遠方マイクで収録した音声を 近接マイク品質にする音響信号処理技術

中谷 智広^{1,a)}

概要：話者から離れたマイク（遠方マイク）を用いて音声を収録すると、他の話者の音声や背景雑音、残響などの不要音が相対的に高いレベルで収録されるため、音声の了解度は低下し、音声認識などの音声アプリケーションの性能は劣化する。音声強調は、収録音からそのような不要音を取り除き、話者の口元のマイク（近接マイク）で収録した品質の音声に近づけることで、これらの問題を解決する技術である。近年、古典的な音響信号処理の発展技術や深層学習に基づく新技術、およびそれらの統合技術などが活発に研究され、音声強調技術の大幅な進展が得られている。特に、複数マイクを用いた音声強調は最新の深層学習に基づく音声認識の性能を大きく改善できることが、国際チャレンジ企画などを通じて多数報告されている。

本講演では、まず、古典的な音響信号処理の発展技術の中でも代表的な技術として、重み付き予測誤差法 Weighted Prediction Error (WPE) に基づく残響抑圧について詳しく解説する。WPE は、未知の環境で収録され様々な音が混在する収録音から残響のみを効果的に抑圧できるため、幅広い状況に対応できる。また、WPE を他の音声強調手法（独立ベクトル分析に基づく音源抽出など）と統合して用いることで、雑音・残響環境下で収録された混合音声から目的音声のみを高精度に抽出する処理が実現できる。さらに、本講演では、より最近の取り組みの中から、少数のマイク（2～3 個程度）で高精度な音声強調を実現するための新技術として、入力音の状態に応じて複数の音声強調処理を動的に切り替えて適用するスイッチ機構や深層学習に基づく仮想マイク推定などについても紹介する。

¹ 日本電信電話株式会社
NTT Corporation, Seikacho, Sorakugun, Kyoto 619-0237,
Japan

a) tnak@ieee.org