



車室内の快適で安全な音環境の実現に資する PSZ 能動騒音抑圧技術と所望音通過技術

本稿では、究極の音空間であるパーソナライズドサウンドゾーン (PSZ) 技術における「不要な音を消す」技術と「必要な音は通す」技術について述べます。この要素技術である「音を閉じ込める技術」と融合した新たな能動騒音抑圧技術、音響イベント定位技術、そしてその応用例を紹介します。これらの技術を車室内音響制御に応用すれば先進安全自動車 (ASV : Advanced Safety Vehicle) の利用者にミラー、レーダやカメラの死角を見ることのできる「優れた耳」を提供することもできます。PSZ 技術は利用者の快適性だけでなく、安全性や信頼性を大いに高める「耳」として利用者と車両・車室外環境・車社会を高度につなぎます。

キーワード：#アクティブノイズコントロール (ANC)、#音響イベント定位 (SELD)、
#先進安全自動車 (ASV)

音空間制御に必要な 2 つの技術

私たちを取り巻く音空間は、技術の進歩により、より便利に、より快適に進化しつつあります。

例えば、「聴きたくない音は遮断でき、聴きたい音だけが聴こえる」、このようなニーズは非常に大きく、市場においてもイヤホンなどのウェアラブル製品においてメーカはさまざまな製品を投入しています。一方で、現在に至っても、これらの製品は 2 つの大きな問題を解決できていません。

1 番目の問題は長時間装着時の耳への大きな負担です。聴きたくない音を遮断するためには技術上は耳を塞ぐことがもっともコストが安く効果的であり、このような機能を持つイヤホンの大半が耳の外耳道にイヤピースを挿入するタイプのインイヤー型のイヤホンです。これらのイヤホンは長期使用において圧迫感によるストレスの増加や外耳道疾患のリスクを高める可能性があります⁽¹⁾、人の社会活動の基本となる耳の健康への懸念があります。したがって、「聴きたくない音は耳を塞がずに遮断できる」ことが音空間を制御する技術に求められているといえます。

そして 2 番目の問題は、「聴きたい音」以外の「聴かなければならない音」を十分に考慮できていないことです。人間は目には見えないところで起こっている物事に対し、耳で聴く音で反応し対応することがで

きます。例えば、後方からくる目に見えない自転車や、ベルの音を聴いて避ける対応ができます。一方で、これらのデバイスではそれらの音が聴けるか否かはユーザに判断を任せており、常時このような危険察知に必要な音を聴くことはできません。本来耳に備わる機能である「聴かなければならない音が聴こえる」という安全上重要な機能が失われているということです。したがって、「聴きたい音が聴こえる」だけでなく、「聴かなければならない音も聴こえる」という音空間の制御もまた社会からの要請として技術的に強く求められているといえます。

車両音空間制御における 2 つの技術「ANC」「SELD」

車両においてはこれらの技術的要請は他のケースと比べより顕著となります。

耳を塞ぐようなウェアラブルデバイスを装着することは道路交通法違反の可能性があるだけでなく、条例によっては違反となります。そのため、車両での移動中にも快適な音空間を実現するためには「聴きたくない音は耳を塞がずに遮断できる」必要があります。また、車体やヘルメットによる外音の遮断、車両移動中のロードノイズ等による外音の増加や徒歩よりも移動速度が速いことにより、危険回避のために必要な音が聴こえにくくなるのが事故の原因と

かまど のりよし かわけ ともこ
鎌土 記良 / 川瀬 智子
やすだ まさひろ さいとう しょういちろう
安田 昌弘 / 齊藤 翔一郎
こづか しほり いたう ひろあき
小塚 詩穂里 / 伊藤 弘章
なかやま あきら
中山 彰

NTT コンピュータ&データサイエンス研究所

もなります。そのため、「聴かなければならない音が聴こえる」必要性はより高い、特に先進安全自動車 (ASV : Advanced Safety Vehicle) においてはその安全性を高めるうえで必須であるといえます。

私たちは、これらの技術的要請にこたえるため、パーソナライズドサウンドゾーン (PSZ) の要素技術としてスポット再生・音響 XR 技術と高度に融合した技術の研究開発を進めています。本稿では、図 1 に示すような耳を塞がずに騒音抑圧効果が得られるようなアクティブノイズコントロール (ANC : Active Noise Control) 技術と、周りの音が聴こえにくい環境においても危険回避のために、必要な音は聴こえやすくする音響イベント定位 (SELD : Sound Event Localization and Detection) 技術について紹介します。

耳を塞がずに騒音を抑圧する ANC 技術

ここでは「聴きたくない音は耳を塞がずに遮断できる」ことを目的とした ANC 技術について説明します。

前述のとおり、現在普及しているイヤホンは耳の外音を遮断するため、耳を塞ぐものが一般的です。図 2 (a) に一般に普及しているインイヤー型のイヤホンがどのように聴きたくない音を遮断するかを示します。インイヤー型のイヤホンは耳の穴を塞ぐか

たちで装着します。イヤホンが耳の穴を塞ぐことで耳栓のような役割を果たすことで、耳の外の音は聴こえにくくなります。

ただし、耳の外の音は完全には消えないため、ANCが用いられます。図2 (b)のようにイヤホン内部に2つのマイク、参照マイクとエラーマイクを装着し、これらのマイクで時々刻々と観測される音を元に、耳元の騒音を消すことのできる打ち消し音をスピーカより再生することで、より高い遮音性能を得ることができるようになります。このとき、参照マイクは遮音対象となる周囲騒音を、エラーマイクは耳内におけるANC処理の消し残りの音をそれぞれ観測するために用いられます。

ここで、耳を塞がずに騒音を消す方法を考えます。耳を塞がないためには、耳から離れた位置にスピーカを配置しなければなりません。したがって、スピーカはイヤホンのような小さなものでは出力が足りないため、より大きな物が必要となります。また、インイヤーイヤホンのように耳元に参照マイクやエラーマイクを置くと耳を塞ぐことになってしまうため、これらのマイクも耳から離して設置する必要があります。このようなシステムを図に表すと図2 (c)のようになります。

このように、耳を塞がずに周囲の騒音を抑制するためには、耳から離れたスピーカ・マイクを用いて耳元の騒音の打ち消しを実現する必要があります。しかし、この方法には3つの大きな問題があります。

- ① 制御の安定性です。耳から離れたスピーカで出る騒音の打ち消し音が周囲騒音のみを収録することを目的とした参照マイクで収録されるため、いわゆるハウリング（フィードバックとも呼ぶ）が発生します。
- ② 図2 (b)のようにイヤホンANCでは耳の入り口近くにマイクがあるため、耳内で聴こえる騒音を正しく観測できます。一方で図2 (c)のようにこれらのマイクが耳から離れると耳内で聴こえる騒音を正

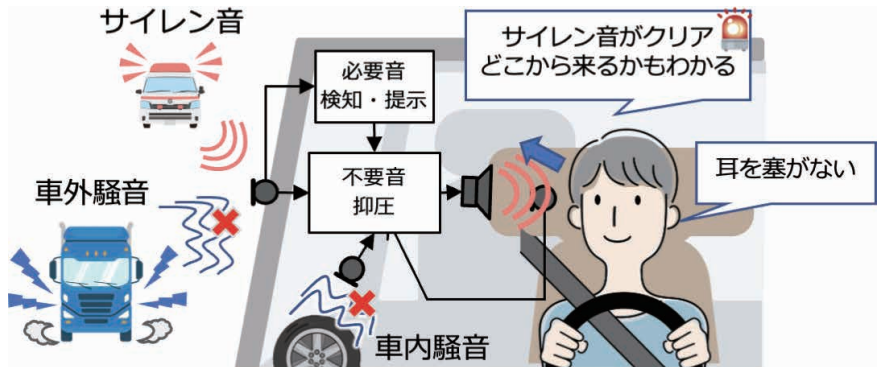


図1 車両におけるPSZ技術の応用例

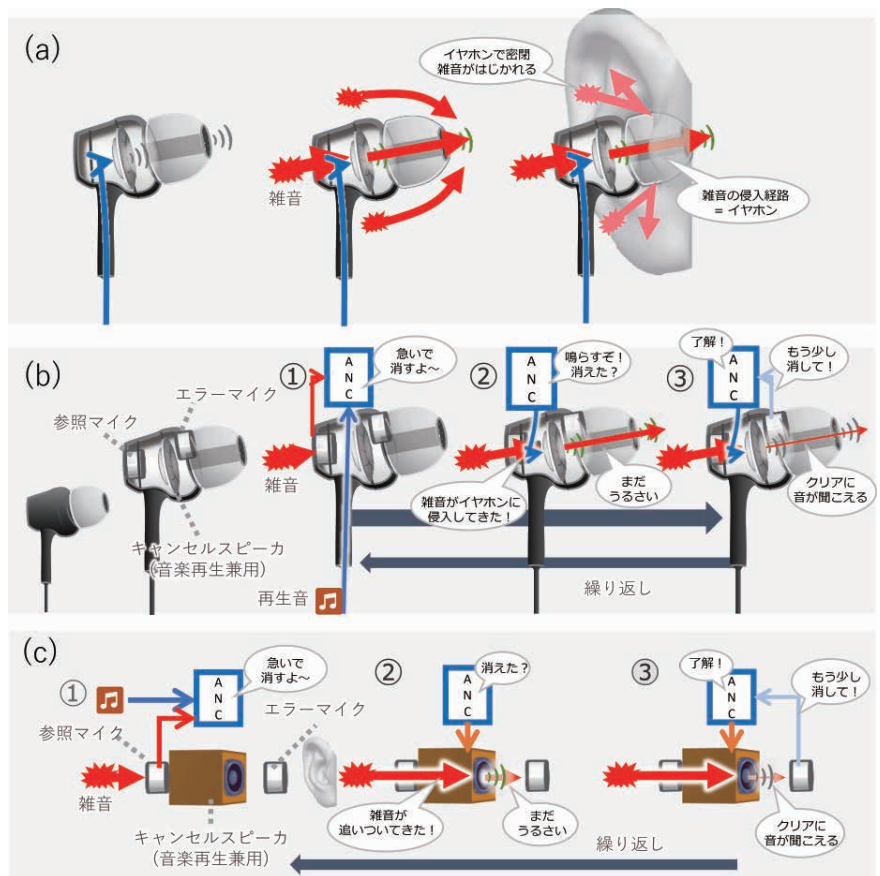


図2 インイヤーイヤホンが必要な音以外の騒音を小さくする仕組み

しく観測できず、騒音以外の音を消そうとしてしまいます。

③ 市販のスピーカやコンピュータでは音の再生と収録にすら時間がかかるため、打ち消し音を生成している間に騒音は耳

元に届いてしまい、打ち消すことができません。また、車室においては大電力の使用は困難であるため、省電力である必要があります。このような用途では従来はデジタルシグナルプロセッサ (DSP)

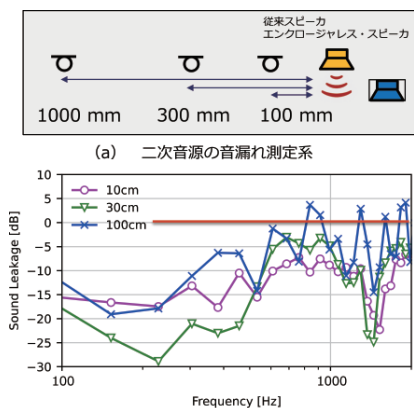


図3 音漏れ抑圧スピーカの効果

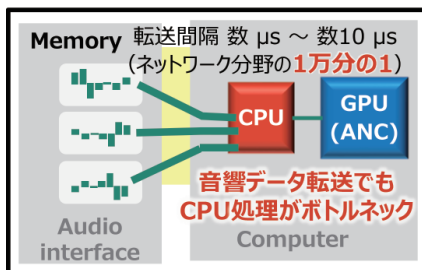
が一般に用いられてきました。一方、耳を塞がないANCにおいては、DSPの演算量が飽和するほどの困難な問題を解く必要があります。さらに、②の問題の対応のためには、一般にスピーカやマイクを多量に配置し、さまざまな補償のための信号処理を実施する必要があります。これには多くの演算量がかかり、打ち消し音の生成が間に合わなくなります。

まず、①の問題の解決のためには、スピーカから参照マイクへの音漏れを低減する必要があります。そこで、前述のスポット再生技術の原理を応用した新たなスピーカを開発しました。図3にその音漏れの低減効果を示します。

このスピーカではスピーカ周囲全体において音漏れが少なくなるだけでなく、特にスピーカの振動板と平行な面では音漏れが非常に小さくなる領域が生まれます。図3において赤線は従来スピーカの音漏れ量を、それ以外の色の線はスポット再生技術を適用したスピーカの音漏れ量を表しています。従来スピーカと比較し、スピーカ近傍100mm~300mmで数dBから数10dBもの音漏れ量の抑圧効果が得られていることが分かります。

②の問題においては、イヤホン同様に参照マイク、エラーマイクを耳元に近づけていくことが必要となります。①のスポット

(a) 従来技術およびその課題



(b) 従来技術の課題を解決する技術ポイント

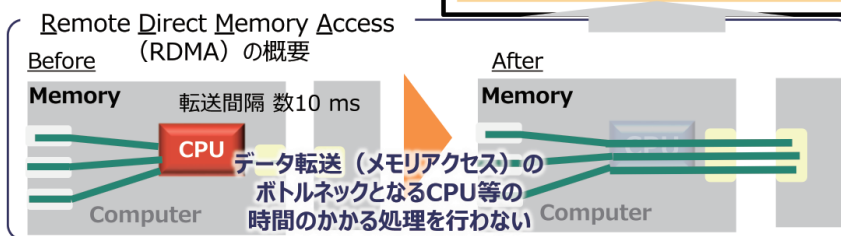
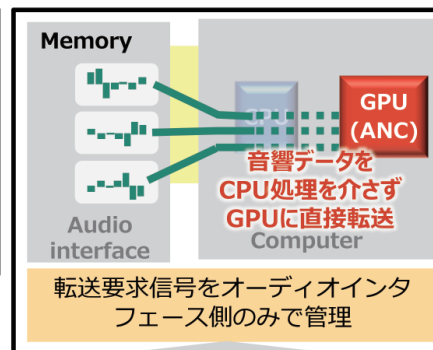


図4 多量のデータを高速低遅延で低消費電力を実現するGPU Direct Audio方式

再生スピーカは音漏れが少ないため、車両ヘッドレストにこのスピーカを埋め込むことで、参照マイクも耳元に近づけることができるようになります。これにより、参照マイクは耳元で聴こえる騒音に近い音を観測できるようになります。

一方で、エラーマイクは参照マイクのようにハウリングの問題は生じないため、参照マイクよりさらに耳元に近づけて置くことができます。しかし、耳元にエラーマイクを置くことは耳を塞がないという前提上不可能であるため、耳からわずかに離しつつ、耳元の騒音を推定する信号処理が必要となります。この処理のためには、一般に複数のエラーマイクを配置しなければならず、この処理を騒音が参照マイクで観測されてから耳元に到達する前の非常に短い時間、数100 μs の間に完了させる必要があります。

これらの理由から、耳を塞がないANCにおいては、信号処理を高速かつ超低遅延に実行する省電力なハードウェアが必須となります。そこで私たちは、ネットワーク・映像処理分野において盛んに活用されてい

るGPGPU (General-Purpose computing on Graphics Processing Units) を音響処理にも応用、最適化することで、この問題を解決しました。

図4にその原理を示します。GPGPUは高速低遅延で単位処理当りの消費電力が少ないことで知られていますが、図4(a)に示すように、データの入力と出力にオーディオ機器を直接接続できず、間にCPUが介入するために処理遅延が増える問題がありました。そこで、図4(b)に示すように、この遅延をRDMA (Remote Direct Memory Access) 技術を利用し、マイク・スピーカの音響信号を、CPUを介さずGPGPUにダイレクトに接続するとともに、1msの遅れも許されない音響処理特有の強いリアルタイム性、他信号処理にない多量の小データ・多量送信パケット(フレームバースト)に対応すべく最適化を実施しました。

この結果、従来の50分の1程度の時間で音響データ転送を実現し、多量の音響信号をリアルタイムかつ低消費電力で処理することが可能なハードウェアを実現しまし

た。このハードウェアにより、DSPでは困難であった耳を塞がないANCを実現することができるだけでなく、ANC分野ではその演算量や処理遅延の問題で導入が難しいといわれている深層学習技術などの導入も実現することができます。

これらの技術を搭載した車両の例を図5に示します。当該車両においては全席のヘッドレスト左右のユーザ視線を遮らない位置にスポット再生機能を持ったスピーカを搭載し、その周囲に参照マイク・エラーマイクを配置しています。シート内部にスポット再生スピーカが搭載されることで、そのスポット再生能力が低下しないよう、ヘッドレストの形状等にも工夫を施し、耳元での騒音抑圧精度を高めています。これにより、車室内空間での快適性を向上させることができます。



図5 耳を塞がないANCシステムの実装例

「聴かなければならない音が聴こえる」SELD技術

前述のとおり、ANC技術では車両走行時の不快な音を抑圧することはできませんが、エラーマイクで観測される音はすべてその抑圧対象となるため、危険回避のために必要な音が聴こえにくくなるのが事故の原因となり得ます。そのため、「聴かなければならない音が聴こえる」ためには新たな必要音通過技術の開発が必要となります。そこで私たちは、SELD技術に着目しました。

SELD技術の概要を図6に示します。SELD技術はマイクで観測される音響信号からいつ、どこで、何があったかを推定する技術です。SELD技術を用いることで、マイクに入力されたさまざまな音から、ユーザに真に必要な音をその到来方位を含め推定することができます。SELD技術は、現在ではエンド・ツー・エンド深層学習技術により構成されるのが一般的で、その内部では「どこで」に対応する音源到来方位(DOA: Direction of Arrival)推定、「何が」に対応する音響イベント検知(SED:

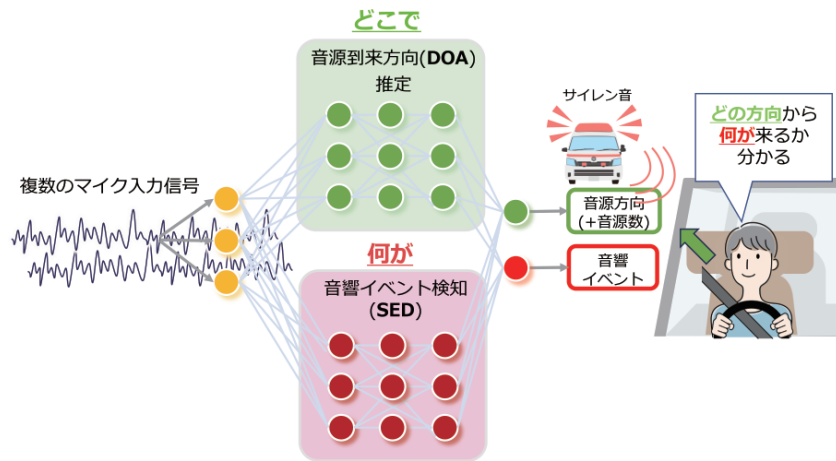


図6 SELD技術の概要

Sound Event Detection) を行います。

SELD技術は深層学習技術を用いるため、その学習には「聴かなければならない必要な音」に関する音響信号が多量に必要となります。例えば、図6のように運転者の死角にいる遠くのサイレン音をいち早く検知し、ユーザにその方位を知らせることで適切な回避を促すような用途では、理想的には死角になるあらゆる方位のサイレン音を、あらゆる場所、あらゆるシチュエーション(例えば周囲の車、建物や天候状況)などを考慮し収録し学習させる必要があります。しかし、現実にはこのような多量の音の網

羅的な収録は困難です。

一方、人間は環境の違いや自己運動に伴う音の変化があったとしても、音の到来方位をある程度推測することができます^{(2)~(4)}。これは、人間が音に含まれる情報から、その到来方位を推定するために必要な情報を取捨選択できるということを意味します。私たちはこの人間の能力^{(5)~(7)}をSELDに模倣させることを考えました。

私たちは、このような人間の能力を模倣する機能をSELD技術に付与するEcho-Aware feature Refinement (EAR)-Comprehensive Anechoic data and

Sparse Multi-environment data (CASM) 技術⁽⁶⁾ および Motion-Aware feature Refinement (MAR) 技術⁽⁹⁾を開発しました。これにより、環境が変わっても、自分が動いても頑健に動作する SELD 技術が実現できます。これらの技術により、さまざまな環境の音の網羅的な収録は不要となり、従来は困難であった車のような移動環境における SELD 技術の適用が現実的なコストで可能となりました。

EAR-CASM 技術と MAR 技術を適用した車載向け SELD 技術の概要を図 7 に示します。EAR-CASM 技術では、SELD の学習データに含まれない音の響きを知るための手掛かりとして、自車から発した音（例えば自車走行音やソナー音など）が周囲に反射して戻るエコーを SELD に与えます。改良された SELD は、このエコーから未知の環境の音の響きの影響をこれまで学習してきた音の響きを活用し抑圧する、人の耳のような動きをします。また、MAR 法では、自己運動の情報として、自車に取り付けられた加速度センサやカメラ情報等さまざまなセンサ入力を SELD に与えることで、自車運動に対する音の変化の影響を低減します。

本技術を ANC 技術とともに ASV に適用することで、「聴きたくない音は耳を塞がずに遮断でき」、「聴きたい音が聴こえる」だけでなく、「聴かなければならない音も聴こえる」、安全で快適な移動のための音空間の創出に貢献しています。

■参考文献

(1) C. Mukhopadhyay, S. Basak, S. Gupta, K. Chawla, and I. Bairy : "A comparative analysis of bacterial growth with earphone use," OJHAS, Vol. 7, No. 2, April 2008.
 (2) D. R. Begault, E. M. Wenzel, and M. R. Anderson : "Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source," J. Audio Eng. Soc., Vol. 49, No. 10, pp. 904-916, Oct. 2001.

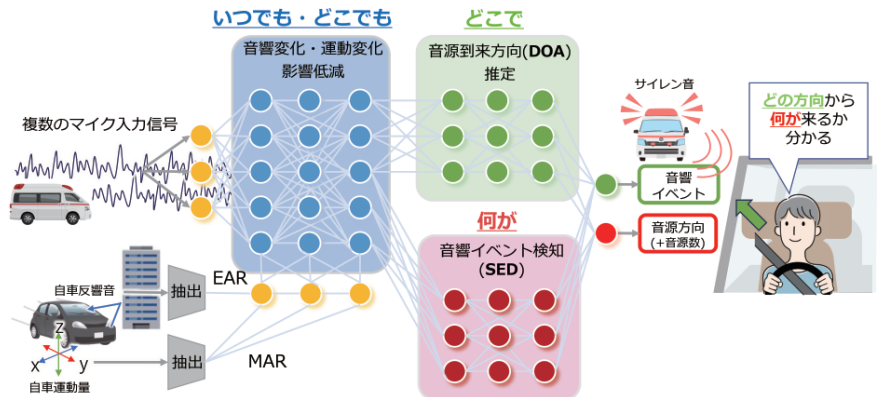


図 7 EAR-CASM と MAR を適用した車載向け SELD 技術の概要

(3) Y. Iwaya, Y. Suzuki, and D. Kimura : "Effects of head movement on front-back error in sound localization," Acoust. Sci. & Tech., Vol. 24, No. 5, pp. 322-324, 2003.
 (4) B. C. J. Moore : "An introduction to the psychology of hearing (3rd ed.)," Academic Press., 1989.
 (5) R. Gao, C. Chen, Z. Al-Halah, C. Schissler, and K. Grauman : "Visual echoes: Spatial image representation learning through echolocation," Proc. of ECCV 2020, August 2020.
 (6) F. Antonacci, J. Filos, M. R. P. Thomas, E. A. P. Habets, A. Sarti, P. A. Naylor, and S. Tubaro : "Inference of room geometry from acoustic impulse responses," IEEE Trans. on Audio, Speech, and Lang. Process., Vol. 20, No. 10, pp. 2683-2695, 2012.
 (7) L. Rosenblum, M. Gordon, and L. Jarquin : "Echo locating distance by moving and stationary listeners," Ecological Psychology-ECOLPSYCHOL, Vol. 12, pp. 181-206, 2000.
 (8) M. Yasuda, Y. Ohishi, and S. Saito : "Echo-aware Adaptation of Sound Event Localization and Detection in Unknown Environments," Proc. of IEEE ICASSP 2022, May 2022.
 (9) M. Yasuda, S. Saito, A. Nakayama, and N. Harada : "6DoF SELD: Sound event localization and detection using microphones and motion tracking sensors on self-motioning human," Proc. of IEEE ICASSP 2024, April 2024.



(後列左から) 中山 彰 / 齊藤 翔一郎 / 小塚 詩穂里 / 伊藤 弘章 (右上)
 (前列左から) 川瀬 智子 / 安田 昌弘 / 鎌土 記良

PSZ は利用者の音環境を周囲の音環境と安全性・信頼性を保ちながら高度に調和・融合し、利用者とその周囲だけでなく、利用者を取り巻く社会環境をもつなぐ、なくてはならないインタフェースとなる技術です。PSZ の利用シーンは広がり続けます。

◆問い合わせ先

NTT コンピュータ&データサイエンス研究所
 企画担当
 TEL 046-859-4003
 FAX 046-855-1149
 E-mail cd-koho-ml@ntt.com