

NTTコミュニケーション科学基礎研究所
上席特別研究員

原田 登 Noboru Harada

「世の中の役に立つこと」をめざして、基礎研究と国際標準化活動の両輪でそれを実現

リモート会議の普及やメタバースという言葉がメディアを賑わせるようになってくると共に、映像と音声を活用したコミュニケーションに社会の注目が集まるようになってきました。こうしたコミュニケーションにおいては、映像に関心が高まる傾向にあります。臨場感をはじめとして音声が大きな役割を果たしています。音声・音響コミュニケーションの研究に長期にわたり取り組み、「伝えたい音・聴きたい音」を選択的に伝える等、「自然で高機能な音響コミュニケーション技術の研究と標準化」を新テーマに研究を行っているNTTコミュニケーション科学基礎研究所 原田登上席特別研究員に研究のめざす方向、研究成果、国際標準化活動と研究者としての姿勢・考え方を伺いました。



音声・音響コミュニケーションに関する研究と国際標準化で世の中に貢献

現在、手掛けていらっしゃる研究について教えていただけますでしょうか。

2022年に「自然で高機能な音響コミュニケーション技術の研究と標準化」という新しいテーマを設定し、人にとって自然な多地点会議や、伝えたい音・聴きたい音だけを選択的に伝える、高機能イマーシブ通信と、個人や状況に合わせて、自然かつ自動的に体感品質（QoE：Quality of Experience）が改善される音響コミュニケーション環境の実現をめざしています。

人どうしが対面で会話をしているとき、周囲の雑音がある中でも、話し相手の声を選択的に聞き分けています。これが在宅のリモート会議の場合、参加者に聞かれたくない周囲の生活音までマイクが忠実に拾い、それが相手のスピーカから流れることで、対面と異なり好むと好まざるとにかかわらず生活音まで意識せざるを得なくなります。逆に遠く離れて暮らす家族や孫とのリモートコミュニケーションでは、この生活音によりコミュニケーションの親密度が高まります。このようにコミュニケーションにおいては、伝えたい音・聴きたい音があります。

また、資料のプレゼンテーションの場合を例にとると、リモート環境においてはリモート環境としてのコミュニケーション手法があり、対面のリアル環境においてはリアル環境としてのコミュ

ニケーション手法があります。リモート環境はコロナ禍により普及が加速しましたが、コロナ禍の終息に伴い、リモート環境とリアル環境が混在するような場面も数多く登場しています。このようなハイブリッドな場としては、ハイブリッドな環境としてのコミュニケーションの手法の実現が課題になります。

さて、目的や状況に応じて、伝えたい音・聴きたい音だけを選択的に伝えることで、こうした課題へ対応することができます。高機能イマーシブ通信の実現に向けては、音響イベント検知、音源方向推定などの音環境理解技術を応用し、入力される音響信号の意味的構造を理解して、メタデータを含む音響オブジェクトを抽出することが必要となり、さらにこのような抽出を行うための表現をデータから自動的に獲得することが新しいテーマのキーテクノロジーとなります。

そして、抽出された音響オブジェクトを再構成し、伝送やVR（Virtual Reality）・AR（Augmented Reality）再生などへの応用の検討を行います。AR・VR通信への応用の場合、複数の音が混在する入力の中から、音の意味的な塊である音響オブジェクトを分離し、クラス情報・空間情報とともに出力する技術であるSemantic segmentationを実現します（図1）。それに加えて、ユースケースにより必ずしも高品質を必要としない、あるいはすべての音響オブジェクトを抽出する必要のない場合もあるため、人の聴感品質評価基準のモデル化によりユースケースに応じたQoEを推定し、それに合うように品質や抽出対象音響オブジェクトを自動改善・チューンすることで、高機能イマーシブ通信が実

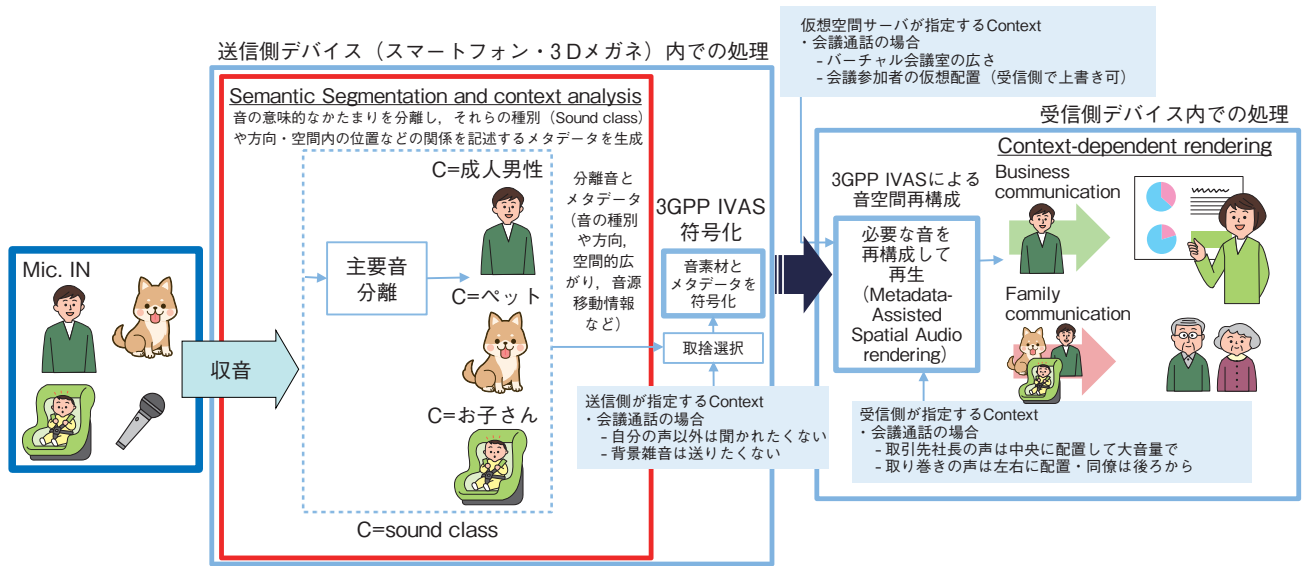


図1 Semantic segmentation技術のイマーシブ通信への応用例

現されます。

これらの技術は、会議通話に限らず、遠隔診療・見守り、重機の遠隔操縦など、遠隔で音を聞くさまざまな利用シーンで活用できると考えており、こうしたユースケースを意識しつつ、国際標準化などを通じて、世界を巻き込んだ実用化をめざします。

新しいテーマということですが、これまでどのような研究をされてきたのでしょうか。

私は、1997年にNTTに入社したのですが、それ以降、事業会社での開発業務や研究マネジメントの仕事をしていた時期を除き、音響音響信号処理、符号化とその国際標準化に取り組み、近年では自己教師あり学習などを用いた表現学習と、それらを応用した音響イベント検知や異常音検知などの音環境理解も専門分野として研究を行ってきました。

その中で代表的な成果として、音響信号のロスレス符号化方式であるMPEG (Moving Picture Experts Group)-4 ALS (Audio Lossless Coding) *への技術提案と標準化などを行ってきました。この技術はNTTエレクトロニクス社製エンコーダ装置に実装され、ハイレゾ音楽配信などに使われています。2018年には、機械等が発する音から異常を検知する技術をNTTデータなどの事業会社と協力して実用化し、地上で収録する列車の走行音

* MPEG-4 ALS : MPEGにおいて、MPEG-4オーディオの一部として規格化された。圧縮前のデータと、圧縮・展開の処理を経たデータが完全に等しくなるデータ圧縮方法であるロスレス圧縮方式。

を解析して車両の異常を検知する車両異常の監視や、発電設備など、機器の損傷や劣化を判別・検知する技術として活用されています。

国際標準化においては、ISO/IEC (International Organization for Standardization/International Electrotechnical Commission) MPEGではMPEG-AのAd Hoc Group議長として音声符号化におけるMPEGアーカイブフォーマット、エディターとしてMPEG-4ロスレス符号化技術の標準化、IECではプロジェクトリーダー・エディターとしてIEC61937-10伝送標準規格、ITU-T (International Telecommunications Union -Telecommunication Standardization Sector) ではエディターとして音声符号化に関する標準規格G.711.0、IETF (Internet Engineering Task Force) ではエディターとしてリアルタイム通信プロトコルにおけるペイロードに関する標準規格RFC 7655等の国際標準化に貢献してきました。

また、2014年に移動体通信に関する国際標準化団体である3GPP (3rd Generation Partnership Project) において、すべてのスマートフォンに実装され、NTTドコモのVoLTE HD+ (Voice over LTE High Definition+) サービスに採用されている高品質音声符号化技術EVS (Enhanced Voice Services) の国際標準化にも貢献し、現在では、双方向イマーシブ通信を想定したオブジェクトオーディオ符号化技術で高品質多対地電話を実現するIVAS (Codec for Immersive Voice and Audio Services) 規格の国際標準化にも取り組んでいます (図2)。

さらに、2020年には音環境理解の主要国際会議DCASE (Detection and Classification of Acoustic Scenes and

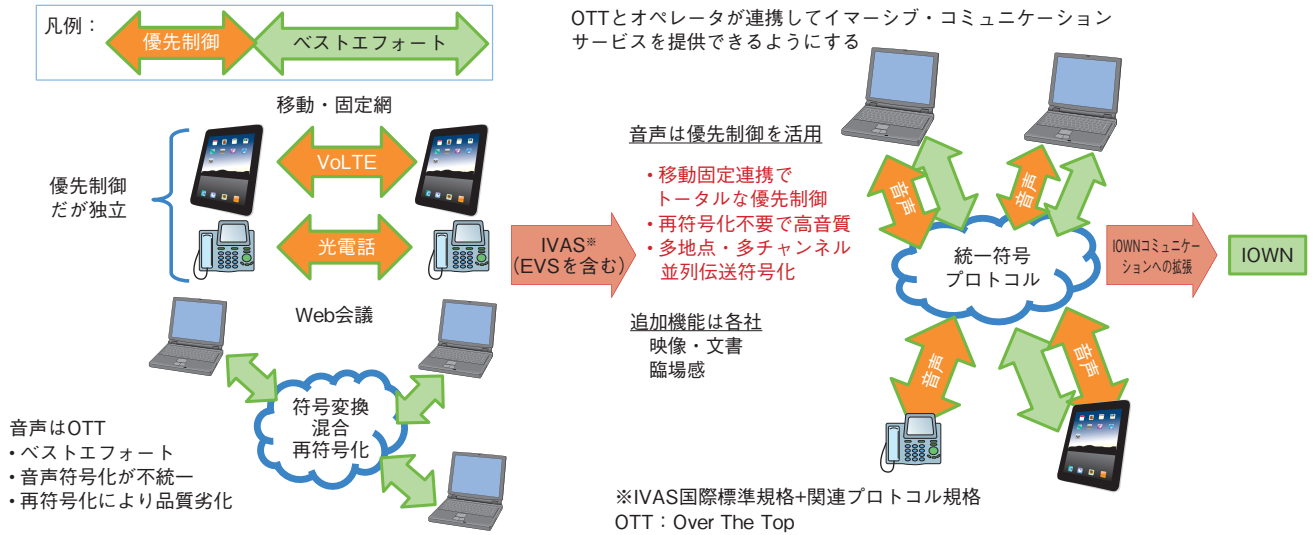


図2 NTTにおける3GPP IVAS標準化のねらい

Events) においてジェネラルチェアを務め、言語と音の距離モデルを提案するとともに、併催の音説明文生成コンテスト(DCASE Challenge) において世界一を獲得しました。

基礎研究の成果を国際標準化をとおして実用化

基礎技術の研究から国際標準化まで幅広くご活躍されてきたのですね。こうした中で研究者として大切にしていることや、それを教えていただけますでしょうか。

私の研究では、実応用を想定して研究課題を見つけ、要求される制約条件や事象をよく観察し、物理的観点、心理的観点、数理・情動的観点から総合的に検討して理に適った仮説モデルを作成します。このモデルを活用することで、実用化に向けて役に立つために必要な機能や、それを実現するための解法を考案し、国際標準化の会議などへの提案を通じて実際に使われるところまでを研究のゴールとして取り組んでいます。このモデルは、状況に応じて精密で複雑なモデルとしても、簡単で、シンプルなモデルとしてもつくることのできることで、ユースケースに応じて実用上必要とされる精度等の要求条件を考慮して適切なモデル表現を探します。

実用面からの要求条件を検討する際、例えば異音検知の場合は発想の転換も図りました。一般的には、音楽や音声は信号で、背景雑音等が不要なノイズとなりますが、異音検知の場合は機械の発する異音(ノイズ)を信号としてとらまえ、正常な機械音をノイズとしてとらえることで、これまで信号と思っていたものをノ

イズとして扱い、ノイズだと思っていたものを信号と見なしてモデル化を行う必要があります。

着目する事象をどのようにモデル化するかという観点において、ロスレス符号化やEVSで用いられている技術とベースとなる考え方は共通なのです。

さて、私たちは企業の研究所として基礎研究から実用化まで取り組んでおり、世の中の役に立つ研究を行い、最終的に何らかのかたちで人類に貢献したいという思いを持っています。そのため、ニーズと新技術の交点にあるテーマ、もしくはその交点を模索していくようなテーマを設定しているのですが、一方で、基礎研究である以上は、他の研究者たちとの競争的な要素も意識しなければなりません。そのためには、1人でできることは限られているので、全部を1人でやろうとはせず、自分たちでなければならないことにフォーカスするということが重要だと思っています。そのうえで、チームとして有機的な役割分担で取り組む、あるいは外部の人とコミュニケーションをとりながら共同で取り組むといったことが大切になります。

参加者の合意で成立する国際標準化の世界はこれとは趣を異にするものなのでしょうか。

国際標準化の世界では、世の中の役に立つということが最重要です。そうでなければ国際標準を制定する意味がありません。したがって、参加者はこれをめざして提案し、議論を進める中で合意形成します。議論の中では、場合によっては意見の対立も起こりますが、それをまとめて賛同者を増やして合意形成していくのが、基本的に選挙で決まる議長やラポータ等の役割で、合意を得

るための要件として参加者からの信頼とその分野に対する広く、深い技術力が求められます。

さて、私がテーマとしている技術は国際標準化の場の中ではほんの一部の領域でしかないのですが、それでも自分の技術が採用されないというのは悔しいところもあるので、採用されるための努力はします。しかし、必ずしも自分の技術が採用されなくても、対抗馬を示すことで、より使いやすく価値のある技術が採用されるように誘導できれば、結果として人類に貢献できると考えています。

研究においては、研究者それぞれが得意な分野、テーマ設定から実用化に向けた各フェーズの中で得意とするフェーズがあり、仮にテーマやゴールが近いところであってもそれぞれ異なったアプローチにより、時間軸に差が出てきます。これが研究における競争の1つではないかと思えます。

私には研究者としての立場と、国際標準化への貢献者としての立場の両方があるのですが、実用化して世の中に役立てることをめざして研究に取り組んでいるので、立場、役割は異なりますが、それぞれが別なものという意識はありません。先述のとおり、ニーズと新技術の交点を常に意識しているので、国際標準化における課題やリクワイヤメントがニーズになり、それを実現するための世の中にない新技術・基礎技術という点でそれが研究テーマになります。逆に研究の結果としての新技術を応用して国際標準化の場へ提案し、そこにおける議論をとおして新技術の応用手法を形成していくこともあります。

最近の活動を例に挙げると、例えばDCASEにおいて、私が主催している異常音検知チャレンジではコミュニティを醸成しており、これをとおして異常音検知に関する研究者の数を増やし、そこから生まれてきた新たな技術を参考にして、私たちも最終的にはビジネスに使えるようなヒントを得るという仕組みもつくりました。もちろんコミュニティへの参加者もこの仕組みを活用して研究を進めており、国際標準化とは異なりますが、世の中の役に立つという流れを大きく加速するものだと思っています。

研究の価値を理解・共感してもらう

後進の研究者へのメッセージをお願いします。

一般的に研究者は、自分の興味をテーマとして研究することが多いと思います。また、興味のあるテーマの研究は楽しいことであると思います。こうした研究は、スポンサー、お客さま、パートナー、事業会社、上司、同僚、ライバルの研究者等、周囲の方々にその価値を理解してもらうことで取り掛かることができます。価値を理解してもらうためには、エビデンスを示していくこととなります。「for What, What, Why, How, Evidence」という言

葉がありますが、どのような目標を達成するために、何をやるのか、なぜそれが重要なのか、どのように解決するのかを説明し、そしてそれらをサポートする証拠・エビデンスを示すことが重要です。ただし、このエビデンスを示すべき相手であるスポンサー、お客さま、パートナー、事業会社、上司、同僚、ライバルの研究者等といった方々は、専門家ばかりではなく、また研究者や技術者でもない人もいます。さまざまな立場の人たちに研究の価値をご理解いただき、共感を得られるような説明をしていただきたいと思います。そのためにも研究を取り巻く周囲の環境に対しても意識を向けていくことが必要ではないでしょうか。

しかし一方で、基礎研究の分野では、先進的過ぎて誰もついてこれないような、破壊的なHowが突然生まれてしまうことがあります。このような場合には、ほかの人の言うことなどに紛らわされず、信じた道を突き進んでください。

そして研究を進めた結果、誰も文句を言えない強烈なEvidenceを示すことができれば、まわりもいつかは理解せざるを得なくなります。もちろん、どんなすごい技術であっても、共感を得る努力もするほうが良いとは思いますが……