

挑戦する 研究者たち CHALLENGERS



守谷健弘

NTTコミュニケーション科学基礎研究所
フェロー



研究者は社会の進歩の源泉である。自らの研究テーマに自信を持って「その次の主役」をめざそう

コロナ禍をきっかけとして、リモートワークが定着してきましたが、そこではPCやスマートフォンを利用して画像や音声によるコミュニケーションが重要な役割を果たしています。約40年にわたり音声・音響信号符号化の研究に携わり、広い意味での通信や生活の質の革新的向上をめざすNTTコミュニケーション科学基礎研究所 守谷健弘フェローに、音声・音響信号符号化領域の技術進展の変遷と研究者としての姿勢を伺いました。



音声・音響信号符号化を追究

40年あまり継続している研究と変遷を教えてください。

音声や音楽の信号をデジタル化し、それを効率良く、かつ再生時の再現性を高く情報圧縮する「音声・音響信号符号化」を追究してきました。例えば、私たちが携帯音楽プレーヤーやデジタル放送などで楽しんでいる音楽は、本来の信号を単純にデジタル化したものではなくデジタルデータ量を10分の1程度に圧縮したものです。

さて、電電公社では、電話の限られた帯域で伝送される、

品質を含むアナログ音声の研究に注力し、ITU-T (International Telecommunication Union Telecommunication Standardization Sector) の前身であるCCITT (Comité Consultatif International Télégraphique et Téléphonique) における国際標準化にも大きく貢献してきました。電話網・端末のデジタル化の研究の進展とともに、アナログ音声のデジタル化、圧縮に関する研究も進み、これがNTTにおける音声・音響信号符号化研究の源流となっています。中継網がデジタル化され、1984年に「INS ネット64」「INS ネット1500」がサービス開始しそれが普及してくると、音声・音響信号符号化技術

がますます重要なものとなってきました。1990年代になり、デジタル化された第2世代の携帯電話が登場しましたが、固定網と比較して伝送ビットレート等の制約が厳しい中で品質を確保するために音声・音響信号符号化技術への期待値が高まり、その重要性が増大してきました。こうした状況の中の研究成果として、低ビットレートで伝送符号誤りがあっても音質を担保できることなどの条件をクリアすることを可能とし、各方式によるコンテストの結果、第2世代携帯電話の日本標準方式として採用されました。その後、第3世代の携帯電話やIP電話においても私たちの成果が要素技術として採用され、世界中の携帯電話の音声品質の向上に貢献しています。

全世界に影響を与える研究成果を上げられたのですね。

全世界への影響という意味では、1990年代からNTTのチームはいくつかの国際標準化に貢献してきました。ITU-T規格ではIP電話用符号化標準、ISO/IEC MPEG規格では低ビット音楽符号化標準、歪みを許さないロスレス符号化標準などがあります。また2010年ごろ、移動通信システムの国際標準化団体である3GPP (3rd Generation Partnership Project) において音声符号化標準規格の制定が開始されました。世界統一の第4世代の移動通信システムにおける音声通信方式であるVoLTE(Voice over Long-Term Evolution)向けに新たな音声符号化技術の制定が強く望まれたからです。これに対し、NTTグループをはじめ、世界の多くの専門家による競争と協力によって、2015年に音声・音響信号統合符号化方式としてEVS(Enhanced Voice Services)が国際標準化されました。

それまでの携帯電話向け音声符号化方式は、人間の発声メカニズムにならったCELP(Code Excited Linear Prediction)という方式が用いられ、人間の声を低ビットレートで高品質に伝送してきました。EVSではCELPに加えて新たに開発された低遅延の音楽向け符号化方式を組み合わせることで、それまで実現できなかった背景雑音や背景音楽を含む音声あるいは音楽を高品質のまま低遅延で伝送できます。標準化の過程では、第三者機関によ

てさまざまな条件・音源・言語での大規模な主観品質評価試験が行われ、従来方式よりも格段に高品質であることが確認されました。

これにより、EVS方式は世界の通信事業者、通信機器メーカー、チップメーカーに一齐に採用されました。これにより、現在世界中で利用されているスマートフォンの通話は、通信事業者、通信機器メーカーにかかわらず、広帯域で高品質になりました。これは、長きにわたって電話の音質を高めるためにNTTのチーム、世界中の有力な研究者やエンジニアと繰り返した試行錯誤の結果なのです。



IVAS標準規格の制定に挑む

最近でもさらに音質の重要性が問われていると聞きました。

新型コロナウイルスの感染拡大の影響で、Web会議などが急激に増加したことはいうまでもありません。こうした現状についてNTTデータ経営研究所とShureが、オンライン会議で行き交うデジタル音声の音質の違いがもたらす生体ストレス反応への影響を検証する実証実験を実施したところ、85%の利用者がWeb会議の音質に不満を抱えていることが分かりました。具体的な不満内容は、音質が悪い会議では、内容が理解されないだけでなく参加者にストレスを与えるというのです。

3GPPにおける国際標準化等により、スマートフォンどうしの通話は高音質になったものの、Web会議などではPCを通じて会話がなされており、通信品質がベストエフォートなインターネットが利用されているため、遅延やパケットロス等によりいまだ音声品質は不安定です。また、PC等のWeb会議のアプリにおいても処理遅延やパケットロスへの対応が不十分で品質劣化は否めません。一方、社会ではメタバース等の新たなコミュニケーションの場も含め、あたかも同じ場で会議しているかのような臨場感に関する技術が求められるようになってきました。臨場感については、高精細な画像をイメージすることが多いと思いますが、実は音声・音響の品質等に依存するところが大きいのです。こうした背景から、EVSを拡張するIVAS (EVS



(EVS Extension for Immersive Voice and Audio Services)
*2023年中の仕様凍結めざし、オープン開発中

- 立体音場の収音、圧縮伝送、再生による**高臨場感双方向通信**
- * EVSはモノラル通信
- 複数ストリームと再生合成機能による**多地点双方向通信**
- * EVSは1対1の地点の通信
- 相互接続機能による**EVSとの符号変換なし双方向通信**
- * EVSは世界中に普及

図1 IVASの目標



図2 音声品質を保証する高機能電話

extension for Immersive Voice and Audio Services) の標準規格制定に臨んでいます。IVASは3GPPにおける2023年中の仕様凍結をめざしてオープン開発中で、立体音場の収音、圧縮伝送、音声による高い臨場感と双方向通信、そして、複数のストリームと再生合成機能による多地点双方向通信、さらに相互接続機能によるEVSとの符号変換なしの双方向通信をめざしています(図1)。

3GPPではネットワークの専門家がこれまでに構築してくれた、ひかり電話(VoIP)網や移動通信(VoLTE)網のような音声通話に適した高信頼通信の利用を実現しつつ、多様な機能拡張が自由にできるWebRTCとの連携プロトコルの設計もめざしています。これにより現在のWeb会議より安定で高品質の音声通話が保証されつつ、逐次高機能化、高臨場化、さらにXR(Cross Reality)やメタバースなども含む多様な通信形態への進展が期待されます(図2)。

新しい技術が次々と生まれる中、臨場感に大きな影響がある音声技術の追究を続けていらっしゃるのですね。

NTTグループは携帯電話サービスも固定電話サービスも提供する企業として、音声関連の技術に責任を持って貢献していくことが大切です。私はその領域を追究してきました。特に電話の音声の研究について重厚な歴史を持つNTTの研究者として、ビジネスシーンに活かせるものをつくり上げたいと研究に臨んできたのです。

NTTグループが2024年の仕様確定、2030年の実現をめざして研究開発を進めているIOWN(Innovative Optical and Wireless Network)において、高速・大容量化技術や高精細画像技術等の華やかさと比較して、音声技術はやや地味に感じるかもしれませんが、私は、高臨場感に与える影響の大きい音声分野において、媒体に左右されない音質の良い通話を最優先に実現したいと考えます。高速・大容量だけでは音質は必ずしも良くなるからです。

現在は例えばAI(人工知能)の研究は大流行していて、多くの研究者がレッドオーシャンでしのぎを削っています。しかし、その中で頂点に立っているのはほんの一握りの研究者です。一方で、世の中には流行に左右されない大切な領域がたくさんあります。その中で興味を持って臨める研究を貫いていくのも、研究者としての1つのあり方だと思います。音声に関する研究は現在では決して華やかなテーマではありませんが、非常に重要なテーマであることは確かです。私は、そこに興味を持って研究を貫いていくことを選んでいるのです。解決すべき課題も多くあります。私の時代では解決できない課題は、先達から受け継いだ匠の技を後進につなげ、解決に導くための道筋をつけていきたいと思っています。

現在、約40年の歳月を経てAIが3度目のブームを迎えています。このように1つのテーマに関する流行はそう長くは続かず、必ず次の波がやってきます。その波を見据えて研究テーマを見つけるのもいいのではないのでしょうか。音声の分野においても1990年代の携帯電話のデジタル化に向けて、私が手掛けてきた音声のデジタル圧縮符号化技術はその重要性が再認識されました。自己満足はいけません

が、自らの研究テーマに自信を持って「その次の主役」をめざして頑張りたいものです。



音声や音質の追究は「命」をもつなく

自らの研究テーマに自信を持つのは大切なことなのです。

世の中に注目されるテーマ、流行に乗っているテーマである以上に、世の中に貢献することのできるテーマが大切だと思っています。その意味において、例えば、交通事故における死亡者数は1995年から格段に減少しているのですが、これには携帯電話の普及によって、事故が発生してもすぐに救急車を呼ぶなどの対応ができるようになったことが影響しているに違いありません。また、コロナ禍にあって、病院や養護施設へ見舞や面会に行くことが制限されています。こうした状況下において、面会や見舞を制限されている場所でも、携帯電話等を使ってコミュニケーションを図ることで遠隔から誰かを励ますことができるようになり、これにより救われた人も多いと思います。携帯電話の普及においては、先達をはじめ、私たちが取り組んできた音声や音質の追究が寄与している部分もあり、それにより単に社会を便利にただけではなく、「命」をもつないでいるのだと実感することができました。

通信の高速化・大容量化に伴い、音声のデジタル圧縮符号化技術の研究領域は先細りするともいわれていますが、必要のない研究ではありません。現に3GPPにおける国際標準化をはじめ、Beyond 5G、6Gを見据えて、研究や開発に取り組んでいる最中です。もちろん世の中への貢献をめざしていることはいうまでもありません。企業の研究者ですから会社のためになることをするのは当然のことですが、さらにその先の世の中への貢献により、自らの研究成果を使っている人の笑顔を思い浮かべて研究に勤しみたいですね。

社会にとって、研究者とはどのような存在だとお考えですか。

研究者は社会の進歩の源泉だと思っています。よく政治家と研究者を比較してお話しをするのですが、例えば、

100しかないものを200人が必要とした場合、政治家であれば、200人のうち真にそれを必要とする人に与えるか、あるいは半分にして全員に分け与えることを考えるかもしれませんが、一方で、研究者は100を200、あるいは1000にするような価値創造、クリエイションに挑みます。これはまさに私が携帯電話のデジタル化の波がやってきたときに電波を倍の効率で使うために、音声を圧縮して情報量を半分にしたのと同じ思考です。

さて、定年を過ぎた私が、フェローとして音声の研究を続けているということは、NTTが音声の研究を重要視していることの証だと思えます。だからこそ、これからは、後進の研究者の皆さんに向けて、いかに音声の研究が重要であるかと価値を説き、魅力を伝えることで音声をテーマとした研究を継続させていくことに努めていきたいと考えています。そのうえで、社会がその重要性をまだ認知していなかったとしても、これは重要であると信念を貫く研究活動はとても重要です。

研究者として、研究の流行に敏感でありながら流行に流されないこと。そして、深く追究することは重要であるけれども重箱の隅を楊枝でほじくるような研究をしないこと。論文を書くことは大切だけれども論文が書ければいいからと自己満足で研究を終わらせないこと、といった心掛けが大切だと考えます。後進の皆さんには、信念を貫けるテーマを掲げて時流を読みながら息の長い研究者となってください。