についてお話させていただきました。その当時と比べると、スマートフォン等に話しかけて操作することは最近では当たり前の光景になってきました。しかし、現在のAI（人工知能）による音声認識技術でも、人がコン

ピュータに話しかけるときには、AIが理解できるように特別な手順に従い、さらに話し方を工夫する必要があります。一方、将来的にロボットなどが、より深く私たちの生活に溶け込めるようになるためには、日常会話の中で人が人に話しかけているのと区別がつかないような自然な会話でも適切に認識できるナチュラル音声インタフェースの実現が必要になります。

ナチュラル音声インタフェース技術の実現に向け、現段階では、音を聞き分ける技術である音声強調をテーマとして私自身は研究に取り組んでいます。日常生活の中で、人は必ずしもマイクのそばで話してくれるとは限りません。

話をしている人から離れたマイクで収録した音声は、壁等に反射して届く音（残響）、複数の人の声（音源）、背景雑音も収録されます。これに対し、音声強調は、残響抑圧、音源分離、雑音抑圧を行うことで、特定の話者の口元にあるマイクで収録したような品質の音声にする技術です。

音声強調では、私たちは2つの研究のアプローチをとっています。1つは音の要素分解（図1）で、音の物理的・統計的性質の違いに基づいて収録した音を個々の要素に分解します。これは非常にコンピュータらしい、人間には不可能だけれどコンピュータが得意な技術領域です。最近の成果の1つとして、残響抑圧、音源分離、雑音抑圧を同時かつ全体最適に処理できる「畳み込みビームフォーマ（BF）」

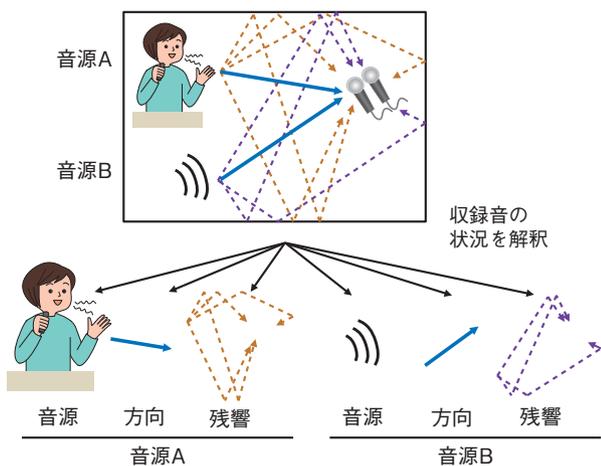
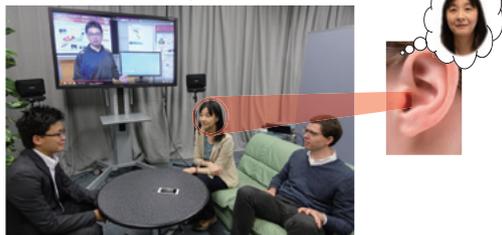


図1 複数マイクを用いた音の要素分解

日常のさまざまな場面で、人の声どうしが混ざることば、よくある



- 人は、声の特徴（声の高さ、声質など）に注目し、聞きたい人の声のみを聞くことができる（選択的聴取）
 - ・コンピュータにはできなかった！

図2 選択的聴取

を実現（世界初）しました。従来は、個別にしか各処理を実現することしかできなかったため、それらを組み合わせたときにベストな性能を引き出すことができませんでした。本技術により各話者の強調音声をさらに大幅に高品質化できるようになりました。

音声強調のもう1つアプローチは、選択的聴取（図2）で、聞きたい人の声に耳を傾けて聞き取る（抽出する）技術です。こちらは、人間は日常生活の中で普通に行っていることですが、コンピュータには最近まで不可能だった技術です。このアプローチの代表的な技術が、2018年9月号の本誌記事でも紹介した、深層学習技術を用いて特定の話者の声の特徴に合致する音声だけを取り出す技術「SpeakerBeam」です。これにより人間のような選択的聴取を世界で初めて実現しました。さらに、この技術を発展させて、時々刻々と変化する会話から「いつ、誰が、何を話したか」を高精度に推定（会話分析）する技術RSANも実現しています。

私たちは、このようにコンピュータが得意な技術分野を活かしつつ、人間が得意な処理の領域にも踏み込んでいくことで、ナチュラル音声インタフェースを実現できると考えています。

自分のやり方と違うと感じても、先輩や周囲に耳を傾けてやってみるのも一手

このテーマに取り組み始めたきっかけはありましたか。また、課題やテーマを探す際に心掛けていることはありますか。

私はNTT入社以来、音源分離のテーマに取り組んできました。NTT入社当時は音源分離の研究はまだ黎明期で、「そんなことでもできるわけない、もっと先に解決すべき課題がある」という雰囲気がありました。それでも、私は人間にできるのだからコンピュータにもできるはずだと思って追究し続けてきました。今の私なら、周りから必ずしも賛同が得られない研究を続けるのは簡単なことではないかもしれませんが、当時の私は「すぐにできることよりも、こんなことができたらいいいと思えること」に対して、

自分の信念や自分の感覚を信じて研究を進めていました。結果として30年間この分野で研究を継続できているのだから、本当に良いテーマであったと思います。

研究テーマや課題を探すのはある意味ではパズルを解くのに似ているかもしれないと思っています。パズルを解く際にさまざまな可能性を模索する感覚です。例えば、最初は複数の声を聞き分ける技術はありませんでしたが、どうやったらそれができるのだろうかと考えて、音声の周波数分布から特性や特徴に関する仮説を立て、他の音声についても同様なことを繰り返し、多くの仮説に共通するコアな部分を見つけてテーマとしました。

このプロセスにはセンスと知識がモノを言います。知識についてはひたすら勉強するしかありません。新しい論文を読み、伝統的な技術に関しては教科書で勉強する。これを積み上げていくことでスキルが上がってきます。それ以外に知識を向上させる方法はありません。センスについては「自分だったらこうする」というビジョンを持っているかどうかではないでしょうか。これは研究者としても人として生きていくうえでも重要な能力だと思います。

上席特別研究員となっても日々勉強なのです。

世の中が発展している以上、勉強しないわけにはいかないのです。特にニューラルネットワークにおいてはコンピュータの速度が格段に上がったこと、研究者の数が非常に増えたことで研究がものすごいスピードで進んでいますから、勉強しなければ全く追いつけません。ただし、ただ闇雲に勉強するのでは生きた研究にはつながっていきません。直観と知識を同時に働かせて、常に自分がやりたかったことにどう役立てることができるかを考えていく必要があります。

また、月並みかもしれませんが、国際会議等で仲間とディスカッションの機会を設けることも大切です。自分のアンテナではキャッチできることは限られていますから、そうした仲間と議論しながら視野を広げていくことは間違いなく重要です。以前は、上司等から国際会議等に赴く際には「人に会ってきなさい」「研究機関を訪問して回ってきなさい」と指示され、半ば強制的に研究者を訪ね歩いてきました。

当時の私にはまだ研究者としての実績も少なく、英語でのコミュニケーションも全く自信がなかったため、相当に高いハードルでした。しかし、この取り組みを通じて、私自身が全く知らなかったことに触れることができ、その研究者等と知り合うことができました。会議で知り合うのも大切ですが、その方の研究機関を訪ねて初めて分かることも多くありますし、より深い議論ができます。このときにできた人間関係は、コロナ禍でもメールや遠隔会議等を通じて活着しています。

知らない研究者に会いたいと連絡することはハードルが高いと思われるかもしれませんが、最初の作業としてはメールを送るだけです。実は研究者の皆さんはとても親切で、ご自身も知り合いを増やしたいと思っていますから基本的には受け入れてくださいます。もちろん、私のところにも初めてのご連絡をいただくことがありますが、存じ上げない方でもお断りすることはありません。

このように、人とのつながりの大切さを教えていただきそれに努めてきたことで、残響抑圧の分野の研究において世界的な研究者とのつながりを持ち、ワークショップを開くこともできました。研究者は他者に認められないと自己満足の世界になり、また、世界的に認められることをめざすうえでも、時には先輩や周囲の声に、たとえそれが自分のやり方と違うと感じても、耳を傾けて参考にするのも1つの手かもしれません。



研究者は「箱」のような存在かもしれない

研究者にとって必要な素養を教えてくださいませんか。

まずは、自分がどういう研究に取り組んでいて何を解決したいと思っているかを、広く社会に知ってもらうことはものすごく大事だと思います。自らそれを宣言して活動をしているとさまざまな情報が集まってくるもので、自分1人では見つけられないような解決策を教えてもらえることもあります。



私は、研究者は「箱」のような存在ではないかと思えます。その箱には研究内容が書かれたラベルが貼ってあり、自分でもさまざまなデータを集めてくるし、他の人がデータや情報を寄せてくれるのではないのでしょうか。この行為はある技術を発展させるためのプロジェクトであり、どれだけ発展させられるかは研究者の能力にかかわってくるかもしれません。

また、良い情報を集めるためには、そのラベルの書き方も重要ですから、どう表現するかしっかりと考えて書かなくてはなりません。

研究者である以上、今取り組んでいる課題は本当に取り組むべきことか、本当に解明しなければならないことは何か、は最大の悩みであり、常に検証して最善のものにしていかなければなりません。幸運なことに、私は最初にとても良い研究テーマに出会ったおかげで、30年間、ほぼ基本的には同じラベル・目標をもって研究に取り組んできました。しかし、特に最近になって、ものすごい速さで世の中が発展しているのを見るにつけ、もう古くなっている部分もあると思い始めています。そろそろ新たなチャレンジに向けて大幅なバージョンアップをする必要があるかもしれません。仮に、65歳まで研究を続けることができるとして、私に残された研究者期間は約10年です。これは生き方にもかかわってきますが、この10年間の研究者人生を価値あるものにするにはどうしたらよいかを考え、研究者として10年後を見通して、社会や当該分野にどう貢献するかも考えていきたいと思っています。

私は、人間と同じように音を理解するコンピュータを実現することに、研究者としての興味、関心を持ち続けています。これは今後も変わらないと思います。人間にできることはコンピュータにも必ずできるはずですが、実際にはそれはなかなか難しいことですし、難しいからこそ面白いことでもあり、気付いたらライフワークのようになっていました。まだまだ解決すべきたくさんありますが、多くの研究者と協力して、あらゆる状況で人とコンピュータが同じ音空間を共有して、周囲の状況を理解しながら適切に会話し自由に協力し合える、鉄腕アトムのような

な世界を実現したいと考えています。

後輩の研究者に向けて一言お願いします。

基本は「仲間」と「自分のラベル」をつくることです。仲間はさまざまなシーンにいるはずで、例えば、職場には信頼できる、議論できる仲間がいます。国内の学会であればいつもライバルとなる仲間もいます。そして、海外であれば、最先端の研究をしている仲間ともリンクできます。繰り返しになりますが、研究活動も世の中の発展のスピードもかなり速くなっていますから、仲間と一緒に取り組んでいかなければ追いつけません。幸いなことに、私たちは海外の研究機関と共同研究をする機会が多くて、お互いを高め合える関係を築いています。

また、人生においても行き詰ったときに誰かに助けてもらうことや、楽しくやれるのは良いことだと思います。研究者は課題と向き合う際に自分と勝負をすることがありますから、どうしても閉じこもりがちになり、放っておくと1人になってしまいます。そうならないためにも、時に先輩方に背中を押しってもらうことも必要かもしれません。

さて、実際には研究活動は失敗ばかりです。前回のインタビューのときは自分の子どもが小さかったこともあり、子育てになぞらえて、失敗してもそこには必ず「研究のタネ」がある、といった話をしたかと思います。その気持ちは今も変わっていません。自分の期待を裏切る結果だったら、そこには自分が信じていたこととは違う自然の規則が存在しているはずで、それを徹底的に検証することで、次の「研究のタネ」を見つけるチャンスが目の前にある（かもしれない）ということです。経験が浅いときには検証するための材料が乏しいために判断に苦しむかもしれませんが、それでも、失敗を繰り返しながら結果と向き合って経験値を蓄えてください。失敗しないと成功はしませんから、失敗にへこたれず、「タネ」を見つけていってほしいと思います。