

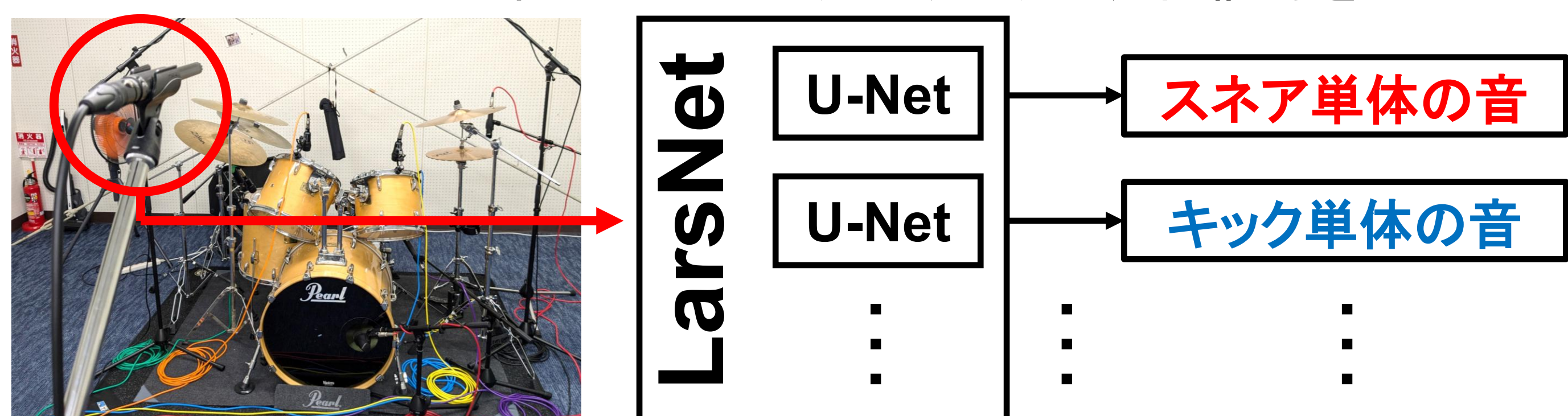
## 1. 研究背景

- ドラムセットを収録する際のマイキング
  - スネアやハイハットなど各ドラムパーツに近接させてマイクを配置  
→ **マルチトラック録音**と呼ばれる
  - 近接しているドラムパーツの音(目的音)のみ録音することを狙う
- ドラムセットの録音における被り音の問題
  - 目的音以外に混ざり込む音を**被り音**という
  - 音源毎の音質調整(ゲイン, EQ, コンプ等)が被り音にも適用されてしまう
    - ミキシング品質が低下
  - 各近接マイクの録音信号中の**目的音**のみを残し, **被り音**は抑圧することが望まれる



## 2. ドラムセットの音源分離の歴史

- ドラムセットの音源分離の研究
  - 同時発音が多い, 全て打撃音で構成 という特徴から高難易度
  - 類似研究: ドラムの自動採譜を目的とした分離処理が過去に研究されてきた ([Lindsay-Smith+, 20012], [Gillet, 2008]等)
- 深層ニューラルネットワーク** (deep neural network: **DNN**)を用いたドラム音源分離: drums source separation (DSS) [Mezza+, 2024]
  - ドラムセットの各ドラムパーツ毎の大規模なデータセットを作成
  - U-Net [Ronneberger+, 2015]を使用したLarsNet [Mezza+, 2024]を作成
  - LarsNetの入力は単一チャンネル(モノラル)の混合信号を与える



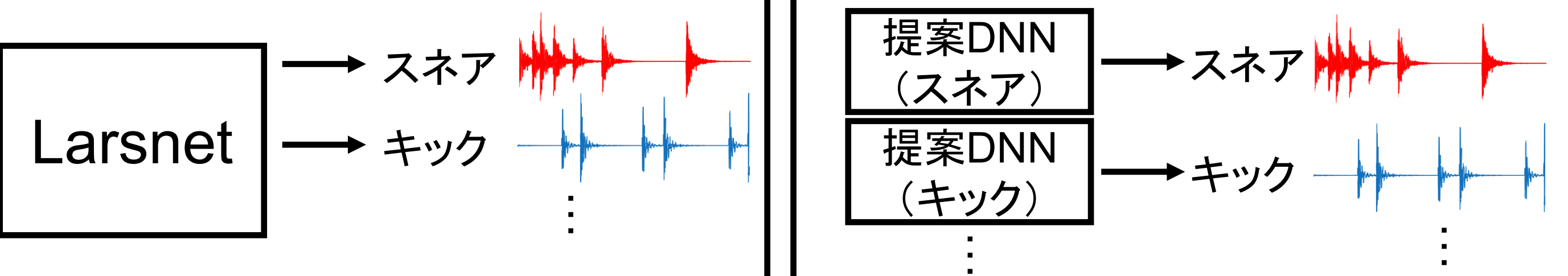
## 3. 提案手法

## 従来手法(LarsNet)

- U-Net [Ronneberger+, 2015]アーキテクチャを使用
  - 時間周波数領域の信号を入力しU-Netで音源分離
- モノラル信号を入力
  - 全てのドラムパーツの音が混合しているため音源分離の難易度が高い
  - 単一チャンネルのためマイク間の相対情報が利用不可
- 単一モデルで全音源を推定
  - 学習と予測の難易度は高い

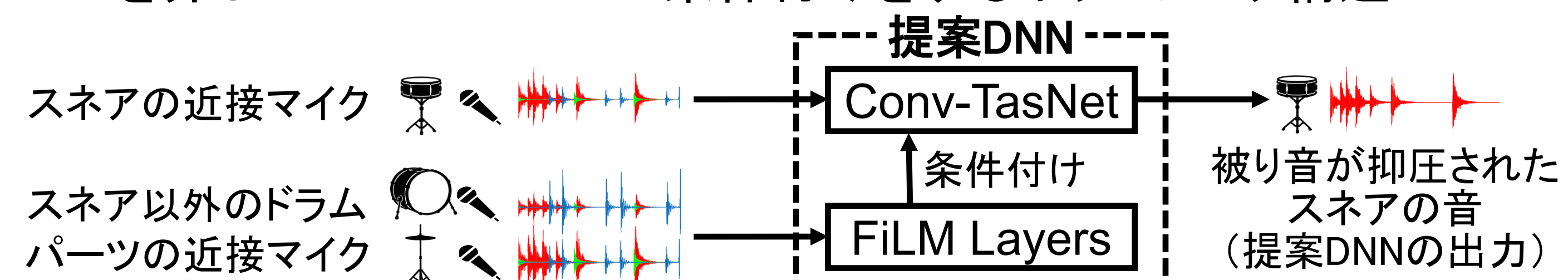
## 提案手法

- Conv-TasNet [Luo+, 2019]アーキテクチャを使用
  - 時間領域の音源分離手法であり強力なモデル
- マルチトラック信号を入力
  - 近接マイクなので観測時点で目的音と被り音のSN比が良い
  - 多チャンネルのためマイク間の相対情報を利用可
- 音源毎に専用モデルを作成
  - 学習と予測の難易度が低下

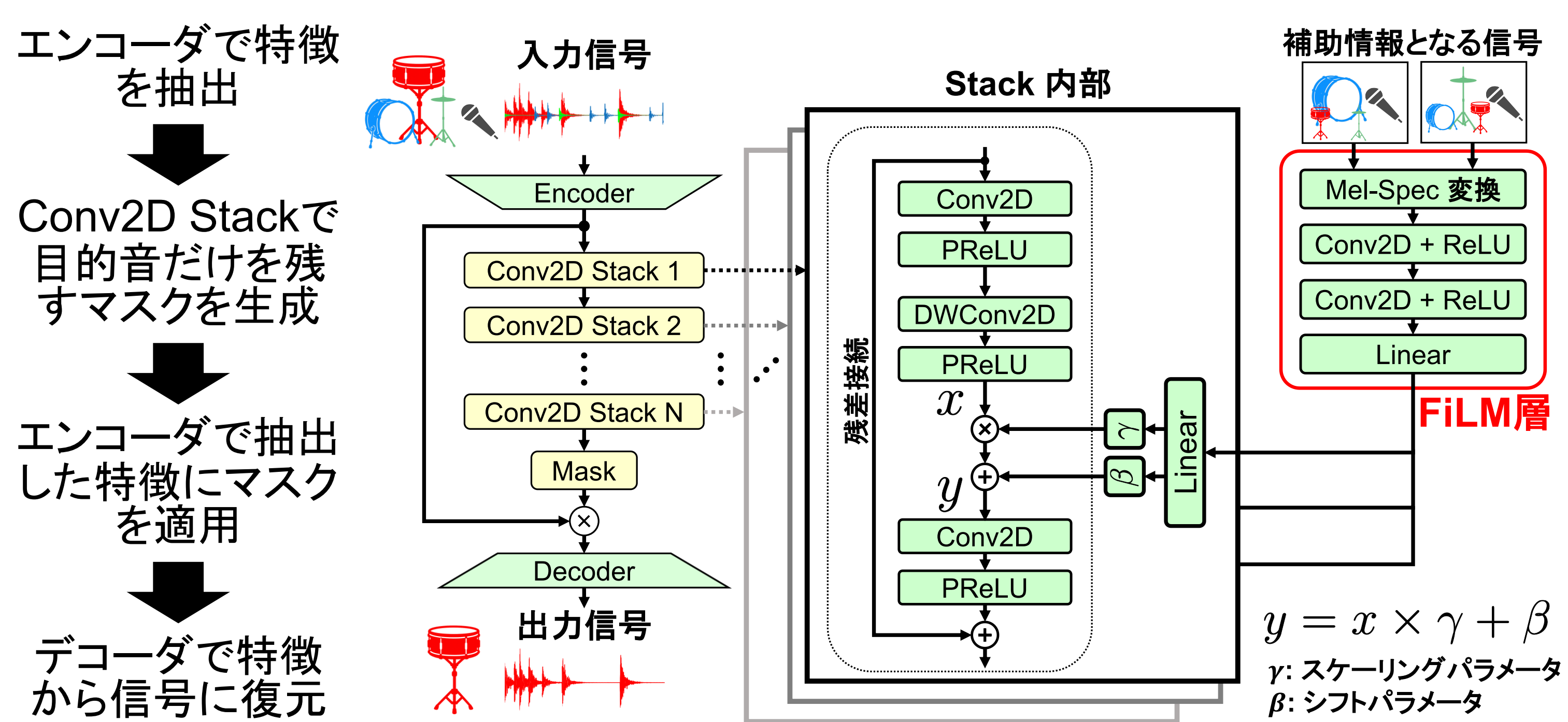


- 提案手法の多チャンネル信号の入力方法(スネア目的音の例)

- メインとなる入力: スネアの近接マイクの信号
- 補助情報となる入力: 他ドラムパーツの近接マイクの信号
  - スネア近接マイク中の被り音を高SN比で録音  
→ **被り音抑圧における強力なヒントとして利用可**
- 補助情報はfeature-wise linear modulation (FiLM) 層 [Ethan+, 2018]を介してConv-TasNetに条件付けをするネットワーク構造



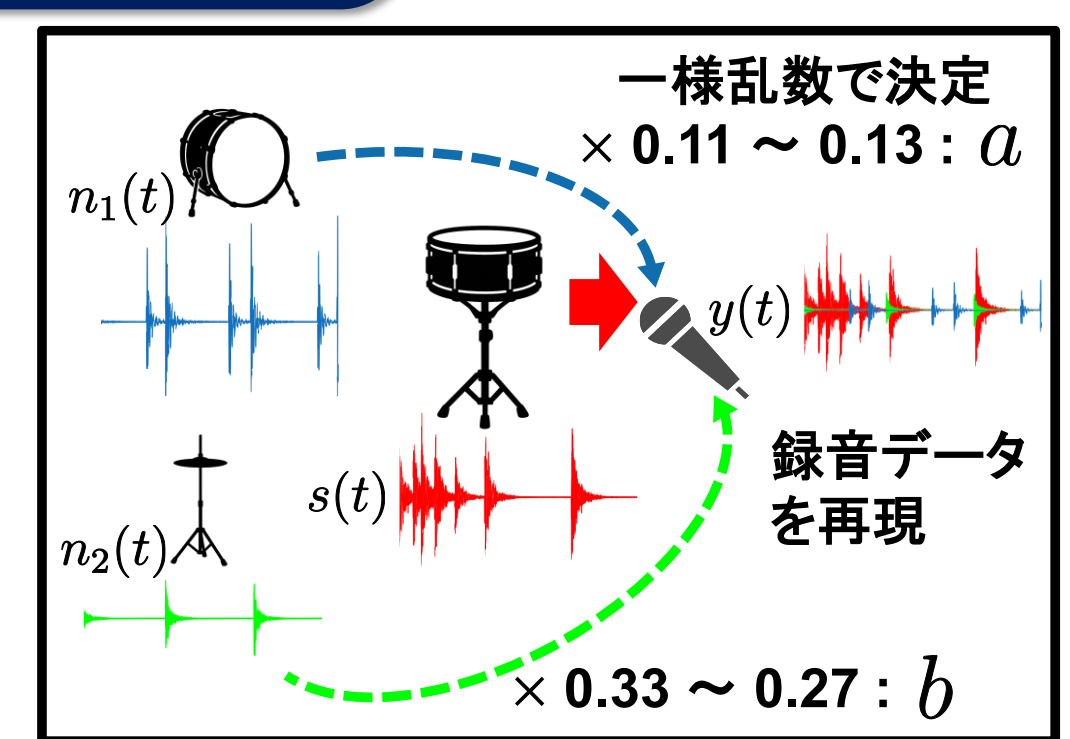
- 提案手法のネットワークアーキテクチャ



## 4. 被り音抑圧実験

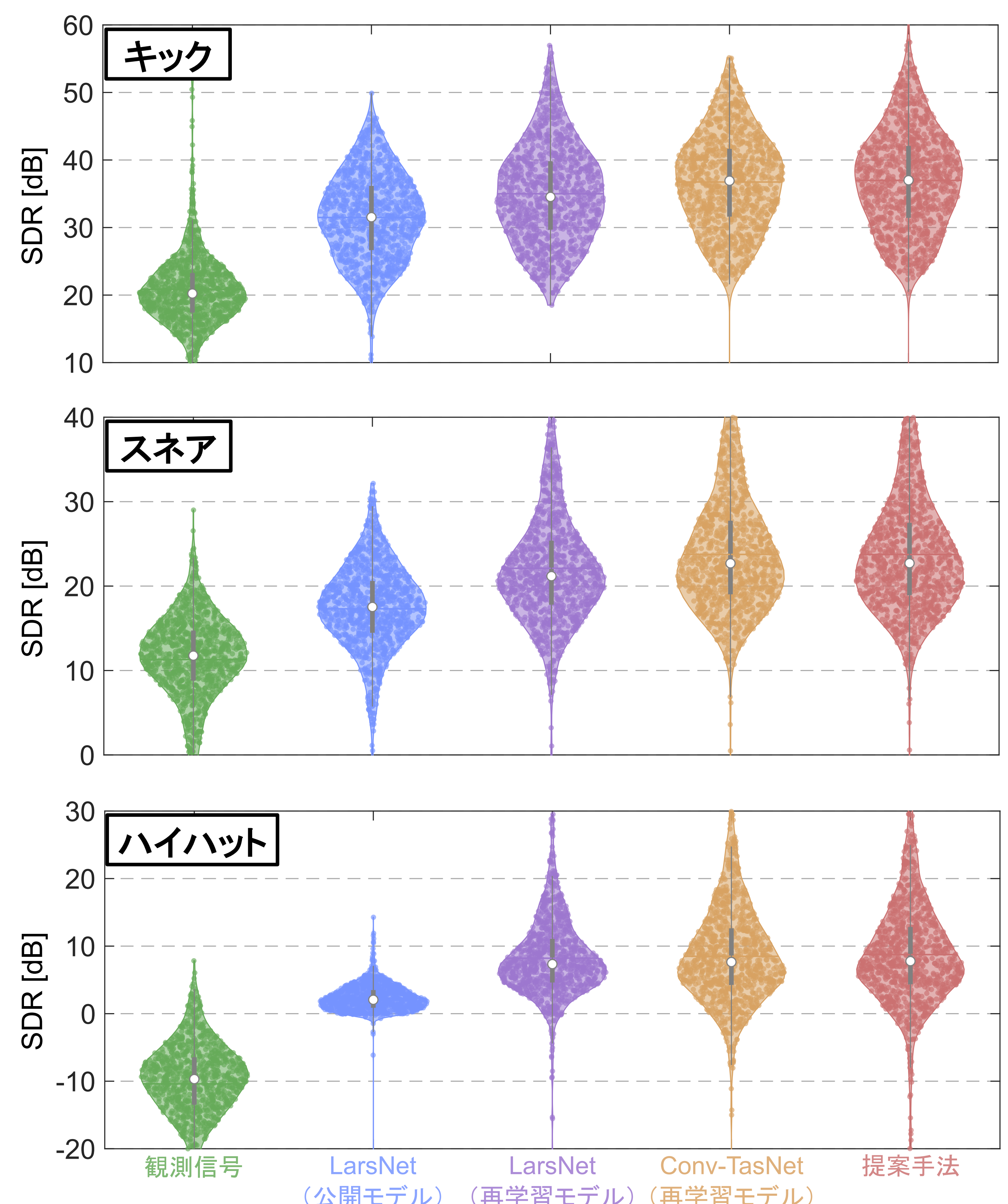
- データセット

- StemGMD [Mezza+, 2024]
  - 人の演奏をMIDIデータ経由で音源化
  - 各ドラムパーツ(9種)の音を収録
- ドラムセットを実際にマルチトラック録音し被り音の振幅を計測
  - 被り音の模擬にその振幅値  $a, b$  を使用  $y(t) = s(t) + an_1(t) + bn_2(t)$



- 実験条件

- キック, スネア, ハイハットの3音源に限定して実験を行う
- 比較手法
  - LarsNet(公開モデル): 単一チャンネルの信号で学習したモデル
  - LarsNet(再学習モデル): 目的音の近接マイク信号で学習したモデル
  - Conv-TasNet(再学習モデル): 目的音の近接マイク信号で学習したモデル
  - 提案手法: 被り音の近接マイク信号を補助情報として入力するモデル
- 評価指標: 信号対歪み比 (source-to-distortion ratio: SDR) [Vincent+, 2006]
  - 「目的音の音質」と「被り音の抑圧量」の両方を加味した評価尺度



- 考察

- Conv-TasNetのSDR値がLarsNetのSDR値より高い
- FiLM層の有無では結果に大きな違いが見られない
  - 被り音抑圧というタスクが音源分離と比較して簡単であり, 提案手法のアーキテクチャでは補助情報による性能改善が十分ではなかった