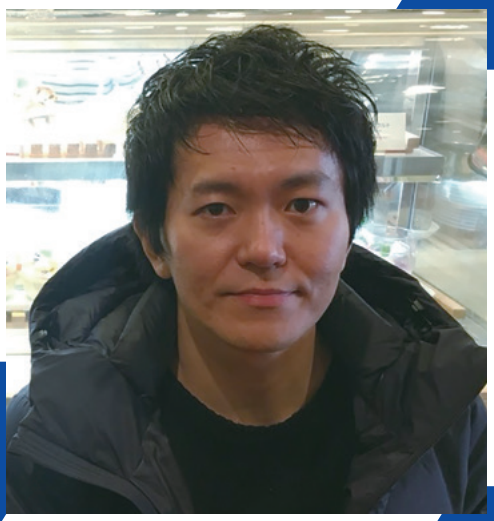
挑戦する
研究者たち
CHALLENGERS



亀岡弘和

NTT コミュニケーション科学基礎研究所
上席特別研究員

エレガントさを追究．研究ゴールという大きな傘をつくる

大学生約1800人を対象にした発音のしにくさに関する調査によると，約 3 割の学生が普段の会話で発音がうまくいかないと感じることが「ある」または「どちらかといえばある」と回答し，発音のしにくさを自覚する人は自分の音声が聞き返されることが多いと感じる傾向がありました．声や話し方を分析，合成，変換することで，コミュニケーションにおけるさまざまな制約の解消をめざす亀岡弘和NTTコミュニケーション科学基礎研究所 上席特別研究員に，現在取り組んでいる研究と研究者としての姿勢を伺いました．

音から状況を理解し，状況に合わせて声を変えて伝える技術開発

現在手掛けている研究から教えてください．

多様なシーンにおいて，人が不自由なくコミュニケーションできる手段の創出をめざして，音響信号の要素分解・情景分析技術，そして，高い品質と自然性を意識した音声生成技術の研究に取り組んでいます．

ミックスジュースから特定の果汁だけを取り出すのが難しいのと同じで，一般に何らかの混合物を要素成分に分解することは容易ではありません．しかし人間は多数の音からなる外界音から各音を聴き分けたり，人の話し声に混在するさまざまな非言語的な成分を読み取ったりすることで，音響的な情景を理解する能力を備えています．この能力は人間が社会生活を営むうえで，特にコミュニケーションにおいて重要な役割を担っています．音響信号の要素分解と情景分析の研究では，音を対象としたこのような要素分解機能，および情景分析機能を計算機に備えさせるための数理モデルやアルゴリズムの実現をめざしています．

外界音を対象とした要素分解・情景分析の研究では，混合音に含まれる複数の音を分離抽出する音源分離，対象音が何の音かを同定する音源同定，対象音がいつ鳴っているかを推定する音声区間推定，対象音がどこで鳴っているかを推定する音源定位，残響や雑音を取り除いて特定の音声

を強調する音声強調といった問題に取り組んできました．従来はこの分離，同定，区間推定，定位，強調の問題はそれぞれ個別の研究課題として取り組まれていましたが，よく考えてみるとこれらは独立しているわけではなく相互依存していることに気が付きます．例えば，何の音かがあらかじめ特定できていれば，それぞれを分離することは比較的簡単になりますし，それぞれの音をあらかじめ分離できていれば定位することが容易になるというように，１つの問題の解が他の問題の手掛かりになっています．この観点から私たちはこれらの問題を個別の問題としてとらえずに同時最適化問題としてアプローチし，まとめて解決する方法を考えました．

外界音に複数の音が混在するように，１つひとつの音声の中にもさまざまな成分が混在しています．私たちは会話の際，言葉に相当する言語情報とともに声の高低を用いて調子や意図などの非言語情報を相手に伝えますが，音声には言語情報に関係する音素，非言語情報に関係するフレーズ成分やアクセント成分などの要素が混在しています．声の高低の時間変化を表す基本周波数パターンは声帯に張力を与える甲状軟骨によって制御されているのですが，フレーズ成分とアクセント成分はその並進運動と回転運動に伴う成分とされています．これらの成分のタイミングと強度を正しく推定できれば非言語情報を定量化する重要な物理量となり得ますが，これらの逆推定は長らく難しい問題とされてきました（**図１**）．音声を対象とした要素分解・情景分析の研究では，基本周波数パターンをフレーズ成分とアクセント成分に分解する問題に焦点を当て，統計的信号処理アプローチにより高速かつ高精度に解決する手法を考案しました．この技術の応用例を紹介するため，標準イントネーションの日本語音声を関西風のイントネーションに変換するデモシステムを実装し，NTTコミュニケーション科学基礎研究所（CS研）オープンハウスやNTT R&Dフォーラムなどのイベントで実演したところ，なじみやすい内容だったからか多くのお客さまや報道関係者にも非常に好評で，テレビ，新聞，インターネット記事などで広く紹介していただきました．

これらの一連の研究は入社前の大学院時代も含めおよそ10年にわたって行ってきたもので，その間，ありがたいことに数々の表彰をいただきました．例えば2009年にIEEEからSPS Young Author Best Paper Awardを，2018年に科学技術分野の文部科学大臣表彰若手科学者賞をそれぞれ受賞しました．IEEEの賞は日本人初だったと聞いています．こうした受賞歴は研究分野でのプレゼンス向上につながりますので，振り返ってみると大きな出来事だった

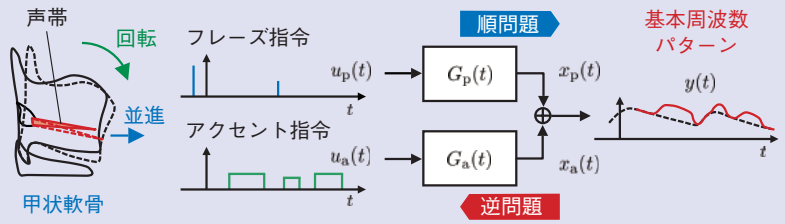


図１　音声の基本周波数パターンの生成過程とその逆問題

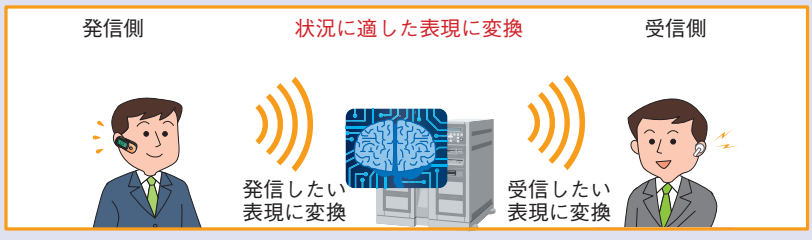
と感じています．

最近は，これらの研究に加え，深層学習のアプローチにより，高い品質と自然性を意識した音声生成の研究にも取り組んでいます．特に力を入れているのは，声の高低パターンだけでなく，声質やリズムといったさまざまな声の特徴を柔軟に変換することができる音声変換技術の研究です．不慣れな言語での会話，緊張状態でのプレゼンテーション，発声機能や聴覚機能に障碍や衰えがある場合の会話など，思いどおりに円滑にコミュニケーションを行うことができない場面は多くあります．音声変換技術の研究を通じて，円滑なコミュニケーションを阻害し得るさまざまな制限を取り除くことをめざしています（**図 2**）．

CS研オープンハウスでいただいた質問から得た新たな視点

大きな成果が得られたのですね．最近の音声変換の研究についてもう少し詳しく聞かせてください．

音声変換の研究をより深く追究しようと考えたのは，実は先ほどお話ししたCS研オープンハウスでのある出来事がきっかけになっています．当時展示で紹介していたのは音声の「アクセント」を変換する技術だったのですが，見学にいらした方から「英語のアクセントも変換できるのか」という質問をいただいたのです．一般用語的には「アクセント」というと「訛り」をさし，英語の場合は発音の訛りをイメージすると思うのですが，まさにその意味でとらえられていたわけです．紛らわしいことに「アクセント」は，音声研究の分野では基本周波数パターンの中の１成分を意味する用語としても使用されていて，私たちが当時意図していたのはこちらの意味だったのです．実は，英語の発音の訛りを変換するには，基本周波数パターンを変換するだけでは不十分で，声の音色に相当する特徴である「声質」を変換する必要があります．つまり当時紹介していた技術



開発中の技術による具体的な変換例

- 混合音声 ⇒ 聴取したい声
- ある話者の声 ⇒ 異なる話者の声
- 訛りのある発音 ⇒ 聞き取りやすい発音
- 発声障がい者の音声 ⇒ 健常者の音声
- 顔画像＋音声 ⇒ 顔表情に合った話し方

図 2　コミュニケーション機能拡張技術のイメージと具体例

では英語の発音の訛りの変換は扱えなかったのです．しかしこれがきっかけで，英語の発音の訛りを変換する問題に少し興味を持ち始めたわけです．音声の分野では入力音声を別人の声に変換する声質変換という技術があるのですが，この問題におそらくもっとも関連するのはこの技術だろうと見込んで，声質変換の既存手法を実装し，実験してみたのでした．ところが，既存手法で変えられるのは話者性のみで，発音の訛りなどの発話様式までは変えられないことが分かったのです．このとき，この問題は想像以上に奥が深く面白いなと感じ，もっと本格的に音声変換の研究に取り組んでみようと思い至りました．ちょうどそのころ，深層学習（いわゆるAI）が台頭してきた時期で，深層学習アプローチを用いた音声合成・変換の研究をスタートさせました．そこから4年が経過した現在に至るまで，同僚や実習生とともにさまざまなアプローチの検討を進め，英語の訛りだけではなく，より広範な音声特徴の変換を可能にする高品質で柔軟な音声変換技術を多く創出しています．

こうした音響信号や音声生成を手掛けようと思ったのはなぜでしょうか．

学生時代，趣味で，バンドでギターを演奏していました．好きな曲があると，音を聴いて譜面を書き起こすいわゆる「耳コピ」を行ってから練習することが多かったのですが，卒業論文ではせっかくなので趣味に関係するテーマに取り組んでみたいということで，耳コピを自動的に行うアルゴリズムの研究をやりたいと考えました．そこで，音を扱う研究室の門を叩いたのがきっかけです．当然，研究分野のことは何も分からない状態でのスタートでしたが，大学の講義で習った信号処理や統計の知識をどう活かせるのかを想像してわくわくしたのを覚えています．また，当時は歌もうまくなりたいという憧れがあったのですが，歌うときの自分の声があまり好きになれず，どうしても自信をもてませんでした．そのため，耳コピの自動化以外にも自分の声を自動的にいい声に変換できないかということも考えたりしていました．音声の研究者は音楽が趣味の方が多く，似たような理由で研究を始められた方も結構いらっしゃいますが，私の場合もこういった単純な動機が今の研究につながっています．今振り返ると，耳コピを自動化したいという動機も歌声を良くしたいという動機も，人間の聴覚機能や発声機能をサポートすることをめざしている現在の研究に通ずるところがあるなと感じます．

音楽活動では耳コピや楽器演奏や歌等の音楽的スキルによってできることが制約されてしまいますが，これと同じように，私たちの日ごろのコミュニケーションにおいても，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ています．今の私の関心は，このような制約を機械学習（AI）や信号処理の力により取り除き，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にあります．このためには，送信者と受信者が置かれている状況や環境をとらえる情景分析技術と，送信者が受信者に伝えたい情報を状況に適した表現に変換するメディア変換技術がカギになると考えています．さらに，音だけでなく動画像やテキストなどの多種のメディアを有効活用した新たなコミュニケーション方式の可能性を模索し，それを実現するための基盤技術を創出していきたいと考えています．

大きな傘はできることやりたいこと，求められることの幅を広げてつくる

上席特別研究員となられてから，何か変化はありますか．

実際の研究生活そのものには今のところあまり大きな変化はありませんが，意識として次の２つに努めていきたいです．１番目は，これまでの研究者人生において大切にしてきたことでもあるのですが，「できること」「やりたいこと」「求められていること」の幅をさらに広げていくことです．２番目は，大きな研究ゴールの「傘」をつくることです．この傘の重要さは，私がNTTに入社したときに配属されたグループで学びました．誰もが世の中の役に立つと納得するような明確な研究のゴールがあることで，目前の困難な課題に専心しつつも，進むべき方向に確信をもてるようになります．研究活動は研究成果を地道に積み重ねる作業で，その１つひとつはいつも大きい成果とは限らないものです．そのため時として「自分は大したことをやっていないのではないか」という不安にふと駆られることがあります．私はそんなとき，先輩たちがつくってくれた「傘」の下で安心して研究することができました．まさに，研究ゴールという傘に守られているようなイメージです．ですから，私も後輩のために一点の曇りもなく，携わる誰もが安心し，躊躇なく邁進できるゴールをつくりたいという思いが上席特別研究員となってより強くなっています．こうした，自身の幅を広げる，傘をつくる視野を持つことは，いずれも成長し続けることでのみ可能になることだと考えているので，現状に甘んじて立ち止まることなく，より一層精進する所存です．

逆に，変化しないように心掛けているものもあります．それは「エレガントさを追究する」研究スタイルで，私の研究者としてのポリシーともいうべきものです．これは，大学時代，NTT出身の指導教員の研究スタイルやお考えにかなり影響を受けたところが大きいです．エレガントさとは，はっきりと定義するのは難しいのですが，「本質を見抜いた鮮やかなアプローチ」に対する美的感覚のようなもので，数学で美しい問題・解き方・証明に出くわしたときに覚える感覚に近いかもしれません．何事もエレガントさを追究しながら取り組むことで，思考が研ぎ澄まされ，その積み重ねにより高みに登ることができるはずと考えています．私自身も研究者生活で手ごたえを感じたことはいくつかありましたが，まだまだと感じていますので，これからもエレガントさにはこだわっていきたいですね．あと，プレイヤーとして研究するのがやはり好きなので，可能な限り主著の論文もたくさん書き続けられるように頑張りたいです．

後輩の研究者に一言お願いいたします．

研究活動はハードで精神的にきついときが多いですが，NTTで研究できる喜びを噛み締めて，とにかく研究を楽しむことが一番大事なのではないかと思います．気分が落ち込んでいるときや邪念に支配されて思考が停止しているときは研究が進みませんし，逆に気持ちが乗っているときは研究も進むものです．心の持ちようで研究のスピードが変わりますから，心を安定させて前を向くことが大事です．例えば，誰かを負かしたい，周囲に自分を良く見せたいとつい思ってしまいがちですが，これは他者を意識しすぎるあまり妬みや焦りといった負の感情に心が囚われている状態なのだと思われます．自分ではコントロールできない問題にはあれこれ悩まず，自らが日々成長しているかという点に意識を向けてみると良いと思います．自らの成長を楽しみにすることで，壁にぶつかってももう少し頑張ろうという気持ちになれると思います．

一方で，研究に没頭しすぎているときも注意が必要です．一見そのときそのときはすべてが順調に進んでいるように思えても，後で振り返ってみると停滞していたと感じることはよくあります．私も陥りがちですが，そういうときは

大抵視野が狭くなっています．客観的に見てそれほど重要ではないことを重要と信じて心血を注いでいる状態です．油断しているといつでもそのような状態に陥ってしまいかねないので，目の前のことに粘り強く打ち込める集中力とともに，常時客観的に自分を見つめる冷静さを養っておくことが必要です．諦めずこつこつ取り組む人格と冷静に自己を客観視する人格を備えて，意識的にこれらの人格間で対話をし続けることが重要なのです．

私は，研究者は世の中を良い方向に動かすための知恵を出す存在ととらえています．NTTの研究者はもちろんNTTのために研究をしますが，よりマクロな視点でみると，すべての研究者の共通の使命は世の中をより良くすることです．他者との競争に負けない逞しさも必要ですが，同じ使命を担う他者の研究や貢献へのリスペクトを欠かさない誠実さも大切にしていただきたいです．

今後のさらなる目標，展望を教えてください．

人間のコミュニケーションをいかに円滑化するかに主眼を置いた研究を引き続き進めていきます．その先に見据えているのは，メディアをまたいだ変換を実現することです．音声，テキスト，動画像はそれぞれ異なる特長を持つメディアです．例えば音声はメッセージを素早く表現して相手に伝えたいとき，テキストは素早くメッセージの要点を読み取りたいときにそれぞれ有効です．また，動画像は音声やテキストでは表現しきれない細部の情報を表現できる点が強みです．このようなそれぞれのメディアの特長を活かし，送信者と受信者が，置かれている状況に合わせて使用メディアを柔軟に選択できるようにすることで，極めて高効率で円滑なコミュニケーションを実現できるようになると考えています．このためには，各メディアの信号を，メッセージやコンテンツを保持するように異なるメディアの信号に変換するクロスメディア変換を扱う必要がありますが，これはとてもチャレンジしがいのある面白いテーマになると思っています．

研究者としては，センスの良い研究テーマとエレガントなアプローチを追求していきたいです．各分野にはいわゆる花形の研究課題があります．重要な一方，難攻不落で，永年多くの研究者が挑戦を続けている課題です．こういった研究課題では大抵ベンチマークや評価系やデータセットがすでに確立されていて，研究や実験に比較的着手しやすいメリットがあります．一方で，多くの研究者が競い合っている中で違いを生み出すにはかなり高度で専門的な知識と技術レベルが必要になります．逆に，誰も着手していないような新しい課題を開拓する取り組みもまた重要です．分野に新たな世界と価値を生み出せるからです．しかし，場合によっては評価系やデータセット等をゼロから構築する必要があるため研究の立ち上げにはかなりの労力を要します．高い専門性と柔軟な発想力を養い，両タイプの取り組みを両輪としてうまくバランスを取りながらさまざまな課題の解決に向けて邁進していきたいと考えています．

■参考文献

⑴ https://www.jstage.jst.go.jp/article/jasj/75/3/75_118/_pdf
⑵ 亀岡：“音声のイントネーションとアクセントを分析，合成，変換，”NTT技術ジャーナル，Vol.27, No.9, pp.10-12, 2015.