

# 多様な残響環境下における深層学習による ブラインド残響ノイズ除去手法に関する研究

片倉 多智 †

† 東京都市大学大学院

荒井 秀一††

†† 東京都市大学大学院

## 1. はじめに

現代社会において、機械を介した音声通信はコミュニケーションの主要な方法となった。そのため、音声信号に重畳された背景ノイズや残響ノイズの除去は、音声通信の円滑化や精度向上を図るために必要なタスクの一つとなっている。本研究の目的はノイズ除去タスクに残されている課題を調査し原因究明と改善である。

## 2. 残響ノイズ除去タスク

残響ノイズ除去タスクとは、屋外からの騒音のような加法性のノイズではなく複雑に重畳された残響ノイズの除去を目的とするタスクである。先行研究では、機械学習をもとに、ブラインド残響除去と呼ばれる残響環境が不明な音声の残響除去や、様々な環境の残響ノイズを一つのアーキテクチャで汎用的に除去する課題も近年研究が行われている。

## 3. 従来研究

本稿では、従来研究にて提案されている具体的な残響ノイズ除去手法について解説を行う。

### 3.1 2段階の状況考慮型 DNN による背景ノイズ除去・残響処理フレームワーク [1]

Mao ら [1] によって提案された背景ノイズと残響を同時に行うことを目的としたフレームワークで、ノイズ除去段階と残響除去段階で構成される。図1はアーキテクチャの全体図である。ノイズ除去段階では、残響を変えずに、背景ノイズのみを除去することが目的である。残響除去段階では、ノイズのない残響音声からクリーンな音声を推定することが目的である。二段階をフロントエンドで学習する。

### 3.2 残響時間認識 DNN (RTA-DNN)[2][3]

3.1 で説明したアーキテクチャの残響除去段階に利用されているアプローチに関して説明する。Wu ら[2]によって提案された RTA-DNN は、RT60 に依存する2つのパラメータを特徴抽出のための DNN に供給する手法である。RT60 とは残響環境を示す指数の一つで、室内の音源が停止した後、音圧レベルが 60dB 減衰するまでの時間である。残響は環境によって異なる重ね合わせとフレーム単位での時間相関が存在するため、そこに注目することで性能と頑健性を高める効果がある。1つめのパラメータのフレームシフトサイズ(R)である。これが小さいほど波形を細かく見ることができる。

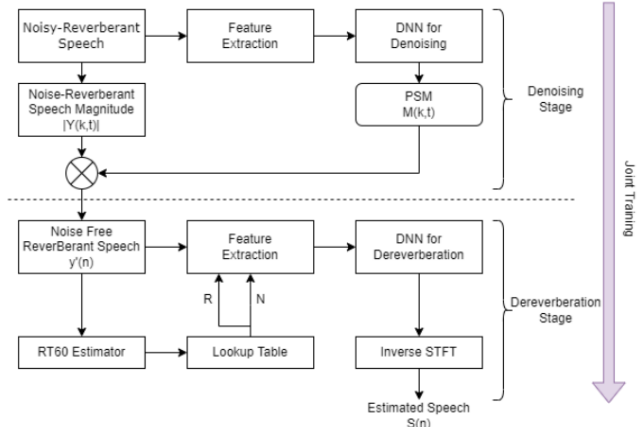


図1. 状況考慮型 DNN アーキテクチャ概要図[1]

2つめのパラメータは音響コンテキストサイズ(N)である。これが大きいほど多くの前後のフレームを見ることができ、弱い残響の場合、高密度なフレームレートによる高い時間分解能が必要になる。強い残響の場合は高い分解能は不要で前後のフレームとの相関を見る必要がある。RT60 に応じて最適なパラメータを選択して残響除去を行う。

## 4. 提案と実験

従来手法の問題点として RT60 が大きい場合に精度が落ちている。可変のパラメータが 0.4 以上の RT60 環境において不変であること、R が増加しているにも関わらず、N も増加しているため、周囲との相関が失われていることが原因と考えた。RとNだけでなくフレームサイズも RT60 に応じて可変なパラメータにし、学習を行うことで上記の問題を解決する。

## 5. まとめ

0.1 から 1.0 まで 0.1 刻みでの残響環境を再現し N を固定値にした上で R とフレームサイズのみを可変にする実験を行い、処理後の音声の明瞭度が向上した。

## 参考文献

- [1] Sidheswar Routray, Qirong Mao "A context aware based deep neural network approach for simultaneous speech denoising and dereverberation"
- [2] Wu, K. Li, M. Yang and C. Lee, "A Reverberation-Time-Aware Approach to Speech Dereverberation Based on Deep Neural Networks," in IEEE/ACM Transactions on Audio, Speech and Language Processing, 2019.
- [3] Y. Xu, J. Du, L. -R. Dai and C. -H. Lee, "A Regression Approach to Speech Enhancement Based on Deep Neural Networks,"