

特別連載

ムーンショット・エフェクト ——NTT研究所の技術レガシー——

第10回 音の技術

ノンフィクション作家の野地秩嘉（のじつねよし）氏より「ムーンショット・エフェクト——NTT研究所の技術レガシー」と題するNTT研究所の技術をテーマとした原稿をいただきました。連載第10回目は「音の技術」です。本連載に掲載された記事は、中学生向けに新書として出版予定です（NTT技術ジャーナル事務局）。

■ガラス越しの声

NTTメディアインテリジェンス研究所は三鷹駅から車で10分の場所、武蔵野研究開発センタにある。

NTTはいくつもの研究所を持っているが、なかでも同研究所の役割は「音、画像、言語を対象としたメディア処理技術をテーマの主要分野として、思考や心理、能力、知的活動に至るまで人を深く理解する技術の研究開発」である。

そのうち、電話と音に関する研究開発は前身の電電公社時代、いやそれより前の戦前、逓信省時代から続けられてきた。長い知的な蓄積があったからこそその研究開発が続き、その結果が現れている。

さて、私が同研究所に足を踏み入れたら、ふたりの研究員が待っていた。鎌土記良主任研究員と小林和則主幹研究員である。ふたりとも心理情報処理プロジェクトの意図理解技術グループに属している。セクション名を聞いただけでは一体、何をやっているかはまったく想像ができない。要は音声に関係するさまざまな技術開発をやっているとのこと。

案内されたところはごくふつうの会議室だった。大きな机があり、椅子は10席ほどである。そこに座って、説明を聞くのかなと思ったら、小林は「部屋の隅に来てください。そうしたら、すりガラスのそばに立ってください。私が外から話しかけます」とのこと。そう言い残して、小林はさっさと部屋から出ていってしまった。

廊下に立った小林がガラスを叩いた。ガラスを叩く音は聞こえた。

実はそれと同時に彼は「僕の声が聞こえますか」と大

きな声で話していたらしい。しかし、分厚いガラスの向こうの声は部屋のなかにいる私にはむろん聞こえなかった。

ところが、一瞬の後、小林が「聞こえますか？」とガラス越しに話しかけてきたのである。

耳を疑った。

どうした？ 何かマジックでも使ったのか、彼は？

■コロナの時代に

部屋から出て、小林のもとへ行った。すると、手に古いアナログ電話機の受話器のような物体を持っていた。

「商品名は『ウインドウスルーTM会話装置』です。名前の通り、ガラスやアクリル板の窓越しであっても、窓がないかのように会話ができるもので、すでに市販しています」。

鎌土もそばに寄ってきて、説明を始めた。

「開発を始めたのはコロナ禍になってからで、半年で完成しました。ガラス越し、車の窓を閉めたまま会話ができれば飛沫感染を防ぐことができるからです。医療従事者がガラス越しに話ができたらずいぶん安全になります。原理は簡単。この技術は「ウインドウトーク[®]」と呼ばれるもので、人が話した声はガラスをわずかに揺らします。わずかな揺れをウインドウトーク[®]のセンサで拾う。ガラスの外にいる人にはガラスから音がするという形で聞こえます（図1）。車の窓を閉めたままでも使えます。ですからドライブスルー型のPCR検査、クリニック等での発熱外来では試験導入されています」。

もう少し、仕組みを説明する。

ウインドウスルーTM会話装置を窓に押し当てると、内

蔵されている振動素子（エキサイター）が窓を振動させ、機器に接続されたマイクで集音した話者の声を窓越しの相手に伝える。

一方、窓越しに聞く相手の声はウインドウスルー™会話装置内蔵の振動ピックアップ（マイク）が集音するから、話者は機器につながったイヤホンで聞くことができる。ガラスを隔てたふたりのうち、ひとりがウインドウスルー™会話装置を持っていれば会話が成り立つのである。

医療現場では医師がウインドウスルー™会話装置を持ち、患者が何も持たないことを想定している。

開発の当初、問題だったのはガラスを振動させるとエコー振動が発生し、集音したい相手の音声以外のノイズが混入することだった。それを彼らは長年、電話で培ったエコーキャンセラ技術（反射音等を打ち消す技術）を応用してノイズのないガラス越しの会話を実現したのである。

だが、なんとといってもウインドウトーク®の成果は半年という短い期間で実用化したことにある。コロナ禍の時代に開発に3年も4年もかかっていたら何の意味もな

い。研究開発におけるスピードを実現させた原動力はこれもまた電話で培った音声研究の賜物だろう。

■在宅勤務に役立つ研究

鎌土と小林は声を合わせて言った。

「次に紹介する技術もまたコロナ禍の時代に役に立つものです」。

私が連れていかれた場所は天井の高い体育館のような広い室内で、そこには飛行機のファーストクラス（乗ったことはない）のような個室が設けられていた。正確に言えば、箱型の個室にソファと足を乗せるオットマンがしつらえてあった。

私たちは個室席の横に立って話をしていた。

鎌土は言った。

「野地さん、音楽は聞こえますか？」。

いやいや、広い室内はシーンとしているじゃありませんか。何も聞こえませんよ。

「じゃあ、そこにある席に座ってみてください」。

はい、わかりました。と席に座ったとたん、両耳に女性歌手の歌声が飛び込んできた。

「あ!」。

- 受話器型の機器（ウインドウトーク®デバイス）を窓に押当て、振動素子（エキサイター）が窓を振動させ、窓越しに音声を届け、振動ピックアップ（マイク）により、相手音声を集音することで、窓越しの会話を実現
- 窓越しの会話なので、濃厚接触とはならず、感染症リスクを低減することができる

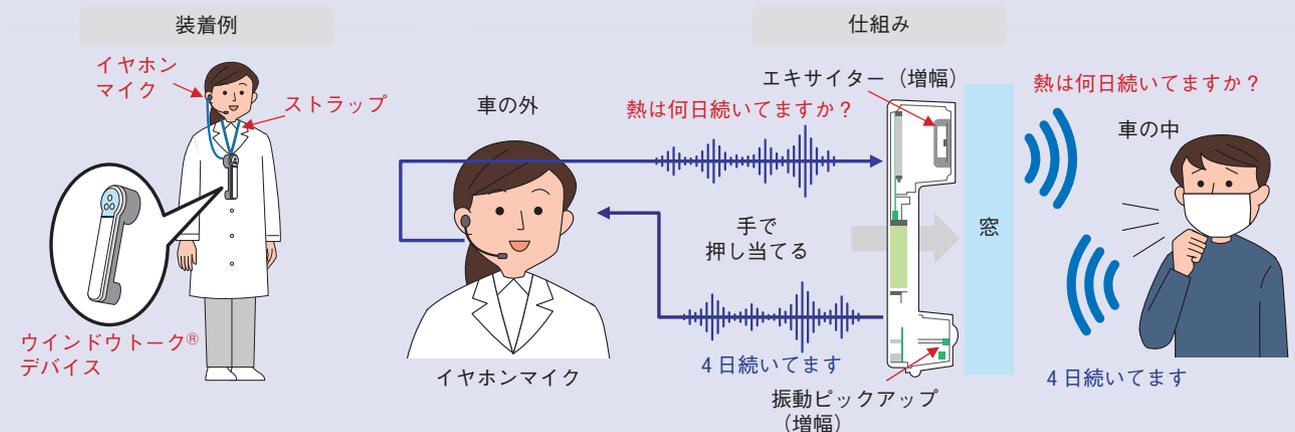


図1 ウインドウトーク®とその仕組み

びっくりして体を起こし、席から体を離れたら、もう音は聞こえない。席に体を押し付けて座っている時にしか聞こえないようになっていた。傍らにいても音は漏れないのである。

鎌土は私がびっくりした様子でいることに満足したようだった。

「これがスポット再生という技術です。顔の周りにだけ音を封じ込めて、外に漏れないようにしてあるのです。『パーソナライズドサウンドゾーン (PSZ: Personalized Sound Zone)』と名付けた究極のプライベート音空間のプロトタイプです。スポット再生は聴きたい人にだけ音を聴かせる技術で、能動サウンド制御技術と我々は呼んでいます」。

これまで周囲に漏れないように音を聞くとすれば、イヤホンやヘッドホンを着けるのが一般的だった。ただし、それであっても少しは外に漏れる。また、長時間にわたってイヤホン、ヘッドホンを使用していると、疲れるし、難聴になるおそれもあった。しかし、スポット再生ではそんなことはない。

■パーソナライズドサウンドゾーン

鎌土の話は続く。

「PSZは周囲の音情報を正確に集音し、周囲の状況を理解したうえで、適切に音を制御する複数の技術を組み合わせます。

主には次の3つです。能動サウンド制御技術、能動騒音制御技術、イベント検知・シーン識別技術です」。

1番目の能動サウンド制御技術はソフトウェアとハードウェアの両方の技術が関わる (図2)。

まず、スポット再生するためには、複数のスピーカを必要とする。再生領域の制御はソフトウェアで実現し、新しく開発した信号処理技術を使う。通常、再生可能な上限の周波数が高く設定されている(再生可能な音の高さが高く設定されている)場合は多くのスピーカが必要となるが、その数を少なくしたのが特徴だ。

また、スピーカの数減らしただけでなく配置も重要だ。個人の自宅では、スピーカを設置できる場所はPCが置かれた机の周りなど、わずかしかない。しかし、この技術では、場所が限られていても、スポット再生ができるように設計してある。

・パーソナライズドサウンドゾーンを実現する技術の1つ「能動サウンド制御技術」は、波形を逆転させた2つの音の回り込みを利用して、スピーカの近くだけで音が聞こえるスポット再生を実現します

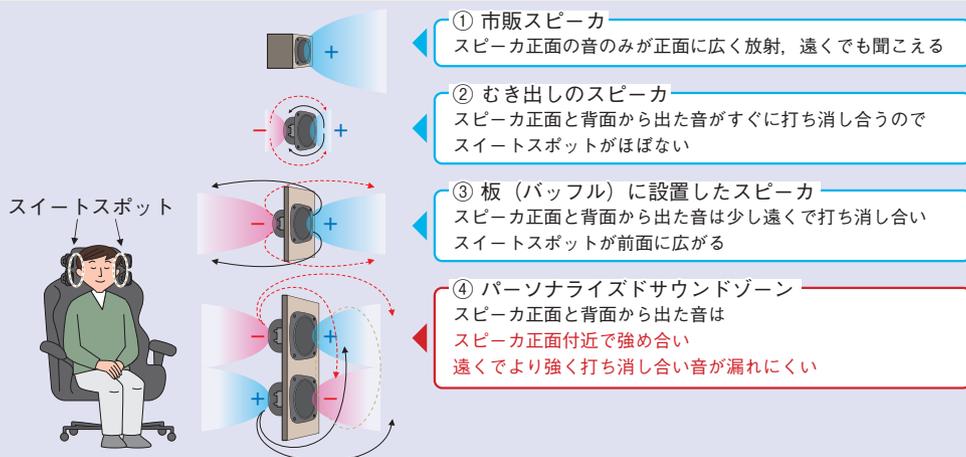


図2 能動サウンド制御技術のポイント

ハードウェア性能向上で特徴的なのは、距離が離れるにつれて音量の減衰が大きいスピーカを開発したこと。スピーカは市販品で、箱に入れずにむき出しのまま利用している。通常、低音域は大型のスピーカが受け持ってきたが、それを低音域まで再生できる小型スピーカにしている。

■能動騒音制御技術

能動騒音制御技術とはつまり、聞きたくない音を消す技術のことで、アクティブノイズキャンセリングとも呼ばれ、航空機の座席にある一部のイヤホンなどで使われている。

それに対して、PSZではスポット再生の場でアクティブノイズキャンセリングを行う。

アクティブノイズキャンセリングとは、ある騒音の波形に、その騒音の波形と逆位相の波形(制御音)を重ねる(音の波形の山・谷が全く逆になった波形を重ねる)ことで、山と谷とを相殺して音を消すことだ。

制御音を出すスピーカの数が多いほど、制御できる点が増え、不要音を消しやすくなる。だが、家庭のなかで利用することを考えると、なるべくスピーカは少ない方がいい。そこで、能動騒音制御技術ではスピーカの数を見直し、騒音を制御できるようにした。

■イベント検知・シーン識別技術

「イベント検知・シーン識別は3つのなかでもっとも難しい技術です」。

そう鎌土は言った。

「イベント検知・シーン識別では何の音かを検知した後、その音が必要なかどうかを状況に応じて判断する技術が必要になります。

例えばですが、在宅勤務をしていると小さなお子さんが部屋のなかに入ってきて、『ママー』と叫んだりすることがあります。また、別の部屋ではお子さんが大きな音でテレビをつけているかもしれません。

リモート会議をしている最中だとしたら、子どもの声が入ってくるのは困る。一方で、状況によっては、子どもの泣き声は聞こえてはいけなくともあります。例えば、家のなかの別の部屋で遊んでいてケガをし

たといったケースだって考えられるからです。

こうした場合、在宅勤務者が聞きたい音だけ聞ける、また通信相手からの音声は在宅勤務者だけに聞こえるといった状況にしなければなりません。つまり、状況に応じてユーザに選択的に音や状況を伝えることが必要なのです。

そのためには、ユーザを取り巻く『環境』を認識する必要があります。『いつ』『何が』『どこで』といった情報を同時推定する『イベント検知技術』、『どのような』『なぜ』といった意味を推定する『シーン識別技術』も必要になります。こうした検知・識別には私たちはAIを利用しました。

つまり、イベント検知・シーン識別には相当に高度な学習能力を持つAIが使われている。

NTTメディアインテリジェンス研究所が進めている技術はもともと快適な音空間を追求したのだろう。しかし、コロナ禍の時代、在宅勤務の状況下ではビジネスパーソンにとって必要不可欠な研究になりつつある。

野地秩嘉 (のじつねよし)

1957年東京都生まれ。早稲田大学商学部卒業後、出版社勤務を経てノンフィクション作家に。日本文藝家協会会員。人物ルポルタージュをはじめ、食や美術、海外文化などの分野で活躍中。著書は



『高倉健インタビューズ』『キャンティ物語』『サービスの達人たち』『ニューヨーク美術案内』など多数。『トヨタ物語』『トヨタに学ぶカイゼンのヒント』がベストセラーに。『TOKYOオリンピック物語』でミスノスポーツライター賞優秀賞受賞。近著は『日本人とインド人』(翻訳 プレジデント社)、『新TOKYOオリンピック・パラリンピック物語』(KADOKAWAから7月14日発売予定)。