中谷智広　上席特別研究員　NTTコミュニケーション科学基礎研究所

研究は子育てと同じ．いくつもの失敗の中にある「研究のタネ」

音声認識を入力に用いるインタフェース．いまや「当たり前」の存在となってきました．しかし，現在，一般的に使われている技術では，周囲が騒がしかったり話者とマイクが離れていたりするとその認識性能が下がってしまうため，音声処理要素技術の高精度化に期待が集まります．適切に動作する音声認識の実現をめざし，昨年，世界一の技術を生み出したNTTコミュニケーション科学基礎研究所の中谷智広上席特別研究員に，最新の研究成果と研究者としての姿勢を伺いました．

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の実現

●現在手掛けていらっしゃる研究についてお聞かせください．

私は，ナチュラル音声入力インタフェースについて研究をしています（**図1**）．例えば，スマートフォン等を音声認識によって操作をした経験はありませんか．音声認識は従来のようにキーボードを打って情報を入力するよりも簡単で便利です．最近では，声で操作できるスマートフォンやタブレット端末が普及したこともあり，音声入力インタフェースの有用性が広く知られるようになりました．しかし，現在のスマートデバイスを用いた場合，マイクを話し手の近くに寄せて，丁寧に話さなければなりません．ところが，日常ではマイクの存在などを気にすることなく，人どうしは自由にコミュニケーションを図っています．

このような機器等を意識せずに話す場合でも情報にアクセスできたり，ロボット等と話せたりする仕組みをつくりたいと，私はNTTに入社以来ずっとこの分野の研究

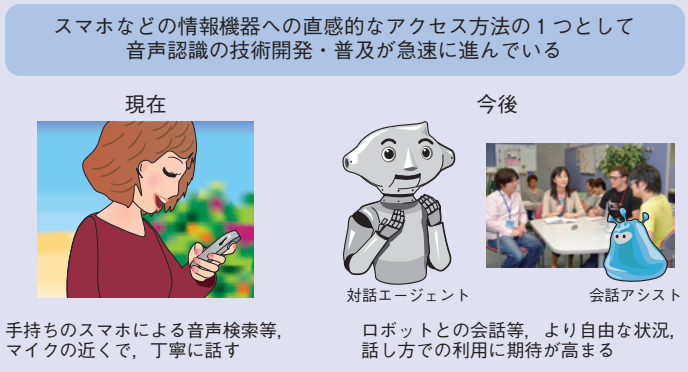

図１　ナチュラル音声入力インタフェース

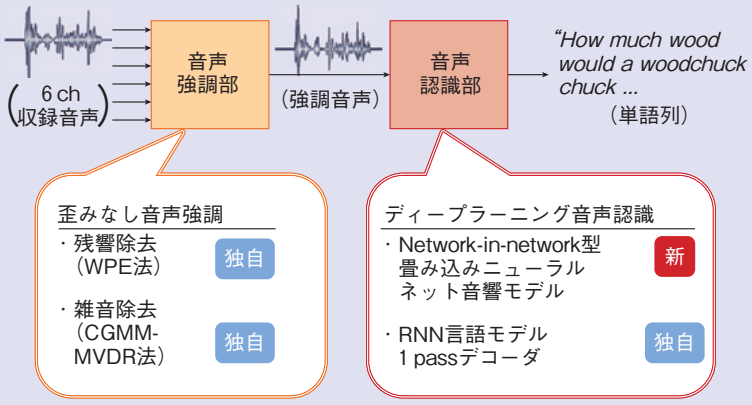

図 2　NTTシステムの特徴

を続けてきました．

研究が進んだ結果，現段階ではリビングの中に置かれたマイクに語りかけるとその指示どおりに家電が動く，といったスマートホームが現実味を帯びてきました．この仕組みのカギとなるのが，マイクから離れた場所で話しても音声を認識する技術です．日常的に複数の人がテーブルを囲んで会話をする場合でも正確にコンピュータが音声や情報を認識する時代はまもなく到来します．

世界一の技術を生んだエキスパートたちの新しい試みと切磋琢磨

●私たちの生活に大きくかかわる技術なのですね．

実は，NTTはすでに周囲の雑音や残響の中にあっても，より正確に音声を認識し，処理する高い技術を開発しているのです．そして，その技術に関して，NTTは世界を牽引しています．昨年，公共エリア雑音下でのモバイル音声認識国際技術評価（CHiME-3）で，2位と大きく差を付けて世界1位の音声認識精度を達成しました．これは，音声認識性能を劣化させる雑音や残響を，音声を歪ませることなく低減する技術や，雑音の影響を受けても精緻に音声をモデル化するディープラーニング音声認識技術の開発に成功した結果です（**図 2**）．世界1位となったこの技術評価において，私たちの誤認識率は約5%でした．ちなみに，従来の標準的DNN（Deep Neural Network：人間の脳を模した機械学習技術）音声認識の誤認識率は33%，そして，人間が聞いたらどれくらい間違うかを中国の大手検索エンジン会社が調べた結果が約11%ですので，かなり優れていることがご理解いただけると思います．この結果は，私たちの技術を用いれば，騒がしい環境において人間よりも聞き取れるという裏付けになりました．

では，技術について分かりやすくお話しします．複数の人が，室内で自由に会話をしている状況を思い浮かべてみてください．マイクを近くに置かずに録音を試みた場合，このような状況では音声認識の精度が低くなります．その要因は大きく分けて2つです．

1つは空調などの雑音や，壁などに反射した音声が少し遅れてマイクに届く残響の影響が大きくなることで，マイクで収録される音声の品質は低下します．また，ほかの人の声が重なり合って収録されることもあります．もう1つは，話し手がマイクを意識せずに自由に話すため，発音がぼやけたり，言葉を省略したりすることも多くなります．

このようなさまざまな要因に対処するためには，雑音や残響，ほかの人の話し声の影響を低減する「雑音除去」「残響除去」「音源分離」等のそれぞれの音声強調技術，そして自由な話し言葉を正確に認識するための「話し言葉音声認識技術」といった技術が求められるのです．この中で，私の主な担当は音を聞き分ける「音声強調技術」です．

CHiME-3では，結果的に1位になりましたので今となっては明るく振り返れるのですが，技術を出す前はどんな評価を受けるのか，正直，恐かったです．しかし，NTTの名前を背負って勝負するのだから何としても1位を獲得したい．自分たちなりに8%以内の誤認識率という目標を立て，新しく難しいタスクにチャレンジしつつ，何とか実績をあげることができたのです．

●こうした技術が誕生するまでにはどんなドラマがあったのでしょうか．

さまざまなドラマがありますが，最初にそのドラマの舞台となった，私が現在グループリーダを務めている信号処

理研究グループのことについて触れさせてください．このグループは，NTTの再編が行われた翌年の2000年にNTTコミュニケーション科学基礎研究所（CS研）に設置されました．この設置は，当時，CS研所長だった故東倉洋一先生や，初代のグループリーダの片桐滋先生（現同志社大学教授）らのご尽力によります．このグループの目標は，音声処理技術を基礎からつくり直すことで，従来の限界を超える新しい可能性を切り拓くことでした．グループには，音響処理と音声認識の2つの異なる技術の基礎研究者の精鋭が集められました．私は2001年にこのグループに一研究員として加わりました．

当初から，このグループの目標はナチュラル音声入力インタフェースの実現でした．その実現には，さまざまな挑戦的な課題を解かねばならず，かなりの時間を要する夢の技術と考えられていました．例えば，複数の人が同時に話していても聞き分けることができる音源分離技術，響く部屋で収録された音声から残響を取り除き音声を聞き取りやすくする残響除去技術，雑音が混ざった収録音から音声の発話区間を検出する音声区間検出技術，多人数会話の中で誰がいつ話したかを推定するダイアライゼーション技術，雑音が混ざった音声でも正確にモデル化し分類できる音響モデル技術，1000万語以上の超大語彙をリアルタイムで処理できる音声認識高速デコーダ技術など，多くの課題を乗り越える必要がありました．

それから約15年，挑戦的だとされていたこれらの多くの課題が解決されました．先日の「NTT コミュニケーション科学基礎研究所 オープンハウス2016」にて披露したように，テーブルに座ってお話しいただいた会話をテーブルの中央におかれたマイクで収録し，その場においてコンピュータで認識するデモシステムも体験していただけるまでになりました（**図3**）．実際に聞いてみてください．

図3　収録した会話がその場で認識されるデモの様子

●雑音や残響が消えて，音声がクリアに聞こえますね．

これらの発展を支えた1つの原動力，そして特徴として，私たちのグループには，音響処理と音声認識の2つの異なる分野の研究者が在籍し，席を並べ研究を進めてきたことがあげられます．一般の方にはこれら2つの技術の違いがよく分からないかもしれませんね．これらの技術分野は，一見近いと思われがちですが，基本技術や応用先が大きく異なっていて，それぞれ別々に発展してきたものです．実はこの2つの研究を一緒に進めているグループは，世界的にもほとんど例がありません．製品としても学術発表の場でも，両者が交流することはあまりありませんでした．世界に先駆けて，この境界領域に重点化し研究に取り組んできたことで，今までになかった新しい考えが次々と生まれ，NTTの音声研究の躍進につながりました．

この2つの分野の出会いの何が大切だったかについて，私が一番深くかかわってきた残響除去技術を例にとりお話しします．残響除去などの研究が古くから進められてきた音響処理の研究分野では，部屋の中をどのように音が伝播するかを数理的にモデル化し，そのモデルに基づいて，収録音から残響などの不要音を取り除く研究が行われてきました．しかし，どんな部屋で音が収録されたかが分からない未知の状況では，この方法を用いてもうまく残響を取り除くことができませんでした．一方，音声認識では，音声信号が取りやすい値の傾向をパターンとして学習し区別するパターン処理の手法（これは，今風にいうと，機械学習の枠組みといえるかもしれません）が広く利用されていました．そこで私たちは，この音声認識で培われたパターン処理の考え方を音響処理に取り込んだ「パターン指向音響処理」の枠組みを提唱し，さまざまな新しいアルゴリズムの開発に挑んだのです．これにより未知の状況における収録音の中で何が起きているかを自動で認識しながら，音響処理で用いられる数理的なモデルを用いて，所望の音の特徴を正確に分析できる技術が生まれました．こうした一連の流れの中で，残響除去の新しいアルゴリズムが，ここ京阪奈の地で2006年ごろに集中的に研究され，世界初の技術として実現されたのです．

信号処理研究グループがこれだけの成果をあげることができたもう1つの背景として，多くのNTTの偉大な先輩たちが切り拓いてきた基礎研究の道を引継ぎ，その延長線上で研究を続けてきたことがあげられます．NTTには，電電公社時代から，携帯電話の符号化方式の発明をはじめとして，世界の音声研究を牽引してきた歴史があります．また，NTT研究者と世界の研究者との間には過去から築か

れてきた強い信頼関係があります．これら偉大な先輩たちと（部分的にでも）世界観を共有しながら研究を進めてきたことは，私たちが基礎研究を進めるうえでとても大きな力になりました．

多くの失敗を検証して，「研究のタネ」を見つけよう

●研究を手掛けるときの着想，発想において大切なことを教えてください．

次の3点を知っておいてほしいと思います．まず，未経験のことを始めるときは思うようにいかないものだと心構えをしておくこと．そして，自分の期待を裏切るような場面に直面しても，その結果は何らかのかたちでその後の研究に活かせるということ．最後に，散々悩んだ挙句にアッという瞬間に求めていた結果が得られるということです．

例えば，京阪奈において世界で初めて実現された残響除去技術の研究も，はじめは失敗の連続でした．残響除去は，音響信号処理の分野ではかなり古くから重要課題とされてきたのにずっと解決することができなかった問題でした．その長い歴史の中で，残響除去を実現するためには，処理の結果，音声のスペクトルが平坦になって音声らしさが損なわれてしまう現象（音声の白色化と呼ばれます）が起きるのをどうやって防ぐかということが最大の課題と信じられていました．私たちもこれを解決するためにあれこれアイデアを出して試していました．しかし，音声の白色化を対策することで残響除去への多少の改善効果はあるものの，なかなか根本的な解決には至らない．本当に暗闇の中での一歩一歩でした．

そんな中で，私たちの研究の羅針盤になったものが，先ほどもお話ししました「パターン指向音響処理」でした．音声認識では常識であった，音の時間変化をパターンとして学習し区別する手法を導入し，残響除去の課題にアプローチしていきました．最初は幼稚な試みから始め，実験しては失敗を繰り返しているうちに部分的に成功するものが出てきては，少しずつ発見を積み重ねていきました．そして，長い間の試行錯誤の結果，ある日，突然，重要な発見にたどり着きました．「スペクトルの平坦化を防ぐことよりも，音声の時間構造が壊れないようにすることが大切なんだ」．これは，私の研究者経験の中でも，本当に目からうろこが落ちるような体験でした．この発見に従って，音声の時間構造を壊さずに残響を除去する仕組みづくりに研究がシフトし，残響除去の研究の急速な発展が始まりました．

●現役の研究者として成果を出すために大切なこと，若い研究者の皆様にも一言お願いします．

自分の期待を裏切る結果だったらそれを徹底的に検証することで，次の「研究のタネ」が見つかります．多くの失敗は面白いものですよ．なぜこうなるのだろうというきっかけをくれるからです．いつも失敗ばかりするという点で，研究と子育てはとても似ていると感じることがあります．私には小学校3年生と5年生の息子がいます．例えば，私が子どもたちに「もっと時間を守れるようになってほしい」と思ったら，決めたルールを守れたらご褒美をあげると約束します．しかし，決めたルールを1回目はうまく守れても，2回目はもう全然ダメなんてことはしょっちゅうあります．そんなときに，自分が期待していたことと実際に起きたことが違ってしまった原因は何だろうかと考えます．この感覚が研究で失敗したときに何が原因であったかを探るのと同じなのです．これをしっかり考えると，次はどうしようというアイデアが出てくる．そのアイデアを使ってうまくいけば良いし，ダメでもまた，次のアイデアを出せば良いのです．失敗とともに，少しずつですが前進していくことですね．逆に失敗するのは当たり前で，失敗してからが本当の勝負くらいに考えると，研究も子育て同様に楽しくやれるのではないかなと思います．

経験を積んでいないときは，検証をする材料が乏しい状況ですから判断をするのに難しいと感じるかもしれません．単に失敗をどんどんすれば良いわけではなく，成功したとしてもなぜ成功したのかと，結果と向き合い考えていく経験を増やしていってほしいと思います．

そして，興味の持てるものを見つけ出す努力をしてください．月並みな言葉ですが，興味のないことは長く続けられません．研究者として成功するためには自分が何に興味を持っているのか，自分の強みを知っているかが大切です．

●今後はどのように挑まれますか．

今回ご紹介した技術を用いれば，カフェや空港などの公共エリアでの音声インタフェースや，オフィスや家庭のリビングでの会話認識など，騒がしい場所で多くの人が話す場面でも快適に動作する音声認識が実現できます．これは，スマホの音声エージェントやコミュニケーションロボットの利用シーンの拡大に大きく貢献すると期待されています．例えば，耳の不自由な方や異なる言語を話す方々が今までできなかったコミュニケーションを支える技術であったり，人と人をつなげるような技術として成長させていきたいと考えています．そう遠くはありません．これらはほんの数年先の未来の姿だと考えながら今後も研究に挑んでいきます．