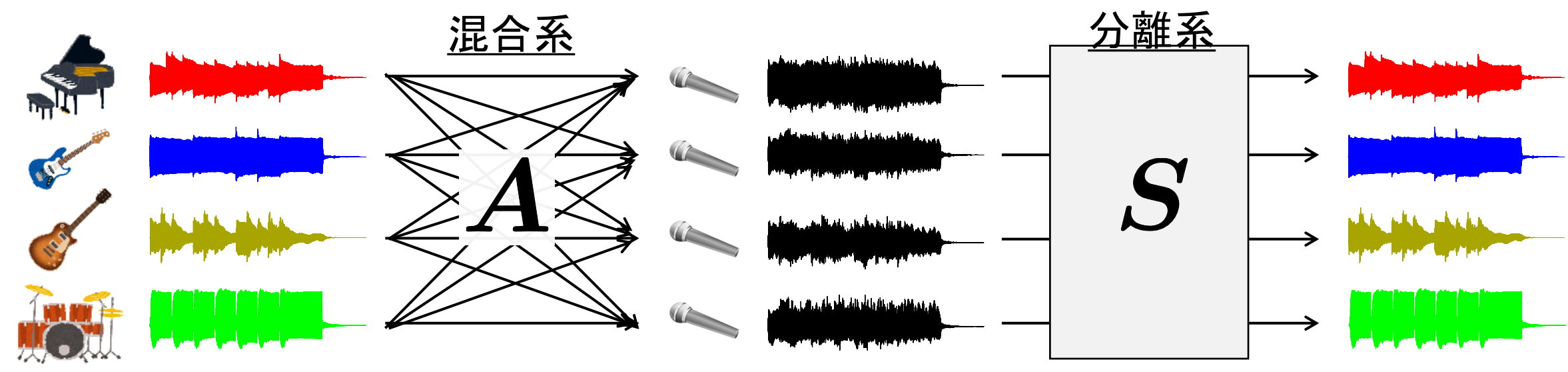
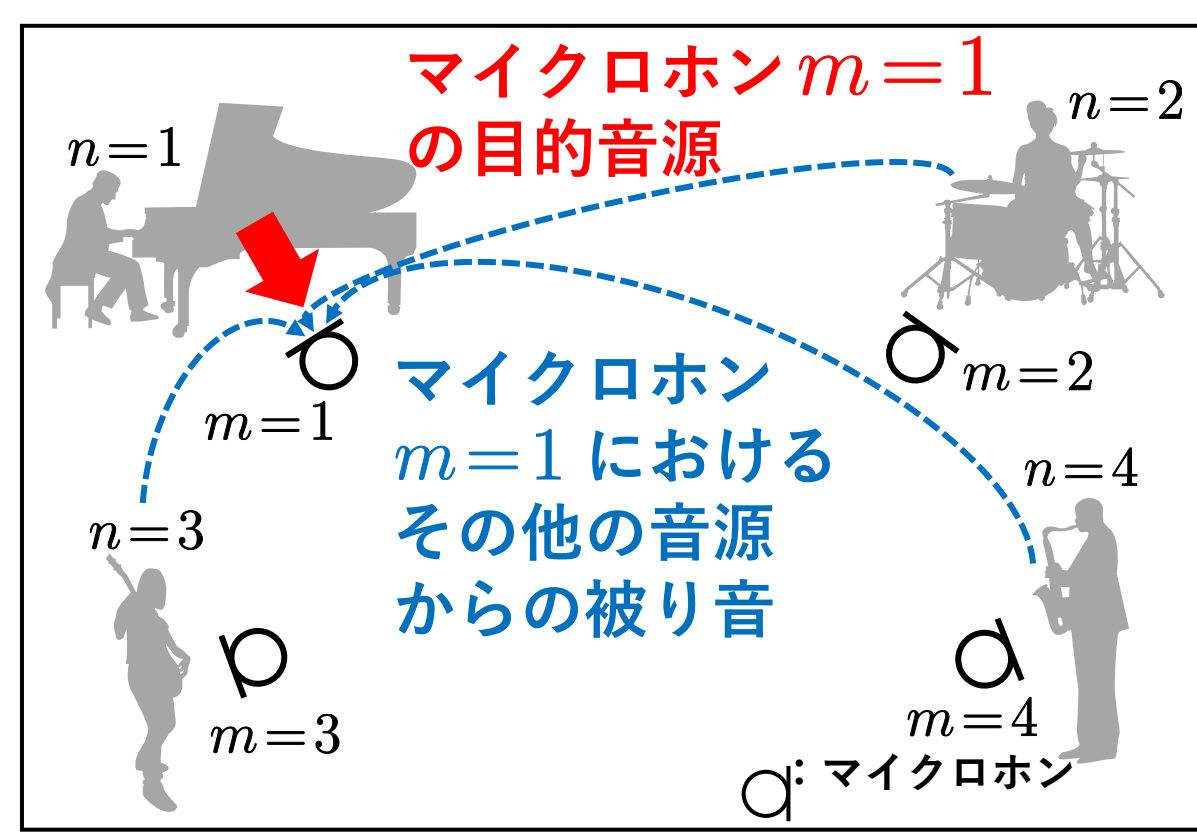
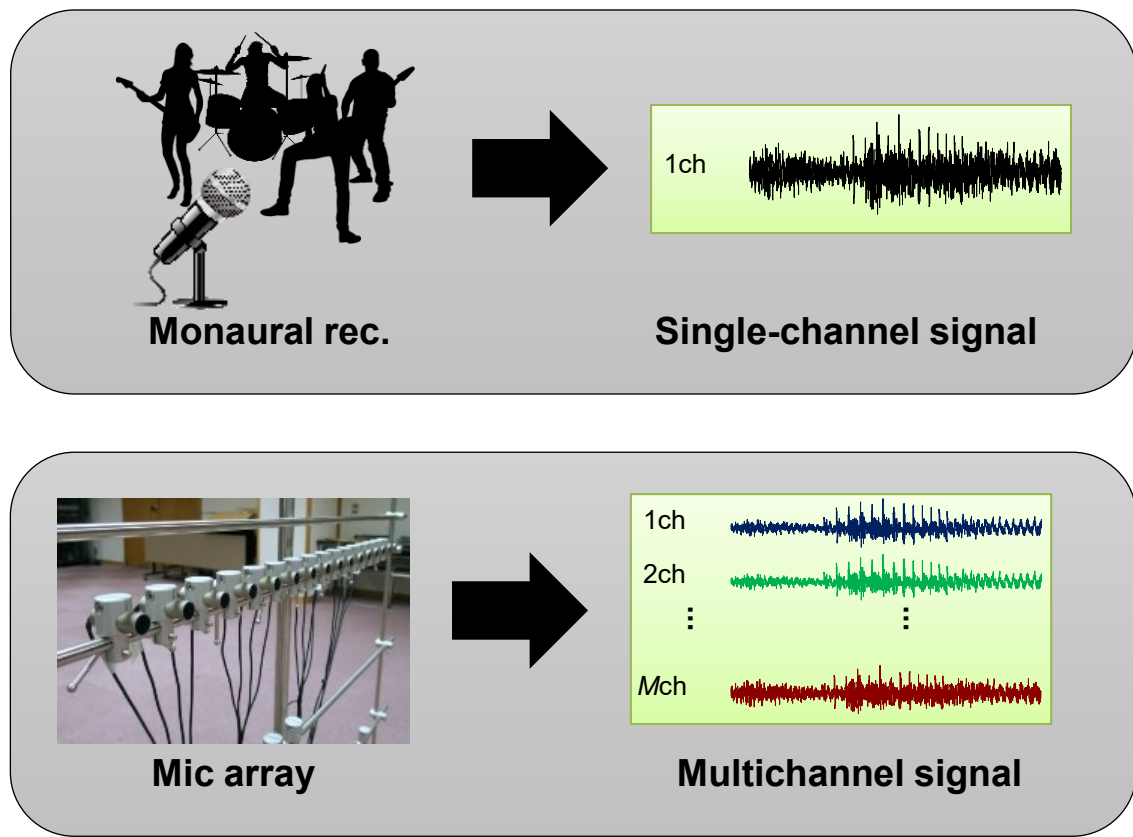


1. 実験内容

- ブラインド音源分離 (blind source separation: BSS)
 - 音源やマイクロホンの空間的な配置 (混合系 A) が不明な状態で分離系 S を推定する技術



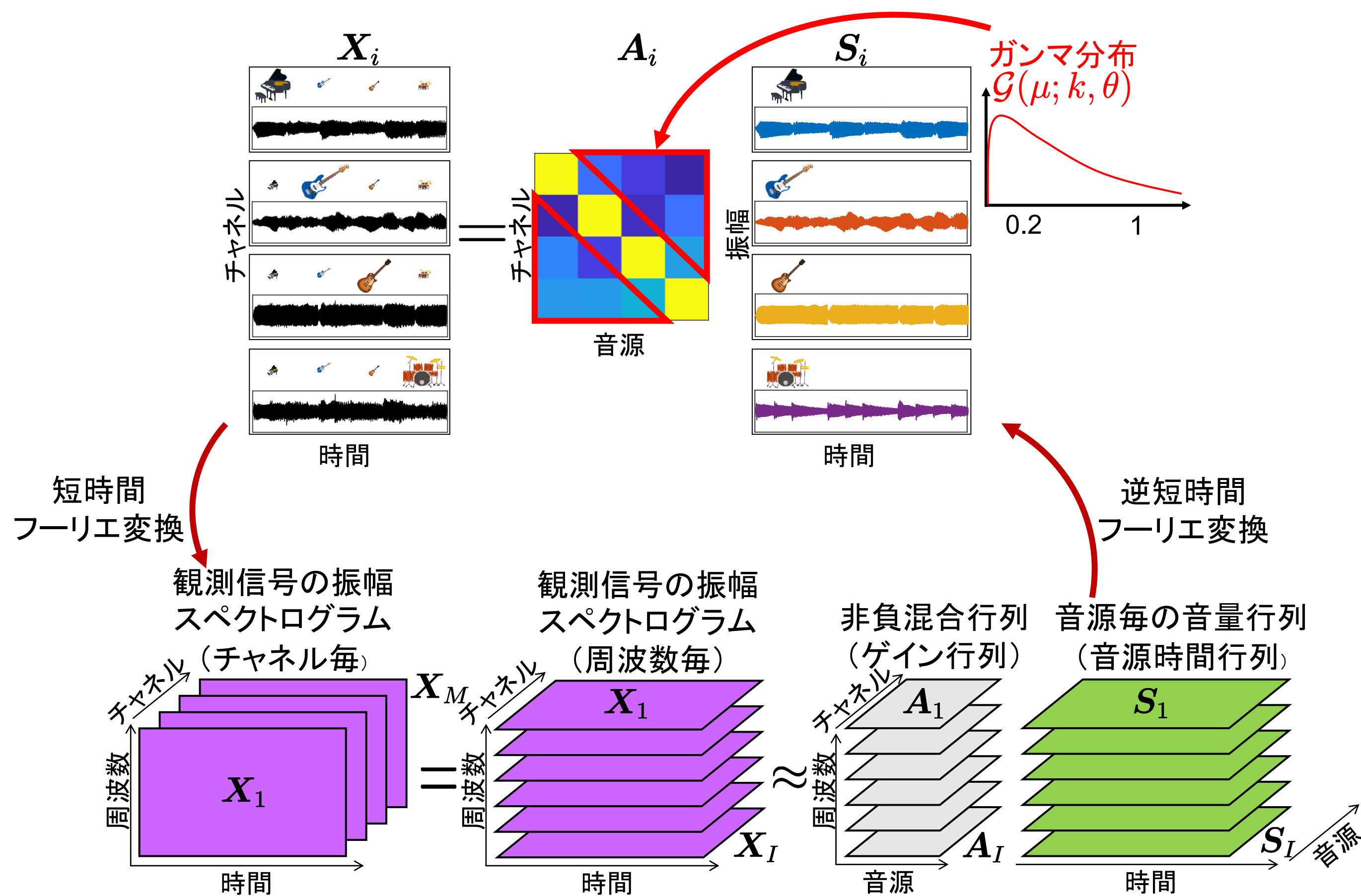
- 被り音 (bleeding sound)
 - 楽器やライブ演奏の録音時, 目的の音のみを録音するために楽器等にマイクを近接
 - 実際には目的音以外の音 (被り音) が混入
 - ミキシングの質や演奏の質低下の原因



- 非負値行列因子分解 (nonnegative matrix factorization: NMF) [Lee+, 1999]
 - 非負制約付きの任意基底数 (K 本) による低ランク近似
 - 限られた数の非負基底ベクトルとそれらの非負係数を抽出
 - 両辺の値が可能な限り近くなるような A と S を推定

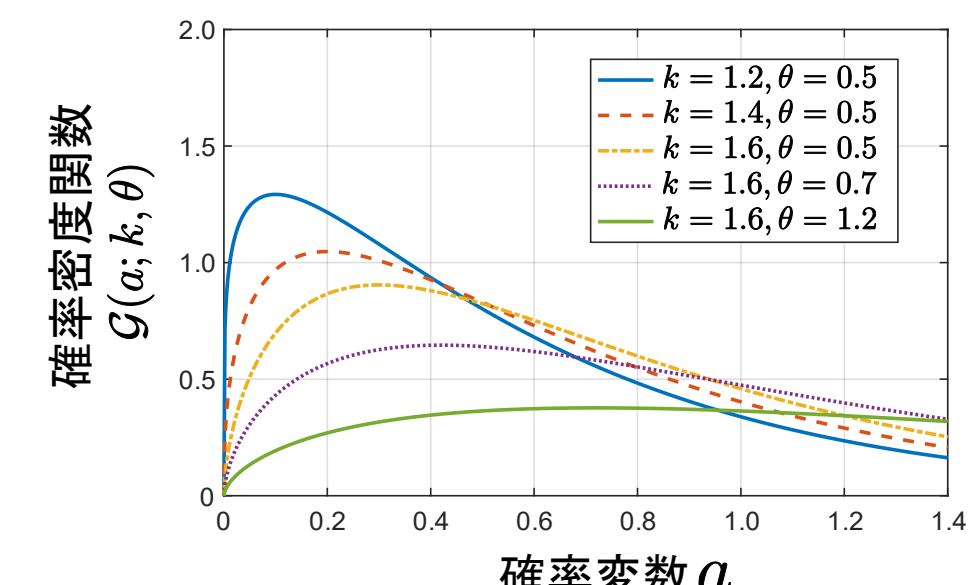
$$\begin{matrix} \text{非負値観測行列} \\ X (M \times J) \\ M \begin{pmatrix} 1 & 2 & 0 & 0 & 1 \\ 1 & 3 & 1 & 2 & 2 \\ 0 & 0 & 3 & 5 & 3 \end{pmatrix} \\ J \end{matrix} \approx \begin{matrix} \text{基底行列} \\ A (M \times K) \\ M \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 3 \end{pmatrix} \\ K \end{matrix} \begin{matrix} \text{係数行列} \\ S (K \times J) \\ \text{アクティベーション} \\ \begin{pmatrix} 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 2 & 1 \end{pmatrix} \\ J \end{matrix} K$$

- 基底行列 A と係数行列 S の求め方
 - 左辺 X と右辺 AS の距離を最小化するような A と S を求める最適化
 - A と S を乱数で初期化して反復的な計算をすることで距離が最小化されるアルゴリズムが提案されている [Lee+, 2000]
- 最大事後確率 (maximum a posteriori: MAP) 推定に基づく時間チャンネルNMF (time channel NMF: TCNMF) [Mizobuchi&Kitamura+, 2021]
 - ゲイン行列 A_i にガンマ事前分布を導入し, X_i にNMFを適用



$$a_{imn} \sim \begin{cases} \delta(a_{imn} - 1) & (m = n) \text{ 対角要素} \\ G(a_{imn}; k, \theta) & (m \neq n) \text{ 非対角要素} \end{cases}$$

$$G(a; k, \theta) = \frac{1}{\Gamma(k)\theta^k} a^{k-1} e^{-a/\theta}$$



目的関数

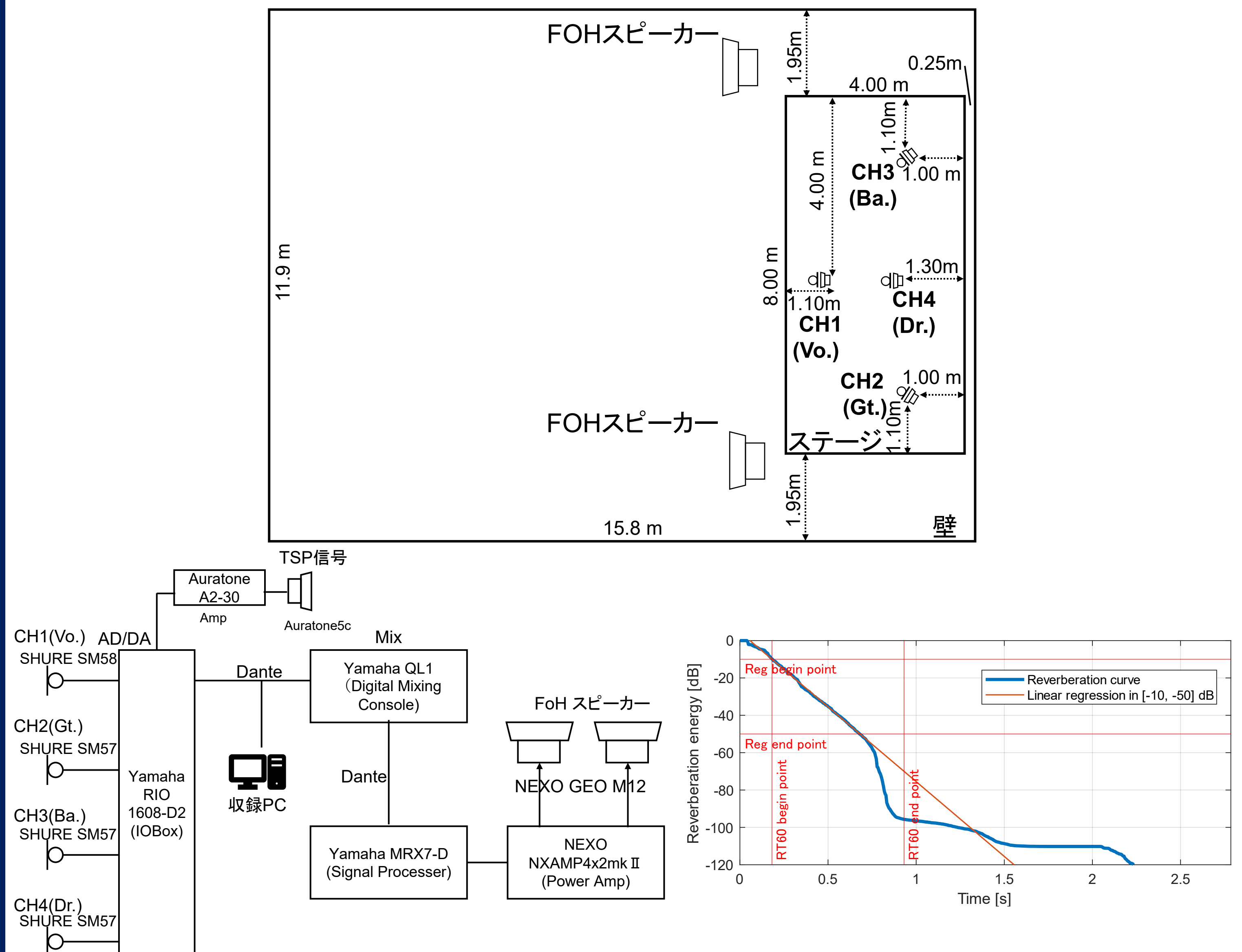
$$\text{Minimize}_{A, S} \sum_i \mathcal{D}_{KL}(X_i | A_i S_i) + \sum_{i, m, n \neq m} \left[-(k-1) \log a_{imn} + \frac{1}{\theta} a_{imn} \right]$$

s.t. $a_{imn}, s_{inj} \geq 0 \forall i, m, n, j$ and $a_{imn} = 1 \forall m = n$

最適化には補助関数法 [Hunter+, 2004]を用いる

2. 実験方法

- あるライブハウスのステージ上でマイクを配置し測定したインパルス応答を使う



- あるデータを使って被り音混合系 (非対角要素) が本当にガンマ分布に従うのかを確認する

3. 実験結果

近接マイクを基準とした被り音の相対的な平均パワー値 対数 (dB)

| 近接マイク基準の相対平均パワー値 (dB) | 音源1 (Vo.) | 音源2 (Gt.) | 音源3 (Ba.) | 音源4 (Dr.) |
|-----------------------|--------------|--------------|--------------|--------------|
| マイク1 (Vo.) | 0 | -14.57444485 | -9.320776897 | 8.739559919 |
| マイク2 (Gt.) | -51.1849096 | 0 | -32.72267359 | -16.2755213 |
| マイク3 (Ba.) | -53.96221219 | -39.78541989 | 0 | -18.99492992 |
| マイク4 (Dr.) | -65.85831146 | -51.55535783 | -46.85608454 | 0 |

ステージ上の配置とゲイン調整の関係でVo.マイクはDr.の方が強い

- ゲイン行列の非対角要素の値は0を取ることが多かった
 - ヒストグラムの左端ビンに属する要素が全ての非対角要素で支配的
 - ゲイン行列には0になるべき要素がかなり多く存在する
- 多くのヒストグラムで指数分布のような形状を確認
 - しかし, 指数事前分布ではTCNMFの自明解を回避できない
 - ゲイン行列の非対角要素の値がほぼ0のビンを無視すればガンマ分布に近いデータもある
 - Dr.音源のVo.マイクへの被り音のように被り音の相対平均パワーが大きくなると「形状母数1より大きいガンマ分布」に近づくように見える
 - 「被り音のパワー」と「ヒストグラム」はかなり相関がありそう

