FY2022/2022年度

Sound Scene Understanding Team

Kazuyoshi Yoshii

音響情景理解チーム　吉井 和佳

RIKEN Center for Advanced Intelligence Project

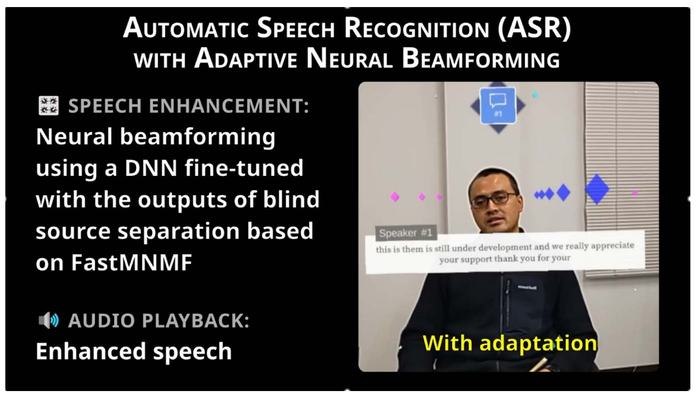
スマートグラスを用いた視聴覚統合環境理解・支援

健常者向け：視聴覚機能の拡張

カクテルパーティにおける会話支援

音声強調・認識・翻訳・AR表示

視線/指差しで話者指定・同時発話認識



開発中のARシステム (HoloLens2)

聴力障害者向け：“知的”健康の回復

屋内外での安心・安全な日常生活の支援

視野内外の音イベント解析・可視化

重複会話の抑制

着目したい話者に対する音声強調・認識

街案内との同時表示

音環境の可視化 (視野内)・注意喚起 (視野外)

実環境を想定した高性能・低遅延システムの開発

高難度リアルデータへの挑戦：EasyComデータセット [Donley+@Meta 2021]

メガネにRGBカメラと4個のマイクを取り付け

装着者がテーブルを囲んで複数人と自然体で会話

頭部動作・音源移動・背景雑音あり

(+ 左右の耳の位置)

保有するブラインド音源分離技術 (FastMNMF) のアプリケーション展開

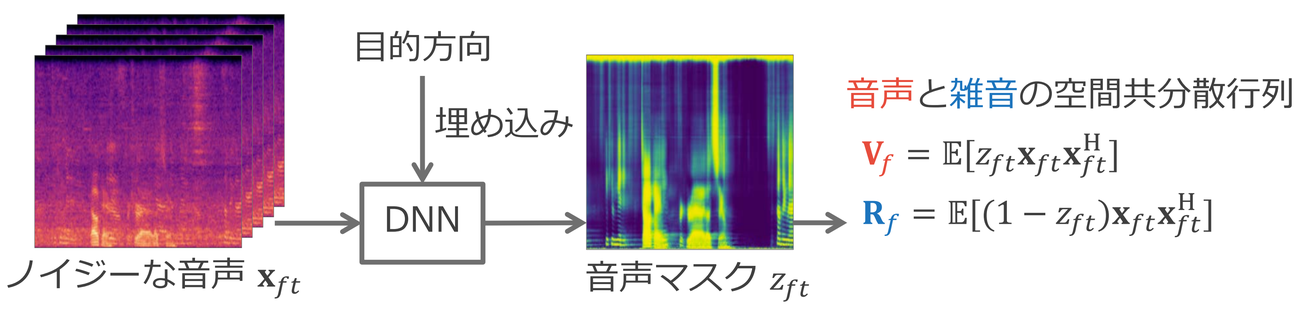
実環境中で頑健に動作する高性能かつ低遅延な音声強調・認識

音声強調：MVDRビームフォーミング

DNNで空間共分散行列を予測

音声認識：RNNトランスデューサー

DNNで音声を文字へストリーミング変換



<u>運用時</u>にオンライン適応することが必要不可欠 → 独自の挑戦的な試み

(オフラインベンチマークからの脱却：正解の音声・認識結果が未知の状況でどうするか？

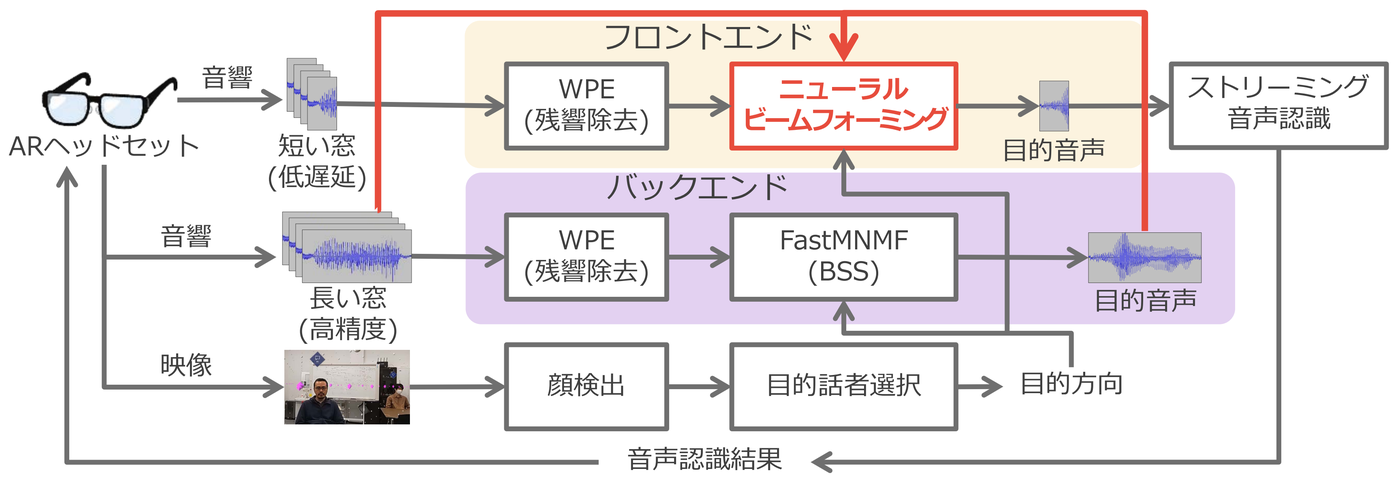
音声強調部のオンライン適応

[IROS 2022/IWAENC 2022]

デュアルプロセスに基づく適応的音声強調

Teacher-Student学習に基づくオンライン適応

単語誤り率が12ポイント以上改善



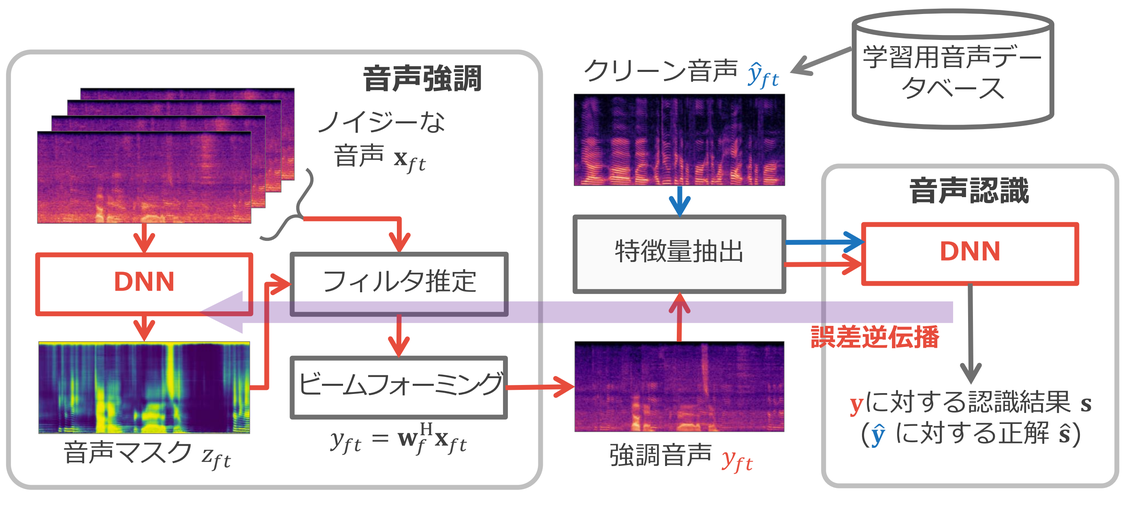
音声強調・認識部の同時オンライン適応

[Interspeech 2022]

音声強調・認識部の同時最適化

信頼できる音声認識結果を学習データに利用

単語誤り率が10ポイント改善



トップカンファレンスで学術発表：同様のシステムを開発中のビッグテックに先駆け