

<p>どんな研究</p>	<p>人間は、様々な音がする中でも、聞きたい人の声や聞きたい種類の音に注目して聞くことができる、選択的聴取と呼ばれる能力を持っています。本研究は、このような人間が持つ選択的聴取の機能をコンピュータ上で実現することをめざしています。</p>
<p>どこが凄い</p>	<p>聞きたい人の声のみを抽出する目的話者抽出技術の抽出対象を任意の種類音に拡張しました。聞きたい音の種類を指定することで、それらの音のみを抽出できます。単一のモデルで日常生活の様々な音を抽出することを可能とし、人間の選択的聴取能力にさらに近づくことに成功しました。</p>
<p>めざす未来</p>	<p>例えばクラクションの音は在宅勤務中には集中を妨げる迷惑な音ですが、道を歩いている際には危険を知らせる重要な情報です。本技術によって、周囲で鳴っている音の中から、状況に応じて聞きたい音・聞きたくない音の選択的聴取を可能とし、快適かつ安全な音環境の創造に貢献します。</p>

人間の選択的聴取

- 日常生活の中では、様々な音が鳴っています。
 - 同じ音でも、状況により、迷惑な雑音の場合もあれば、情報や危険などを知らせる重要な音である場合もあります。



運転中に、消防車のサイレンは重要な音
 仕事中に、消防車のサイレンは迷惑な音

- 人間は状況によって、聞きたい音に集中して聞くことができます。(=選択的聴取)

本研究の目標

- 様々な音が鳴っている混合音から、聞きたい音の種類を指定することで、それらの音のみを抽出できる技術を目指します。

→ コンピュータによる任意の音の選択的聴取を可能に



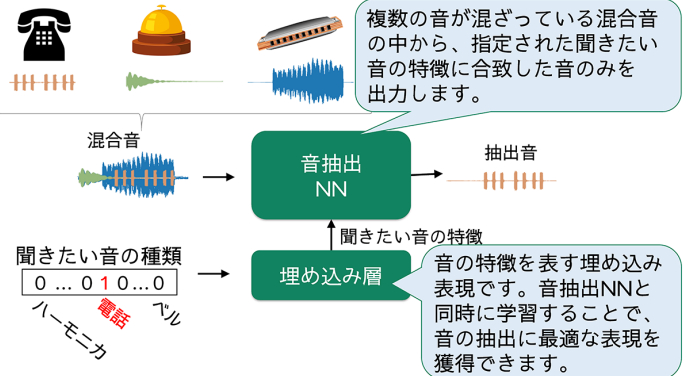
→ 状況によって聞きたい音が変わっても、自由に選べます。

- 応用先
 動画のオーディオの編集や、環境や人の好みに合わせて周囲の音の中から所望の聴取音を制御できるヒアリングデバイスなど。

任意の音の選択的聴取SoundBeam

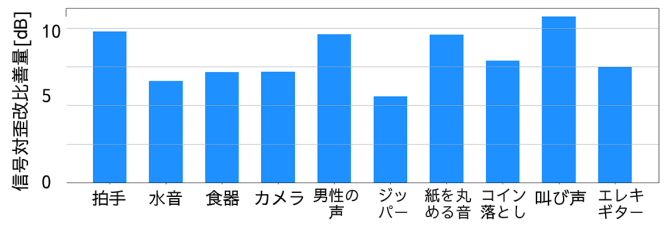
オープンハウス2018年で紹介した深層学習による声の特徴に基づく音声の選択的技術SpeakerBeam [1] を任意の音に拡張

- 深層学習・ニューラルネットワーク(NN)に基づいて、任意の音の選択的聴取を実現します [2,3]。
 - 様々な音の種類を含む人工的に作った混合音と、聞きたい音の信号を用いて、NNを学習します。



指定した聞きたい音の種類を変えることで、同じモデルで、様々な音の種類を抽出可能!

- 10個の音の種類に対する抽出実験
 - 10種からランダムに3つの音を選んで作成した混合音で評価。
 - 人の声だけではなく、日常の様々な音の抽出ができることを確認できました。



信号対雑音比改善量は、混合音に対して聞きたい音がどの程度大きく聞こえるようになったかを測る尺度です。この値が正の場合、提案法が聞きたい音を抽出できていることを示しています。

関連文献

[1] M. Delcroix, K. Zmolikova, 木下慶介, 荒木章子, 小川厚徳, 中谷智広, "SpeakerBeam: 聞きたい人の声に耳を傾けるコンピュータ——深層学習に基づく音声の選択的聴取," *NTT技術ジャーナル*, Vol. 30, No. 9, pp. 12-15, 2018.

[2] T. Ochiai, M. Delcroix, Y. Koizumi, H. Ito, K. Kinoshita, S. Araki, "Listen to what you want: Neural network-based universal sound selector," in *Proc. Interspeech*, pp. 2718 - 2722, 2020.

[3] M. Delcroix, J. B. Vázquez, T. Ochiai, K. Kinoshita, Y. Ohishi, S. Araki, "SoundBeam: Target Sound Extraction Conditioned on Sound-Class Labels and Enrollment Clues for Increased Performance and Continuous Learning," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, Vol. 31, pp. 121-136, 2023.

連絡先

デルクロア マーク (Marc Delcroix)
 メディア情報研究部 信号処理研究グループ