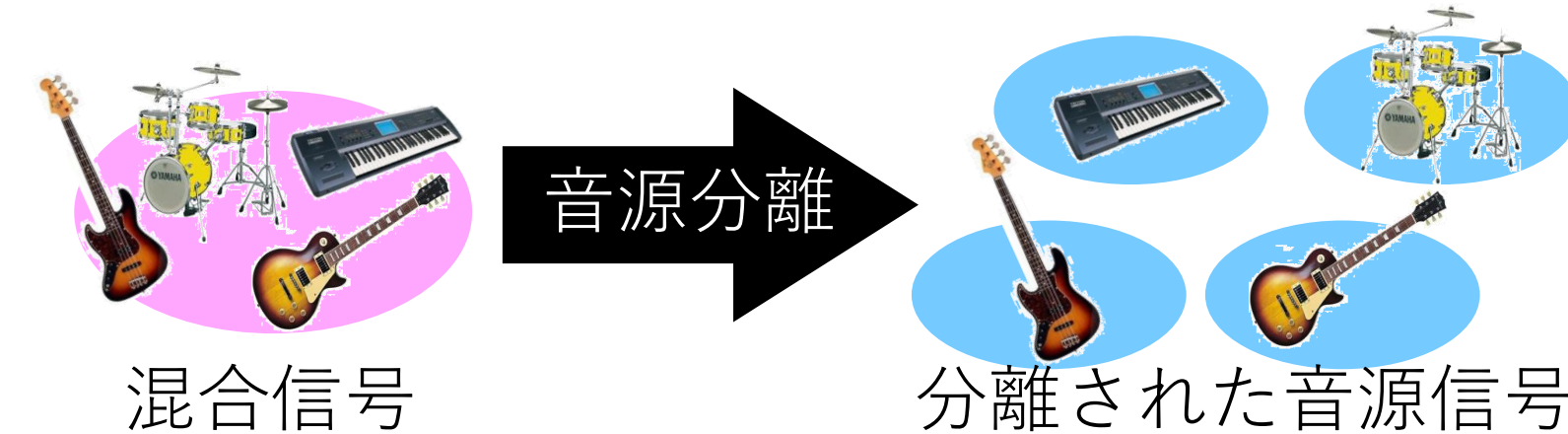


1. 研究背景

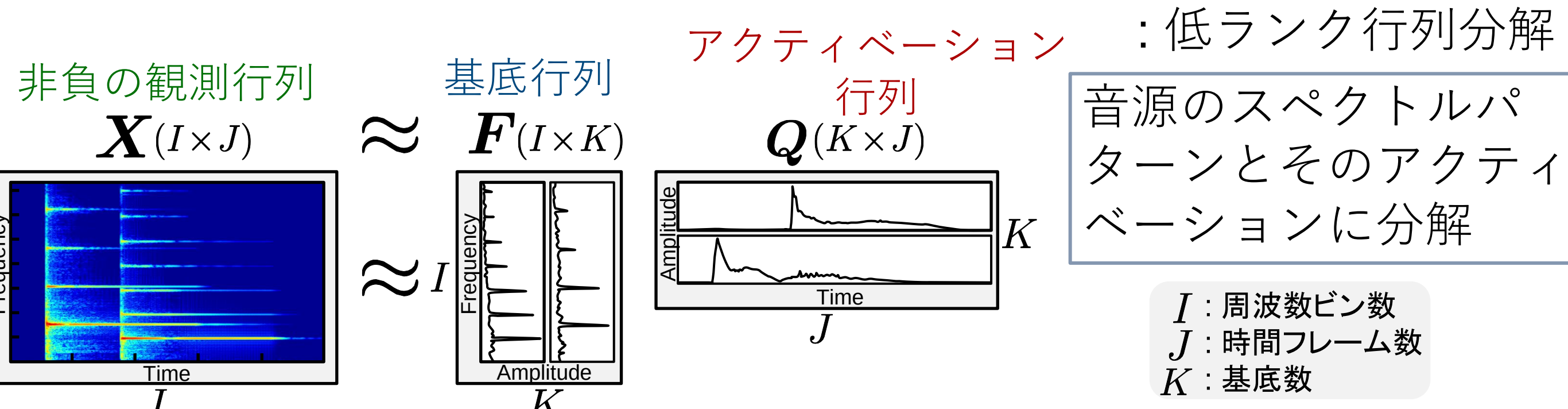
- 音源分離：複数の音声や楽器音の混合音から個々の音源を推定する技術



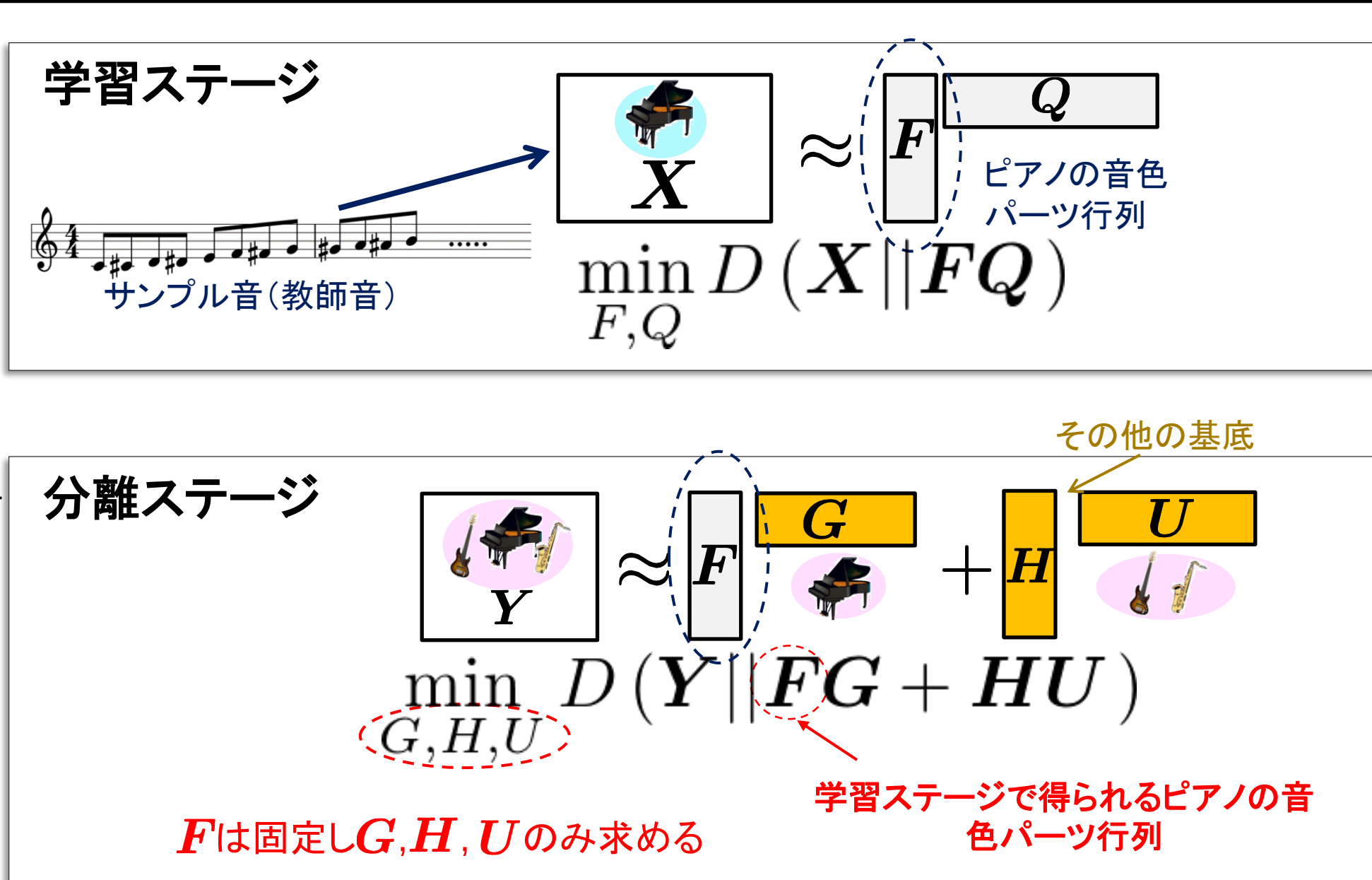
応用例

- 自動採譜の前段処理
- 音声認識の精度向上etc.

- 非負値行列因子分解 (nonnegative matrix factorization: NMF)



- 半教師ありNMF (semi-supervised NMF: SNMF)
事前学習した教師基底行列を用いて混合信号の振幅スペクトログラムを分解し目的音源成分 FG と非目的音源成分 HU を推定



解決すべき問題

混合信号中の目的音源と非目的音源が類似した音色を含む場合、2つの基底行列 F と H に類似したスペクトルパターンが現れ、分離精度は劣化

本研究の目的

非目的音源の基底行列 H に、教師基底行列 F と類似したスペクトルパターンが現れないように最適化する新しい罰則条件付きSNMF (penalized SNMF: PSNMF) を提案し、性能を調査

2. 従来手法

- 内積型PSNMF** [Kitamura et al., 2014]
- 教師基底行列 F と非目的音源の基底行列 H ができるだけ直交するように、内積に基づく罰則条件を付与したSNMF

$$\min_{G,H,U} D(Y \| FG + HU) + \mu \|F^T H\|_{Fr}^2$$

内積罰則項

- 小さな値 α を用いて $H \leftarrow \alpha H$ 及び $U \leftarrow (1/\alpha) U$ とすれば、 $FG + HU$ の値を変えずに内積罰則項 $\|F^T H\|_{Fr}^2$ の値を最小化可能

➡ 本質的には、内積罰則項は最適化に影響を与えていない

3. 提案手法

- Log-cos型PSNMF** [岩瀬ら, 2020]
- 内積ではなく、大きさに非依存なコサイン類似度の対数を罰則条件に用いる
- 内積型PSNMFの内積罰則項が最適化に影響しない問題を解決

$$\min_{G,H,U} D(Y \| FG + HU) + \mu \sum_{k,l} \log \frac{\sum_i f_{ik} h_{il}}{\left(\sum_i f_{ik}^2\right)^{\frac{1}{2}} \left(\sum_i h_{il}^2\right)^{\frac{1}{2}}}$$

対数コサイン類似度罰則項

- コサイン類似度の「対数」を用いた理由は、補助関数法に基づく最適化の反復更新式導出を容易にするため（分数関数を対数で引き算に分解すると簡単）

➡ 対数により罰則項の下限値は $-\infty$ となり、最適化が不安定

- Cos型PSNMF** [岩瀬ら, 2020]
- 対数を用いない純粋なコサイン類似度を罰則条件に用いる
- 上記の内積型PSNMF及びlog-cos型PSNMFの問題を同時に回避可能
- 補助関数法に基づく反復更新式の導出は多少複雑になる
- 最適化問題は次式

$$\min_{G,H,U} D(Y \| FG + HU) + \mu \sum_{k,l} \frac{\sum_i f_{ik} h_{il}}{\left(\sum_i f_{ik}^2\right)^{\frac{1}{2}} \left(\sum_i h_{il}^2\right)^{\frac{1}{2}}}$$

コサイン類似度罰則項

4. 比較実験

実験条件

- 罰則条件のないSNMF, 内積型PSNMF, Log-cos型PSNMF, 及びCos型PSNMFの4手法でsource-to-distortion ratio (SDR) を比較
- 手法毎に最適な重み係数 μ を開発データから求め、テストデータで使用する

窓関数	ハミング窓
窓長	92.9 ms
シフト長	46.4 ms
基底数	目的音源用: 27 非目的音源用: 50
初期値	乱数行列
反復回数	100 回
使用する楽器音信号	公開データセットsongKitamuraより2楽器音の混合信号を90曲作成
最適な重み係数 μ を探索するための開発データ	90曲の中からランダムに45曲選出
最適な μ を使う音源分離テストデータ	学習データ以外の残りの45曲

- 内積型PSNMF及びLog-Cos型PSNMFにおいて、更新毎の基底正規化 (basis normalization: BN) を行う

BNの処理

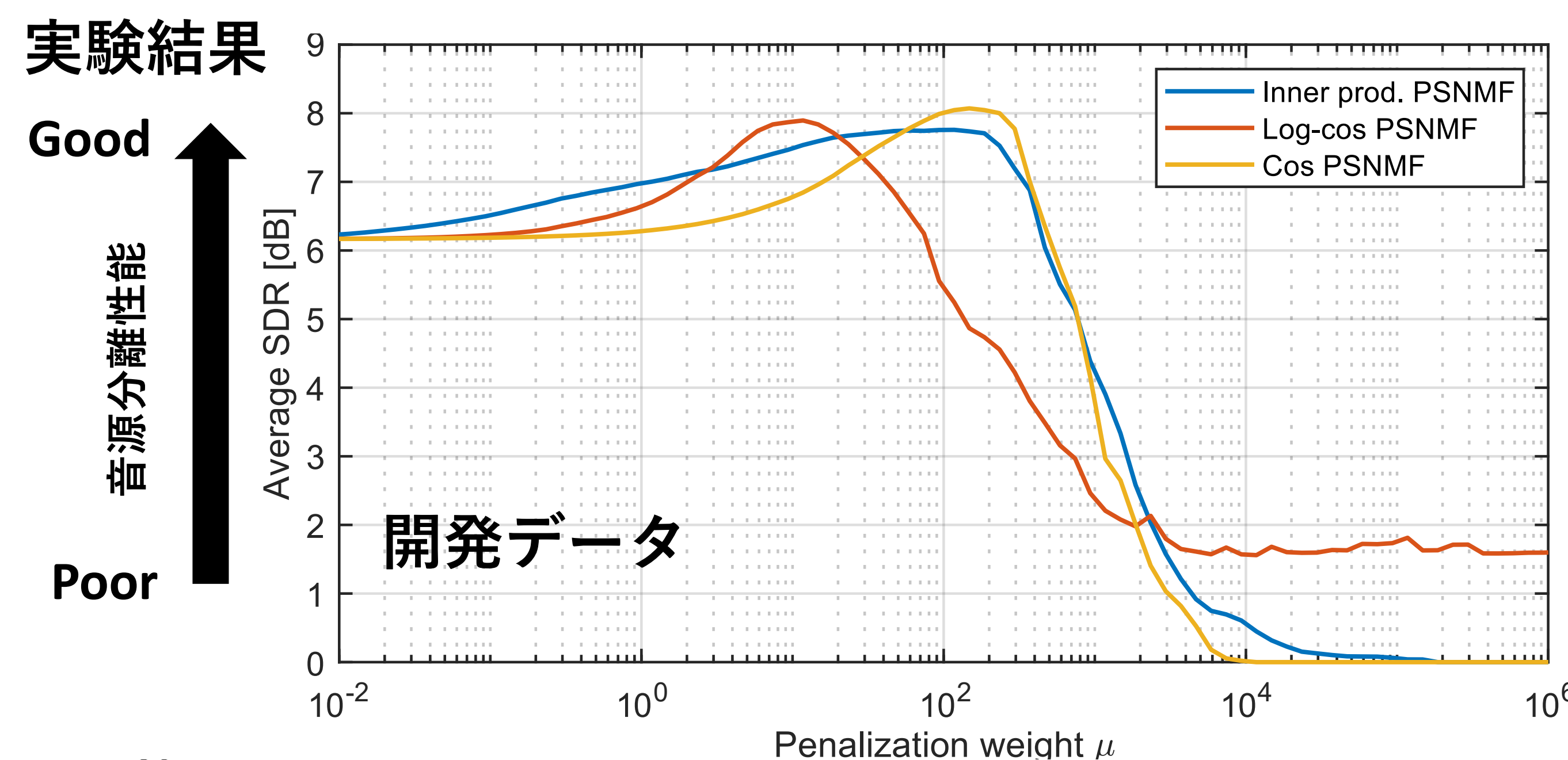
H を更新する度に、 H の全基底ベクトル (列ベクトル) の長さを1に正規化
このとき、基底ベクトル毎の正規化係数 δ_l の逆数を U の行ベクトルに乗じて HU を不変とする

実験結果

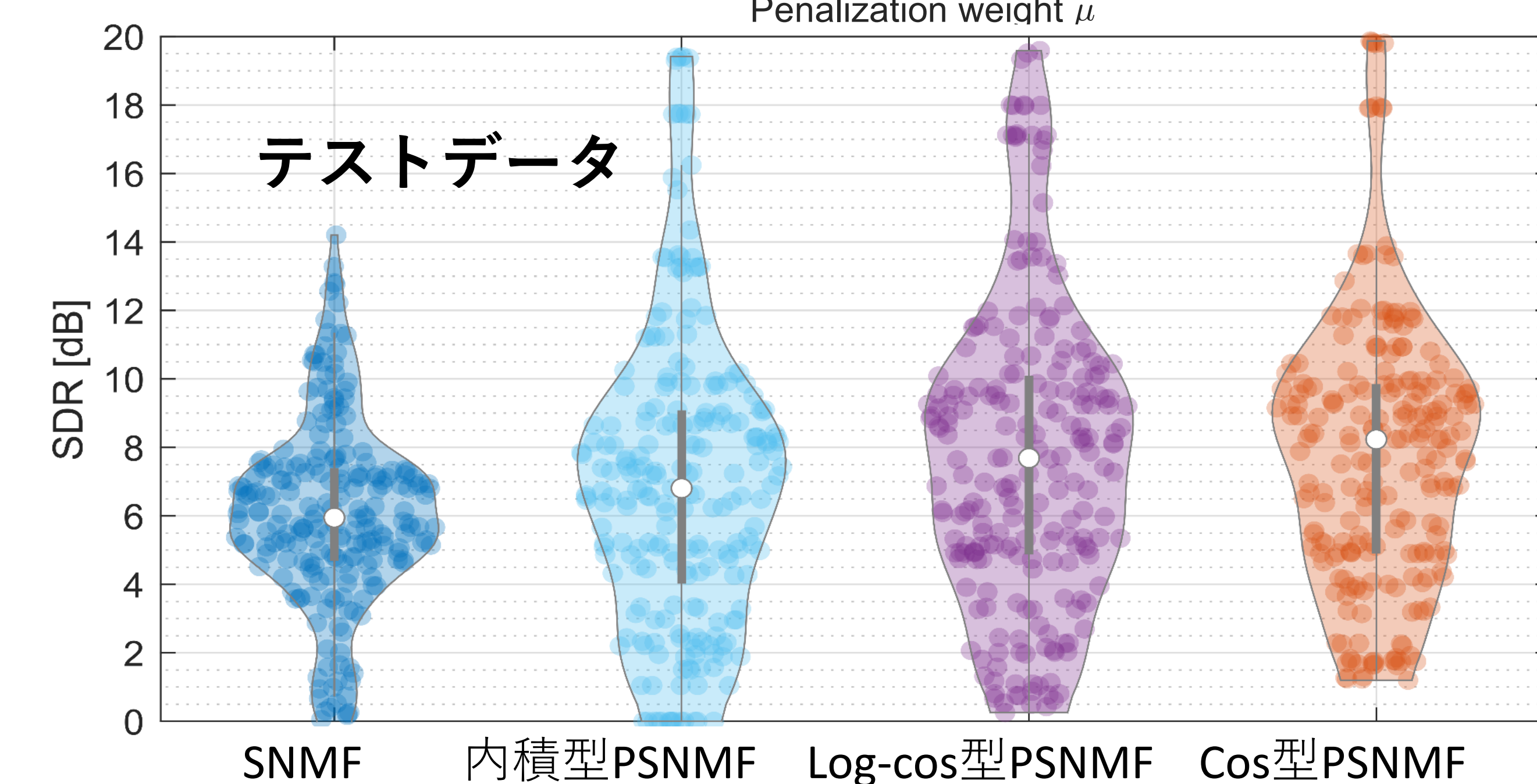
Good

音源分離性能

Poor



開発データ



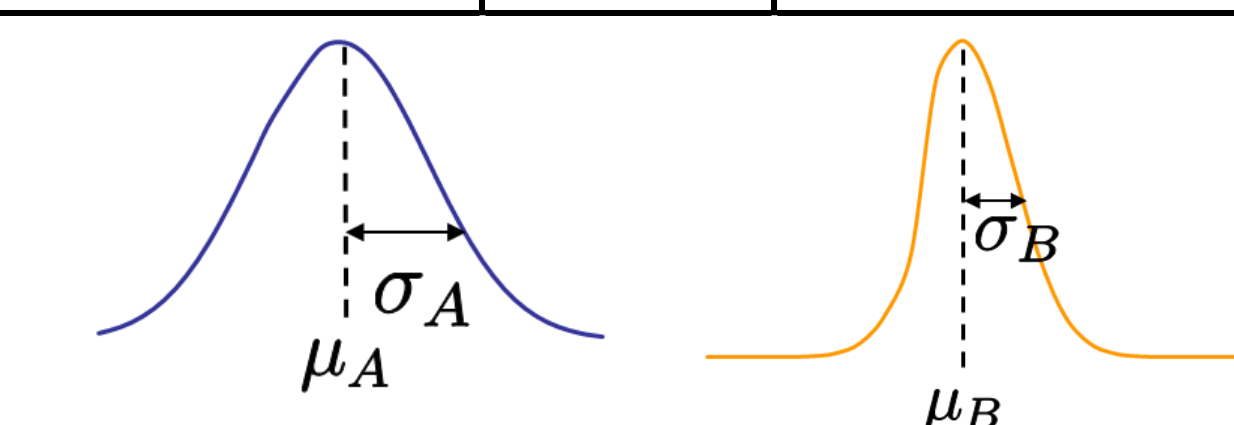
テストデータ

5. 統計的検定

- Welchの (片側) t 検定** [Welch,1947]

- 帰無仮説：各々の正規分布に従う2つの母集団 A 及び B の平均値をそれぞれ μ_A 及び μ_B とするとき、 $\mu_A \geq \mu_B$ である

	SNMF	内積型PSNMF	Log-cos型PSNMF	Cos型PSNMF
SNMF	-	0.0034	0.0000	0.0000
内積型PSNMF	-	-	0.0394	0.0206
Log-cos型PSNMF	-	-	-	0.4290
Cos型PSNMF	-	-	-	-



- Brunner-Munzel検定** [Brunner+,2000]

- 帰無仮説：2つの母集団 A 及び B から抽出した標本をそれぞれ s_A 及び s_B とするとき、 $s_A \geq s_B$ 及び $s_A < s_B$ となる確率が等しい

	SNMF	内積型PSNMF	Log-cos型PSNMF	Cos型PSNMF
SNMF	-	0.0316	0.0001	0.0000
内積型PSNMF	-	-	0.0591	0.0098
Log-cos型PSNMF	-	-	-	0.5977
Cos型PSNMF	-	-	-	-