

# 究極のプライベート音空間を実現する メディア処理技術

NTTメディアインテリジェンス研究所では、働き方改革等で注目されているテレワークのような、多様な空間におけるデジタルトランスフォーメーションの推進に向け、究極のプライベート空間をつくるメディア処理技術の研究開発を進めています。その実現に向け、もっとも重要な要素の1つである音に着目し、周囲の状況を音から理解する技術（イベント検知・シーン識別技術）、聴きたい人にだけ聴かせる技術（能動サウンド制御技術）、および聞きたくない音を消す技術（能動騒音制御技術）を確立することをめざしています。本稿では、これらの技術への取り組みについて解説します。

ふくい まさひろ さいとう しょういちろう  
**福井 勝宏 齊藤 翔一郎**  
 こばやし かずのり  
**小林 和則**

NTTメディアインテリジェンス研究所

## はじめに

政府が推進する働き方改革および新型コロナウイルスの影響により、従来のようにオフィスに出勤する働き方が見直され、場所や時間にとらわれない柔軟なワークスタイルが注目を浴びています。こうした新しいワークスタイルで重要となるのが、どんな場所でも快適に仕事をするための音環境が整えられることです。ここで、在宅勤務について考えてみましょう。家の中には、エアコンが発するノイズや屋外の自動車走行音、時には宅配便を知らせるチャイムなど、いろいろな音が存在します。家族がいる場合は、その人たちの声やテレビからの音もあるかもしれません。エアコンなどのノイズや家族・テレビが発する音は在宅勤務者にとって「聞きたくない音」です。しかし、状況によっては、チャイムや赤ちゃんの泣き声は「聞きたい音」になる場合があります。在宅で電話会議などを行う場

合、こちら側で発生するノイズは通信相手に届けたくありません。反対に、通信相手からの音声は、他の人に聞かせたくありません。このように、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、といった究極のプライベート音空間をつくり出すことができれば、快適な在宅勤務ができるようになります。

現在、NTTメディアインテリジェンス研究所では、「パーソナライズドサウンドゾーン（PSZ: Personalized Sound Zone）」と名付けたこの究極のプライベート音空間の実現をめざしています（図1）。PSZでは、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、など複数の技術を組み合わせて実現します。NTTメディアインテリジェンス研究所ではこれまで集音技術について多くの知見を蓄積しており、それをさらに発展させた音の「状況理解」や「制御」の技術に現在取り組んで

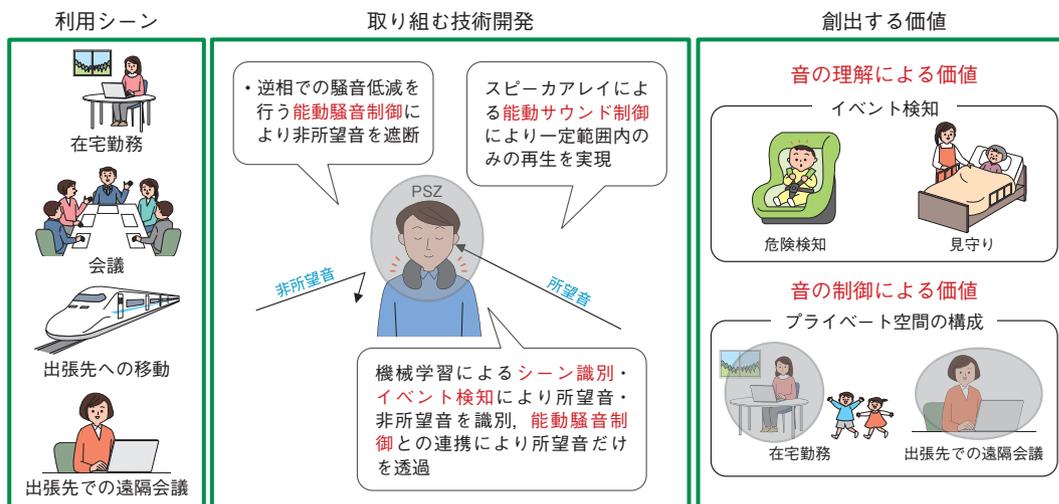


図1 パーソナライズドサウンドゾーンコンセプト

います。以下に、大きく3つの取り組みについて紹介します。

### 周囲の状況を理解する技術 (イベント検知・シーン識別技術)

人は状況によって聞きたい音が異なります。例えば自宅内だと愛犬が鳴く声は聞きたいかもしれませんが、外出先での他の犬の鳴き声は聞きたくないかもしれません。このような場合に、「外出先」において「犬が鳴いている」ということを検知できれば、その音は聞きたくない音である、ということをお判断することができます。

このように、PSZを実現するためには、単純にすべての周囲音を抑制するのではなく、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重要です。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。そのために、「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」や、「どのような」「なぜ」といった情報の意味を推定する「シーン識別技術」に取り組んでいます。

イベント検知技術の難しい点は、同一の場所に到達する音であっても、周囲の多種多様な環境によって音がさまざまに変化する点です。例えば、音が「どこで」発生したかを求める音源定位は、近年、ディープニューラルネットワーク (DNN) を用いた手法が主流ですが、この環境の多様性によりDNNであっても学習データでカバーしきれないことが課題となっています。それに対し、音場の空間対称性を利用したり、物理量推定の手法と組み合わせるなどの工夫で推定精度を向上させる取り組み<sup>(1)~(3)</sup>などに取り組んでいます。一方で、特定のイベントのみを高速・低演算に検知する、というアプリケーションの要請を満たす手法の検討についても進めています<sup>(4)</sup>。

シーン識別技術は、イベントや音源位置より上位の情報として、ユーザの置かれた「状況」の情報を推定することを目標としています。例えば、「車の走行音」というイベントだけでなく、ユーザがどういう状況なのか、また「遠方にある不要な音なので抑圧する」のか「自分に近づいているのでユーザに提示して注意を促す」のか、というところまで判断できるシステムをめざしています。現在その要素技術として、音信号を自然言語で記述する「音説明文生成技術」<sup>(5)</sup>について取り組んでいます。

### 聴きたい人にだけ聴かせる技術 (能動サウンド制御技術)

周囲に影響を与えないように音を聞く場合、これまではイヤホンやヘッドホンを装着する手段が用いられてきました。しかし、着用の煩わしさ、長時間使用による疲れや難聴のおそれ、周囲の状況や危険の察知しづらさ

など、多くの問題がありました。このため、イヤホンやヘッドホンをいわずに対象の受聴者のみが聞こえるようなスポット再生ができれば、これらの問題を解消でき、より便利になります(図2)。NTTメディアインテリジェンス研究所では、このような再生技術の実現をめざし、ソフトウェアとハードウェアの両面で研究開発に取り組んでいます。以降では、それぞれについて課題と取り組みを説明します。

#### ■ソフトウェア性能向上の取り組み

スポット再生するためには、複数のスピーカを必要としますが、再生領域の制御自体はソフトウェアで実現され、能動サウンド制御と呼ぶ信号処理技術を用います。この技術では、通常、再生可能な上限の周波数が高く設定されているほど多くのスピーカを必要とします。また、各スピーカの配置についても制約が発生する場合があります。NTTメディアインテリジェンス研究所がめざすPSZは、

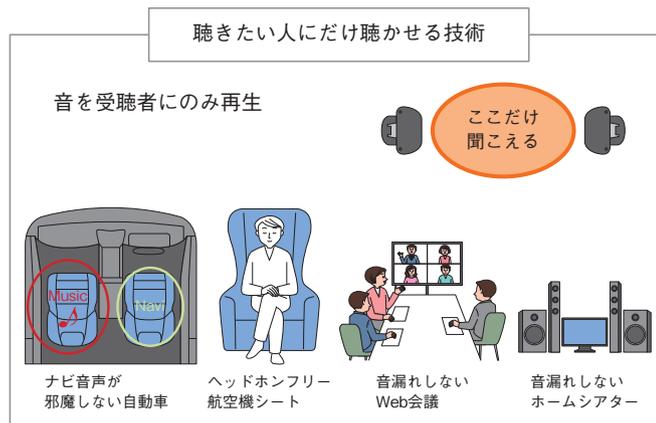


図2 能動サウンド制御技術の適用域

個人向けの空間である性質上、一般的な問題設定と異なり、スピーカの数・配置自由度が著しく制限されます。例えば自宅であれば、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、わずかなスペースに限られます。このような厳しい制限の中、少数のスピーカと限られたスペースでのスポット再生をめざします。能動サウンド制御技術では、フィルタ設計に必要な条件をすべて洗い出して、全条件を同時に満たすよう全体最適化を行っています。

### ■ハードウェア性能向上の取り組み

信号処理技術の検討だけでなく、制約のあるスピーカ数や設置場所において音漏れを最小にできるスピーカ配置を検討するとともに、通常のスピーカより離れるにつれ音量の減衰の大きなハードウェアの検討にも取り組んでいます。ほかにも、上記の取り組みと並行して小型のスピーカで低音を再生する研究開発も行っています。高い音質を保ちたい場合は低音が重要になります。低音を十分な音量で再生するには、スピーカ本体の物理的な大きさを必要とします。しかし、前述のとおり、PSZの実用化にはスペース的な制約がある

ため、サイズの大きいスピーカの利用は現実的ではありません。本研究では、小型スピーカの低音限界がこれまでより低くできるようハードウェアの改良を行っています。

### 聴きたくない音を消す技術 (能動騒音制御技術)

PSZでは、到来する音をイベント検知・シーン識別技術により識別し、不要な音は聞こえない空間の実現をめざします。

現在、広く実用化されているイヤホンなどのノイズキャンセリングは、音を消す空間が狭く、かつ、固定的なため、実現が容易です。しかし、長時間イヤホンを装着するのは、耳が痛くなるなどストレスがたまります。身体に装着しなくてもよい機器で、不要な音を消す技術が実現できれば、より便利になり利用シーンも広がります(図3)。

ある空間で音を消す能動騒音制御技術は、制御音を発生させる制御用のスピーカ、制御点の誤差信号を観測するエラーマイクロホン、騒音信号を参照するリファレンスマイクロホン、そして制御音を生成するための適応アルゴリズムを計算させる制御器で構成されま

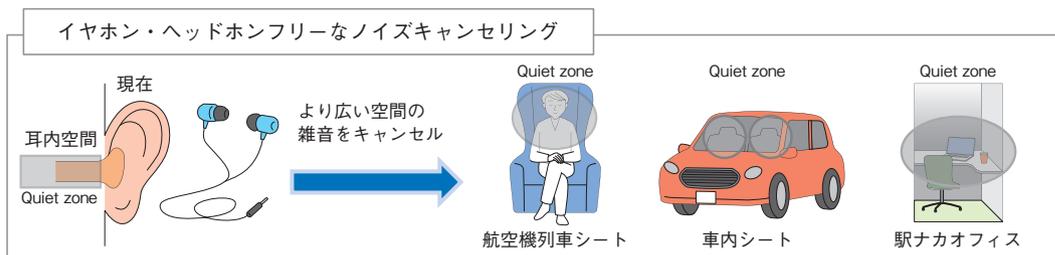


図3 能動騒音制御技術の適用域

す。エラーマイクロホンで観測される誤差信号が小さくなれば、不要音を低減できています。

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ります。しかし家庭内での利用を考えた場合、少数スピーカでの実現が望まれます。また、日本の住宅事情を考慮すると、これらのスピーカ・マイクロホンが近くに配置されることとなります。これまでの能動騒音制御技術で想定していなかった、制御音がリファレンスマイクロホンに回り込むことによる性能劣化など、解決しなければならない問題が残っています。

「聴きたくない音」は、個人や環境によって変わってきます。例えば屋内にいるときは、自動車走行音は聴こえないほうが快適ですが、屋外では走行音が聴こえたほうが安全です。イベント検知・シーン識別で何の音かを検知した後に、その音が今必要かなど、状況に応じた要否を判断するための技術も必要になります。「聴きたい音だけ聞ける世界」を実現するためには、複数の技術が必要になり、それらを高い次元で連携させる必要があります。

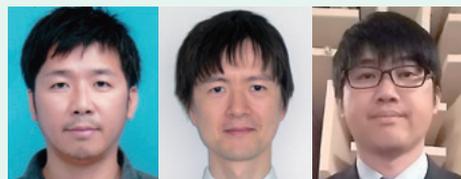
## 今後の展望

本稿では、PSZの概要を述べ、これを実現するための要素技術である「イベント検知・シーン識別技術」「能動サウンド制御技術」および「能動騒音制御技術」における現状の取り組みについて説明しました。技術的な課

題はまだ残されており、NTTメディアインテリジェンス研究所では今後も研究開発を継続的に行っていきます。また、PSZの実現に向け、研究開発だけでなく社内外との連携にも取り組んでいきます。

## 参考文献

- (1) L. Mazzon, Y. Koizumi, M. Yasuda, and N. Harada: "First order ambisonics domain spatial augmentation for DNN-based direction of arrival estimation," Proc. of DCASE 2019 Workshop, New York, U.S.A., Oct. 2019.
- (2) M. Yasuda, Y. Koizumi, S. Saito, H. Uematsu, and K. Imoto: "Sound Event Localization based on Sound Intensity Vector Refined By DNN-based Denoising and Source Separation," Proc. of ICASSP 2020, Barcelona, Spain, May 2020.
- (3) 佐藤・丹羽・小林: "物理的な対称性を保証したアンビニクス領域 DNN による音響イベント検知・方向推定," 日本音響学会秋季研究発表会, 2020.
- (4) 村田・齊藤・小林・中川: "決定木に基づく軽量の音響イベント検知の検討," 日本音響学会秋季研究発表会, 2020.
- (5) Y. Koizumi, R. Masumura, K. Nishida, M. Yasuda, and S. Saito: "A Transformer-based Audio Captioning Model with Keyword Estimation," Proc. of INTERSPEECH 2020, Shanghai, China, Oct. 2020.



(左から) 福井 勝宏 / 齊藤 翔一郎 / 小林 和則

NTTメディアインテリジェンス研究所では、パーソナライズドサウンドゾーンの実現に向けて、企業間連携も推進しながら研究開発を進めていきます。

## ◆問い合わせ先

NTTメディアインテリジェンス研究所  
 心理情報処理プロジェクト  
 意図理解技術グループ  
 TEL 0422-59-4907  
 FAX 0422-60-7811  
 E-mail masahiro.fukui.xe@hco.ntt.co.jp