

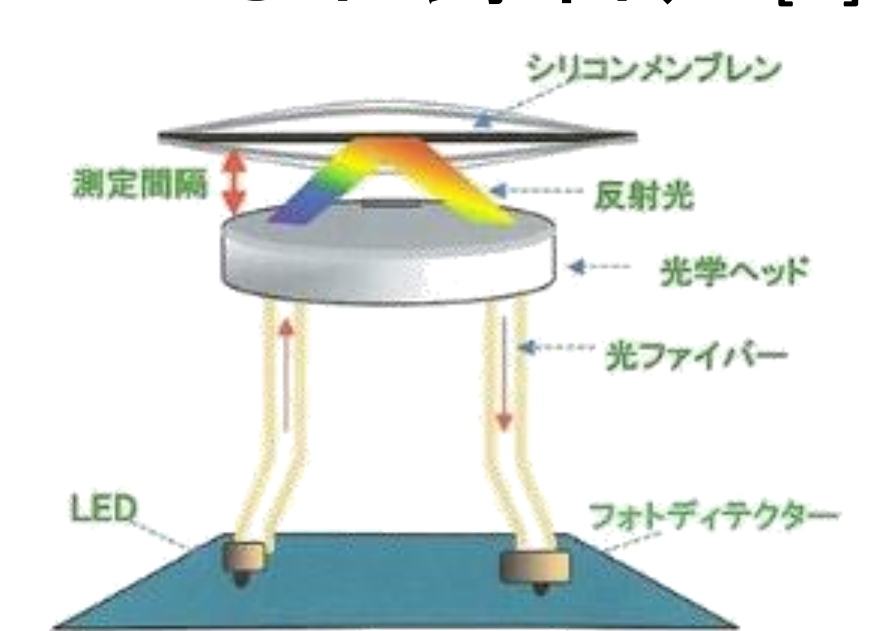
# U-Netを用いた光骨伝導音の音質改善

川本 大貴, 中山 仁史  
 広島市立大学 大学院情報科学研究科

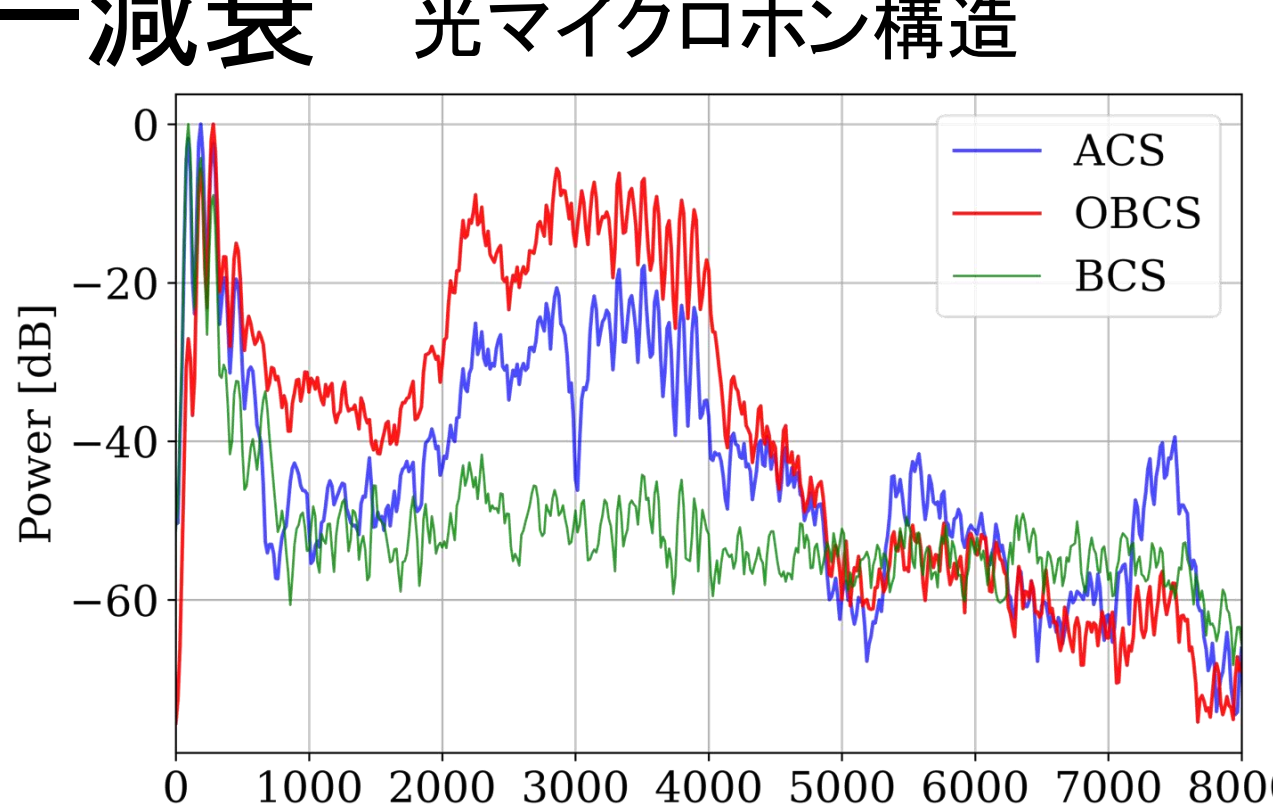
## 1. はじめに

- MRI: 強磁場が発生し体内断面画像を撮像
    - 筐体の振動による約100 dBAの雑音
    - 被撮像者への身体的, 精神的苦痛を伴う
    - ↳ 強磁場・高騒音下でも利用可能な音声コミュニケーションが必要
  - 気導音 : 雑音に脆弱, 磁場環境で使用難
  - 骨伝導音: 耐雑音性
    - ↳ 高周波成分減衰, 磁場環境で使用難
    - ↳ 機器が磁性体を含むため
- ➡ **光骨伝導音の利用**


- 光骨伝導音: 接触型光マイクロホンによる収録音声[1]
  - 気導雑音に頑健
  - 非磁性体で構成
  - 骨伝導音より高音質
    - 高周波数域が減衰しにくい
  - 基本周波数帯エネルギー減衰
    - ↳ **声質が変化**



世界で特許登録済 [1] 光マイクロホン構造



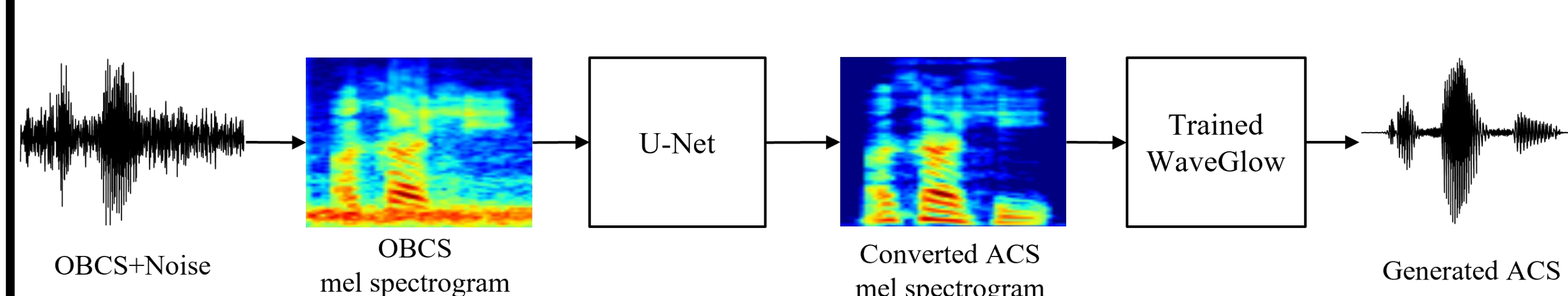
各音声の周波数応答



使用した光マイクロホン

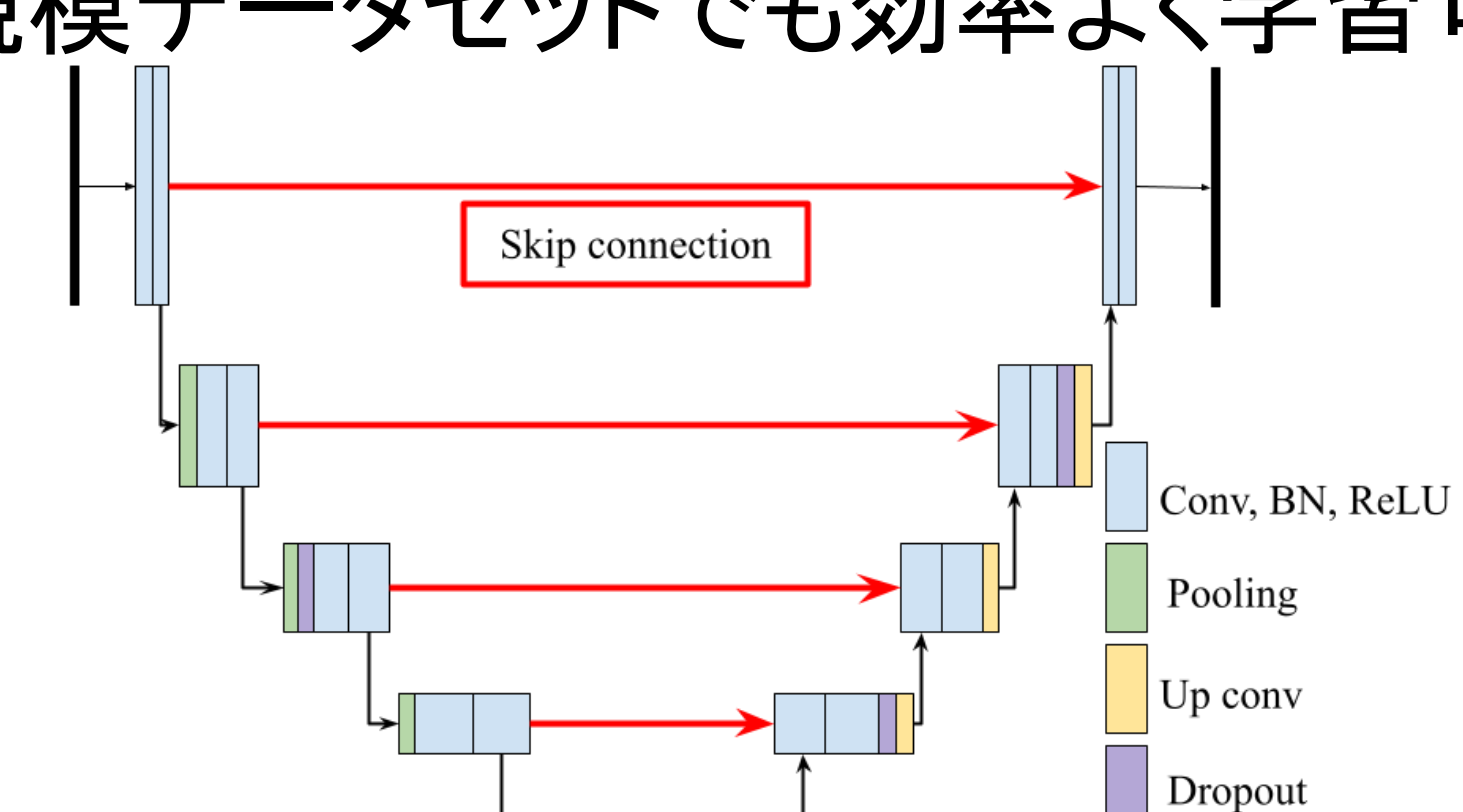
## 2. 提案手法

- 深層学習を用いた音質改善システム
- 雑音付加した光骨伝導音からクリーンな気導音に変換
  - メルスペクトrogramベースの音質改善
  - U-Net型モデルによる変換
    - 雑音除去&声質変換
  - 学習済みWaveGlow [2] による気導音生成



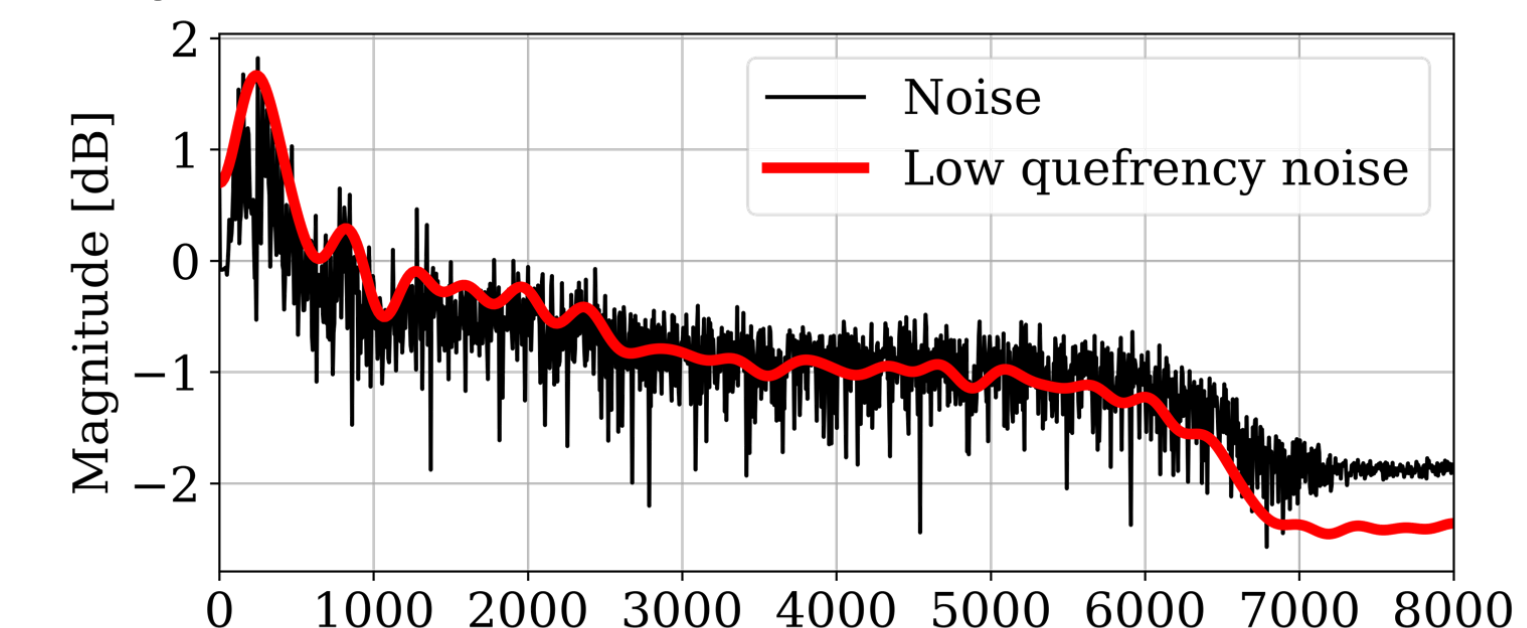
提案手法のフロー

- U-Net型変換モデル
  - エンコーダデコーダモデル
    - エンコーダ部: 高分解能特徴抽出
    - デコーダ部: 得られた特徴の正確な位置情報
  - スキップ接続による入力特徴の付与
    - 小規模データセットでも効率よく学習可



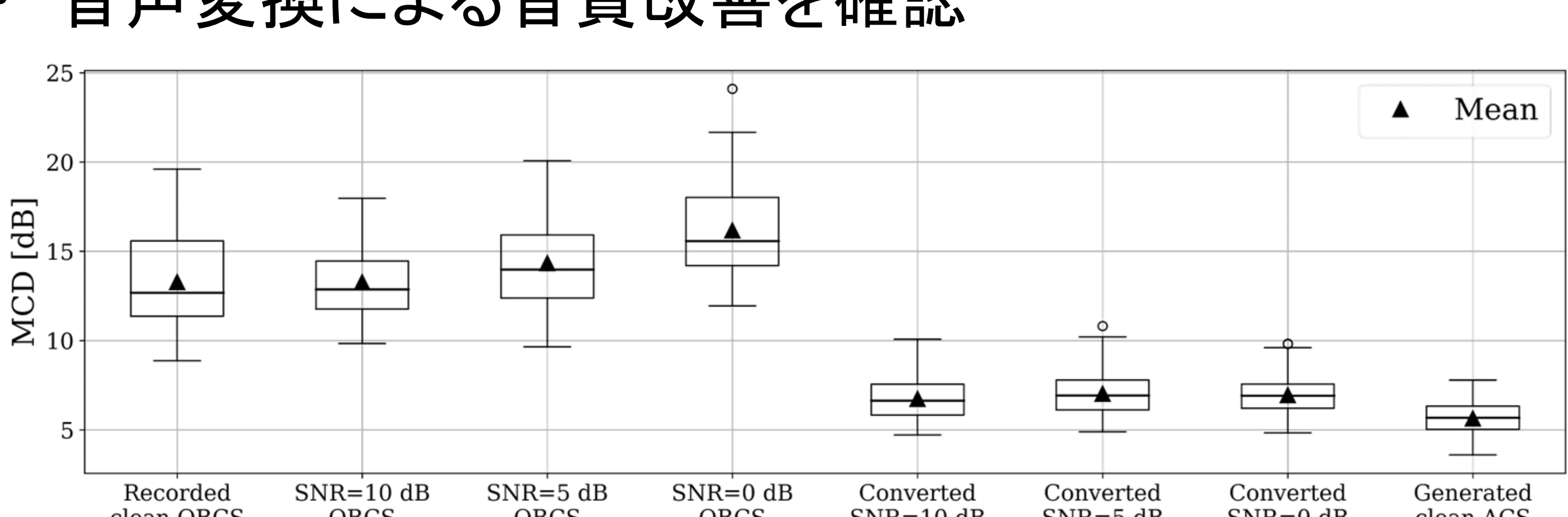
使用モデル構造

## 3. 収録

- 無響室で気導音, 光骨伝導音, 骨伝導音を同時収録
    - 男性1名
    - 東北大-松下单語音声データベース212単語×3回
    - パラレルデータの作成, 学習, 評価に利用
  - MRI室内で光骨伝導音を収録
    - MRI駆動雑音の推定に利用
  - MRI環境下での疑似雑音の作成
    - MRI室内光骨導音の無発声区間からMRI駆動音を抽出
    - ケプストラムによるスペクトル概形推定・フィルタ作成
    - ホワイトノイズにフィルタリングして疑似雑音を作成
  - 光骨伝導音に疑似雑音付加
    - SNR=10, 5及び0 dBの3パターン用意
- 
- 雑音の周波数特性

## 4. 実験

- 訓練500, 評価100データを対象に実験
- メルケプストラム歪み(MCD)による客観評価
  - 収録気導音を基準とした各音声の距離評価
- 音声変換による音質改善を確認



MCDによる音質改善結果

## 5. まとめ

- MRI室内の光骨伝導音の音質改善システムを提案
  - 客観評価における音質改善を確認
  - 前処理の追加の影響調査
  - 複数話者に対応する汎用モデル構築

## 参考文献

[1] Optoacoustics, "OPTIMIC™ MICROPHONES", <https://www.optoacoustics.com/industrial/optimic-microphones>  
 [2] Pytorch, "WAVEGLOW", NVIDIA, [https://pytorch.org/hub/nvidia\\_deeplearningexamples\\_waverglow/](https://pytorch.org/hub/nvidia_deeplearningexamples_waverglow/)