

## 騒がしい生活環境で、音声だけ高品質に取り出す

<b>どんな研究</b>	がやがやとした生活環境において、音声端末が人間の声を高精度に音声認識するには、マイクに混入する雑音や残響を取り除き(雑音抑圧・残響抑圧)、複数の人の声を聞き分ける必要があります(音源分離)。この展示ではこれら <b>3つの問題を統合的に解決</b> する新技術を紹介します。
<b>どこが凄い</b>	現在の最先端技術は、音源分離・雑音抑圧・残響抑圧の3技術を別々に最適化するため、雑音や残響が大きい環境において、音声を高品質に取り出すことができませんでした。これら <b>3技術を統合し全体最適化</b> することで、音声を高品質に取り出し、音声認識性能を大きく改善しました。
<b>めざす未来</b>	マイクで収録した音から各話者の音声を高品質に取り出す技術は、スマートスピーカや対話ロボットといった <b>音声端末(コンピュータ)の基盤をなす技術</b> です。駅・街頭・ショッピングモールなどの生活環境において、人とコンピュータが円滑に対話できる世界の実現に寄与します。

### 目的：マイクで収録した音声をきれいにする

生活環境において遠隔マイクで収録した音は、音声認識が難しい

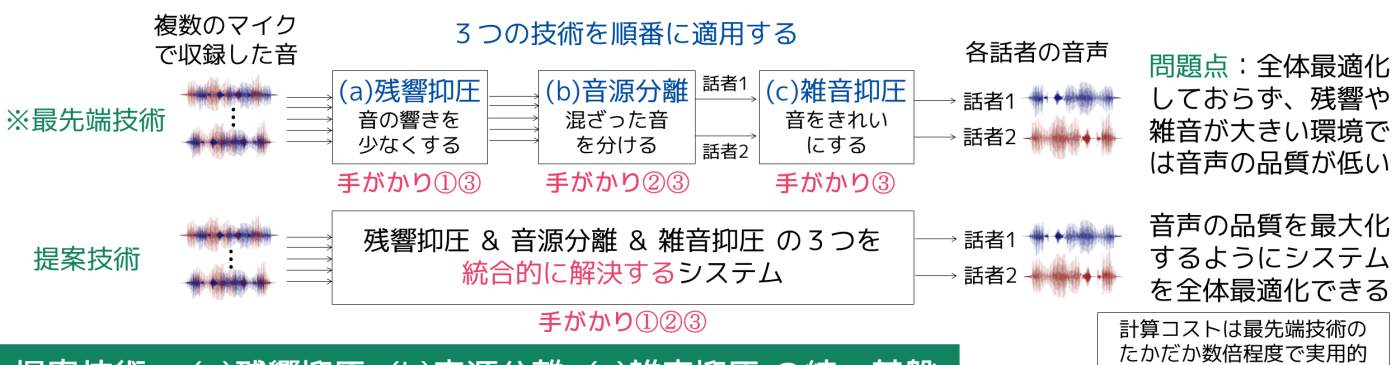
難しさの要因	必要な技術
残響(壁に反射して届く響いた音)が大きい	(a)残響抑圧
複数の人の声が混ざっている	(b)音源分離
背景雑音のレベルが大きい	(c)雑音抑圧



複数マイクを用いた**信号処理**で、それぞれの人の声を高品質に取り出したい  
音声認識には DNN よりはるかに高性能

### 従来技術の問題点と本研究の成果

※ 音声認識の国際コンペ CHiME3 における NTT の技術  
 現在もトップ技術として広く利用されている



### 提案技術：(a)残響抑圧, (b)音源分離, (c)雑音抑圧の統一基盤



### 関連文献

- [1] T. Nakatani, C. Bøddeker, K. Kinoshita, R. Ikeshita, M. Delcroix, R. Haeb-Umbach, "Jointly optimal denoising, dereverberation, and source separation," in *Proc. IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 2267-2282, 2020.
- [2] R. Ikeshita, T. Nakatani, S. Araki, "Block coordinate descent algorithms for auxiliary-function-based independent vector extraction," in *Proc. IEEE Trans. Signal Process.*, 2021, to appear.
- [3] R. Ikeshita, T. Nakatani, "Independent vector extraction for fast joint blind source separation and dereverberation," in *Proc. IEEE Signal Process. Lett.*, 2021, to appear.

### 連絡先

池下 林太郎 (Rintaro Ikeshita) メディア情報研究部 信号処理研究グループ  
 Email: cs-openhouse-ml@hco.ntt.co.jp