

## こんなにガヤガヤした場所でも聞き取ります

～遠隔発話音声認識における雑音除去・深層学習技術の最前線～

## どんな研究

音声認識技術は近年飛躍的に向上していますが、展示会場のような騒がしい場所で口元から離れたマイクを使う場合、雑音や残響の影響を大きく受け、未だに音声認識は難しいという課題があります。この展示では、騒がしい環境でも精度のよい音声認識を実現するための基盤技術を紹介します。

## どこが凄い

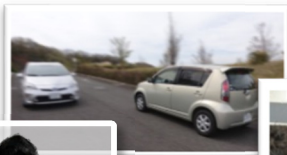
音声認識性能を劣化させる雑音や残響を、音声を歪ませることなく低減する技術や、雑音の影響を受けても精緻に音声をモデル化するディープラーニング音声認識技術を開発しました。これらの技術で、公共エリア雑音下でのモバイル音声認識国際技術評価(CHiME-3)で、世界1位の音声認識精度を達成しました。

## 目指す未来

カフェや空港などの公共エリアでの音声インターフェースや、オフィスや家庭のリビングなどでの会話認識など、騒がしい所で多くの人が話す場面でも快適に動作する音声認識が実現できます。これは、スマホの音声エージェントやコミュニケーションロボットの利用シーンの拡大に大きく貢献すると期待されています。

## 音声認識の利用シーンの拡大へ

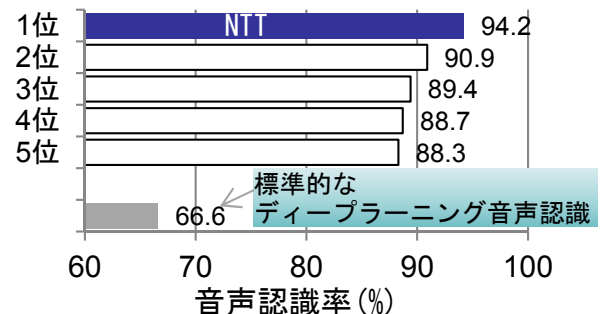
騒がしい公共エリアで  
(道案内、店頭案内など)



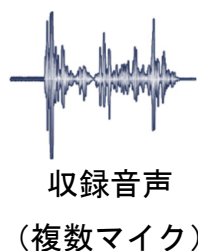
屋内で  
(家電操作、会議録作成など)

## 世界トップ性能を達成

公共エリア雑音下での音声認識(CHiME-3)



## 技術の特徴



ひずみなし  
音声強調

ディープ  
ラーニング  
音声認識

音声認識結果

- 収録音声に含まれる音声をひずませることなく、雑音や残響の音量を低減
- 後段のディープラーニング音声認識に好適な強調音声を生成

- 雑音に影響を受けた音声も精緻にモデル化 (Network-in-Networkに基づく多層畳み込みニューラルネット音響モデル)
- 長い文脈を考慮したより正確な音声認識 (再帰的ニューラルネット言語モデル)

## 【関連文献】

- [1] T. Yoshioka, N. Ito, M. Delcroix, A. Ogawa, K. Kinoshita, M. Fujimoto, C. Yu, W. J. Fabian, M. Espi, T. Higuchi, S. Araki, T. Nakatani, "The NTT CHiME-3 system: Advances in speech enhancement and recognition for mobile multi-microphone devices," in *Proc. ASRU2015*, 2015.
- [2] S. Araki, M. Okada, T. Higuchi, A. Ogawa, T. Nakatani, "Spatial correlation model based observation vector clustering and MVDR beamforming for meeting recognition," in *Proc. ICASSP2016*, 2016.

## 【連絡先】

荒木 章子 (Shoko Araki), 木下 慶介 (Keisuke Kinoshita) メディア情報研究部 信号処理研究グループ E-mail : {araki.shoko, kinoshita.k}@lab.ntt.co.jp