



# 卒業研究論文

## 論文題目

コサイン類似度罰則条件付き半教師あり非負値行列因子分解と  
音源分離への応用

提出年月日	令和2年2月28日
学 科	電気情報工学科
氏 名	岩瀬 佑太 印
指導教員（主査）	北村 大地 助教 印
副 査	太良尾 浩生 准教授 印
学 科 長	辻 正敏 教授 印

香川高等専門学校

# Semi-supervised nonnegative matrix factorization with cosine similarity penalty and application to audio source separation

Yuta Iwase

Department of Electrical and Computer Engineering  
National Institute of Technology, Kagawa College

## Abstract

I aim to improve the performance of audio source separation for monaural mixture signals that consist of two or more sound sources. In audio source separation for monaural signals, nonnegative matrix factorization (NMF) has been utilized. In particular, semi-supervised NMF (SNMF) can achieve separation by employing simple supervised signals such as scale sounds of each instrument. In recent years, penalized SNMF (PSNMF) with orthogonality was proposed. This method forces two basis matrices for target and interference sources to be orthogonal each other. Since basis matrices have spectral patterns of each source and the similarities between these patterns are maximized by orthogonalization, the separation quality is improved. However, in conventional PSNMF, the orthogonality penalty does not affect correctly because of the scale ambiguity between basis and activation matrices in NMF. In this thesis, I propose a new method that uses cosine similarity between the basis matrices of the target and interference sources to avoid the scale ambiguity. I compare the audio source separation performance of the proposed and conventional methods and show the efficacy of the proposed cosine similarity penalty.

**Keywords:** nonnegative matrix factorization, cosine similarity, audio source separation

## (和訳)

本論文では、2つ以上の音源からなるモノラル混合信号の音源分離の性能の向上を目指す。モノラル信号の音源分離では、非負値行列因子分解 (NMF) と呼ばれるアルゴリズムに基づく手法が有名である。各楽器の音階信号のような簡易的な教師情報を NMF に活用する手法の1つに半教師あり NMF (SNMF) がある。特に、SNMF の分離精度を改善する手法として、直交化に基づく罰則条件付き SNMF (PSNMF) がある。直交化罰則項は、目的音源の基底行列と非目的音源の基底行列を直交化する。基底行列には各音源のスペクトルパターンが含まれており、互いの基底行列の非類似度を直交化によって高めることにより分離精度が向上する。しかし、従来の PSNMF では、基底行列とアクティベーション行列間のスケール不定性により、直交化の罰則条件が正しく作用しない定式化となっている。そこで本論文では、PSNMF をスケール不定性に依らない形で再定式化する。提案手法では、目的音源と非目的音源の基底行列のコサイン類似度を罰則条件として付与する。提案手法と従来手法の音源分離の性能を比較し、従来手法よりも優れた音源分離が達成できることを実験により示す。

# 目次

第 1 章	緒言	1
1.1	音源分離の背景 . . . . .	1
1.2	本論における主題 . . . . .	1
1.3	本論文の構成 . . . . .	3
第 2 章	NMF を用いた音源分離	4
2.1	NMF . . . . .	4
2.2	短時間フーリエ変換 . . . . .	6
2.3	NMF の音響信号への適用 . . . . .	6
2.4	SNMF による音源分離 . . . . .	8
第 3 章	コサイン類似度に基づく PSNMF	12
3.1	動機 . . . . .	12
3.2	コスト関数 . . . . .	12
3.3	反復更新式の導出 . . . . .	13
第 4 章	音楽信号の分離実験	18
4.1	実験条件 . . . . .	18
4.2	実験結果 . . . . .	18
第 5 章	結言	51
	謝辞	52
付録 A	補助関数法で利用される不等式	56
A.1	Jensen の不等式 . . . . .	56
A.2	接線不等式 . . . . .	56



# 第 1 章

## 緒言

### 1.1 音源分離の背景

音源分離とは、複数の音源が混合した状態で観測された信号から、混合する前の個々の音源を分離・抽出する技術である。すなわち、人混みや雑踏の中でも自分に関係のある会話や興味のあるキーワードを自然に聞きとることができるカクテルパーティ効果をコンピュータを用いて実現することである。音源分離は、Figs. 1.1 及び 1.2 に示すように、コンピュータを用いて音楽信号から楽譜を自動生成する自動採譜や、コンサートのような複数の楽器が同時に演奏されている音楽信号から特定の楽器のメロディのみを抽出する等の活用方法がある。また音声認識へも応用されており、ノイズを抑制し音声の特徴のみを捉えるために音源分離が活用されている。

### 1.2 本論における主題

音源分離は、予め用意しておく音源分離達成のための手掛かりとなるような情報（これを教師情報と呼ぶ）を用いない「教師なし音源分離」と、教師情報を用いる「教師あり音源分離」の 2 つに大別できる。教師なし音源分離は、教師情報を必要としないことから、あらゆる音響信号への活用が可能である [1]。その反面、高い音源分離精度の実現は困難である。一方、教師あり音源分離では、混合信号中の各音源の種類や楽譜、声色のサンプル等の教師情報を事前に学習し、学習で得られる何らかのモデルを音源分離に活用する。

教師あり音源分離は、取り扱う教師情報の数によってさらに「全教師あり音源分離」と「半教師あり音源分離」に分けられる [2, 3]。音源分離には、全ての音源を個々に分離したい場合と、特定の音だけを全体の中から抽出したい場合の二種類がある。教師あり音源分離を適用する場合、前者では教師情報は全ての音源分必要となる。このような教師あり音源分離を全教師あり音源分離と呼ぶ。また後者では、抽出したい音（目的音）だけの教師情報が必要となる。よって半教師あり音源分離では全教師あり音源分離よりも予め学習する教師情報の数が少ない。すなわち学習コストが低いといえる。

また、教師情報が大量にある場合、深層ニューラルネットワーク (deep neural network:

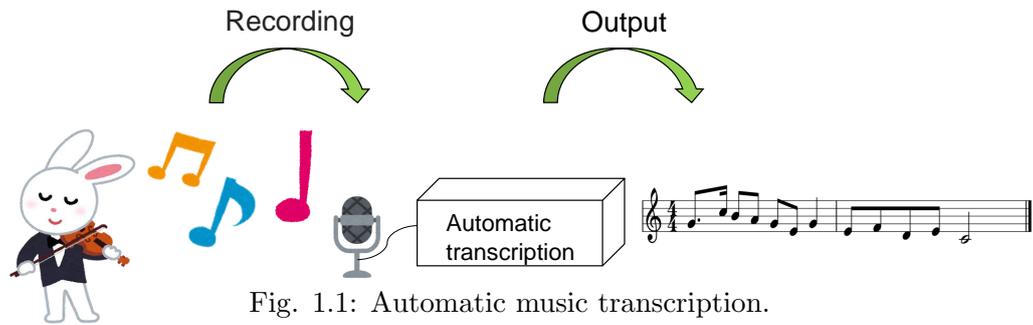


Fig. 1.1: Automatic music transcription.

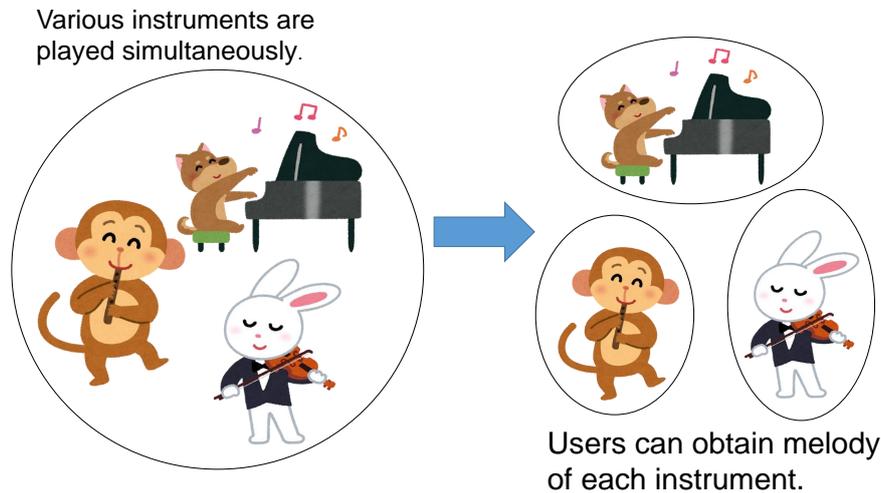


Fig. 1.2: Separation of instrumental sounds.

DNN) と呼ばれる機械学習モデルが有用である [4, 5]. しかし, DNN を学習するためには, 各音源のソロの録音信号を数百時間分用意する必要がある等の現実的な問題がある. 教師なしの状況や, 分離対象となる音源 (目的音源) の学習用データセットが極僅かしか用意できないような状況においては, 非負値行列因子分解 (nonnegative matrix factorization: NMF) [6, 7] に基づく手法がこれまで数多く提案されてきた [1]–[3], [8]–[10]. 特に, 省サンプルの学習データから特徴量を学習しそのまま活用する教師あり NMF (supervised NMF: SNMF) [2, 3] は, 現実的な学習コストと音源分離性能を両立しており, 依然として強力である.

本論文では, SNMF に基づく音源分離のみを取り扱う. 既存手法である直交化罰則項を付与した罰則条件付き SNMF (penalized SNMF: PSNMF) [3] では, NMF におけるスケール不定性により罰則条件が正しく影響しない問題がある. 本論文では, この問題を解決するとともに, 更なる分離精度の改善のためにコサイン類似度を新たな罰則項として用いる PSNMF を提案する. Fig. 1.3 では従来手法, 提案手法及び DNN に基づく音源分離の位置付けを示す. 本論文で提案する手法では, DNN ほどの分離精度は見込めないが, 学習コストは従来手法同様低コストである. さらに, 分離精度は従来の PSNMF より高くなることを実験で示す.

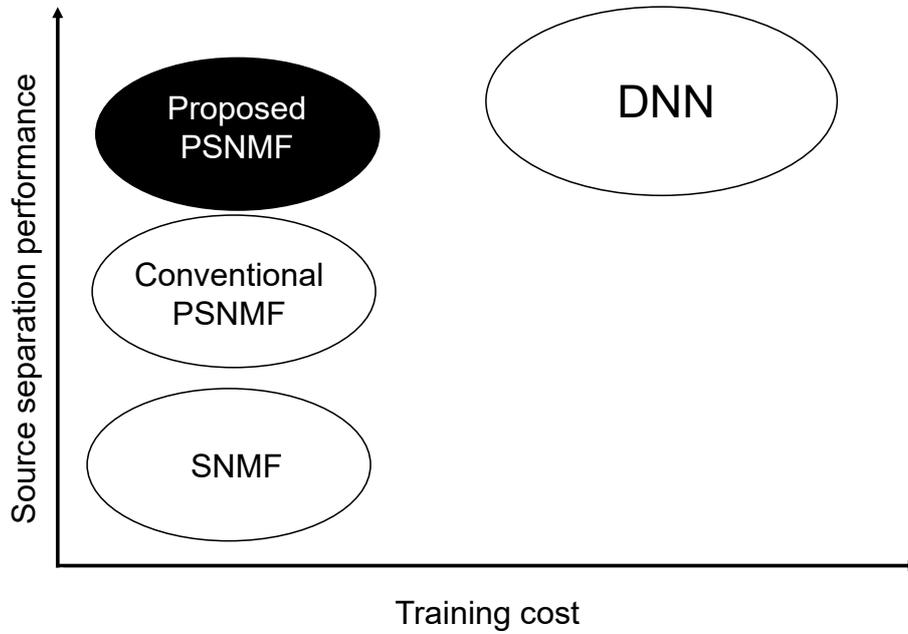


Fig. 1.3: Scope of this thesis.

### 1.3 本論文の構成

本論文の構成は次の通りである．2章では，NMFの基礎理論について述べ，音源分離への適用方法を示す．また，既存のPSNMFの概要の反復更新式を示す．3章では本論文で新たに提案するコサイン類似度罰則付きPSNMFについて説明し，提案手法の反復更新式を導出する．4章では，提案手法の有効性を確認するために，音楽信号の音源分離性能を罰則項なしのシンプルSNMF，従来の直交化罰則付きPSNMF及び提案手法に関して比較し，その結果を述べる．5章にて，本論文のまとめを述べる．

## 第 2 章

# NMF を用いた音源分離

### 2.1 NMF

NMF とは非負行列を次元圧縮することで少数の有用な特徴量を抽出するアルゴリズムである。実世界ではパワースペクトルや画素数といったような非負値のデータが多い。NMF は、そのような非負値のデータを加法的な構成成分に分解することを目的として用いられており、その解釈の容易さから商品の購買データや音源分離といったパターン認識を活用する分野で有用である [3, 11]。パターン認識において NMF は非負の行列データを低ランクの非負行列の積に分解することで、非負の行列データ中の潜在的なパターンを抽出することが可能である。

NMF は次式に示す行列の因子分解を行う。

$$\mathbf{X} \approx \mathbf{F}\mathbf{Q} \quad (2.1)$$

ここで、 $\mathbf{X} \in \mathbb{R}_{\geq 0}^{I \times J}$  は非負観測行列である。また、 $\mathbf{F} \in \mathbb{R}_{\geq 0}^{I \times K}$  及び  $\mathbf{Q} \in \mathbb{R}_{\geq 0}^{K \times J}$  はそれぞれ基底行列及びアクティベーション行列である。 $I$  及び  $J$  は観測行列  $\mathbf{X}$  の行数と列数である。また  $K$  は基底数である。このように、NMF は非負のデータ行列（観測行列） $\mathbf{X}$  を別の二つの非負行列の行列積  $\mathbf{F}\mathbf{Q}$  で近似するアルゴリズムである。基底行列  $\mathbf{F}$  には、観測行列  $\mathbf{X}$  中の頻出パターンが基底ベクトルとして含まれ、それらの強度変化がアクティベーション行列  $\mathbf{Q}$  に含まれる。

式 (3.5) を要素ごとに表すと、

$$x_{ij} \approx \sum_k f_{ik} q_{kj} \quad (2.2)$$

となる。 $x_{ij}$ ,  $f_{ik}$  及び  $q_{kj}$  はそれぞれ  $\mathbf{X}$ ,  $\mathbf{F}$ , 及び  $\mathbf{Q}$  の要素である。さらに、 $i = 1, \dots, I$ ,  $j = 1, \dots, J$ , 及び  $k = 1, \dots, K$  はそれぞれ  $I$ ,  $J$ , 及び  $K$  のインデックスを表す。一般に基底数  $K$  は  $I$  や  $J$  よりも十分に小さい値に設定されるため、NMF は観測行列  $\mathbf{X}$  を  $K$  個の基底で表現する低ランク近似と解釈できる。NMF において  $\mathbf{F}$  及び  $\mathbf{Q}$  は次式の最適化問題で推定される。

$$\min_{\mathbf{F}, \mathbf{Q}} \mathcal{D}(\mathbf{X} \parallel \mathbf{F}\mathbf{Q}) \text{ s.t. } f_{ik}, q_{kj} \geq 0 \forall i, j, k \quad (2.3)$$

ここで、 $\mathcal{D}(\cdot\|\cdot)$  は二つの行列の類似度を測る任意の関数であり、NMF の文脈では二乗ユークリッド距離 [7]，一般化 Kulback–Leibler (KL) ダイバージェンス [7]，Itakura–Saito (IS) ダイバージェンス [12] 等がよく用いられる．従って式 (2.3) は，非負観測行列  $\mathbf{X}$  を良く近似するような非負のモデル行列  $\mathbf{FQ}$ （但しモデル行列  $\mathbf{FQ}$  のランクは  $K$ ）を推定する問題となる．

上記の 3 種類の類似度関数はそれぞれ次のように定義されている．

- 二乗ユークリッド距離

$$\mathcal{D}(\mathbf{X}\|\mathbf{FQ}) = \sum_{i,j} \left( x_{ij} - \sum_k f_{ik}q_{kj} \right)^2 \quad (2.4)$$

- KL ダイバージェンス

$$\mathcal{D}(\mathbf{X}\|\mathbf{FQ}) = \sum_{i,j} \left( x_{ij} \log \frac{x_{ij}}{\sum_k f_{ik}q_{kj}} - x_{ij} + \sum_k f_{ik}q_{kj} \right) \quad (2.5)$$

- IS ダイバージェンス

$$\mathcal{D}(\mathbf{X}\|\mathbf{FQ}) = \sum_{i,j} \left( \frac{x_{ij}}{\sum_k f_{ik}q_{kj}} - \log \frac{x_{ij}}{\sum_k f_{ik}q_{kj}} - 1 \right) \quad (2.6)$$

式 (2.3) の  $\mathbf{F}$  と  $\mathbf{Q}$  に関する解は閉形式では与えられないため， $\mathbf{F}$  と  $\mathbf{Q}$  を何らかの方法で初期化 (例えば [9] 等) したうえで，反復計算を行うことで式 (2.3) の類似度関数を小さくするような  $\mathbf{F}$  と  $\mathbf{Q}$  を求める．式 (2.4)–(2.6) の類似度関数を用いた NMF の，majorization-minimization (MM) アルゴリズム [13] に基づく反復更新式を以下に記す．

- 二乗ユークリッド距離

$$f_{ik} \leftarrow f_{ik} \cdot \frac{\sum_j x_{ij}q_{kj}}{\sum_j q_{kj} \sum_{k'} f_{ik'}q_{k'j}} \quad (2.7)$$

$$q_{kj} \leftarrow q_{kj} \cdot \frac{\sum_i x_{ij}f_{ik}}{\sum_i f_{ik} \sum_{k'} f_{ik'}q_{k'j}} \quad (2.8)$$

- KL ダイバージェンス

$$f_{ik} \leftarrow f_{ik} \cdot \frac{\sum_j \frac{x_{ij}}{\sum_{k'} f_{ik'}q_{k'j}} q_{kj}}{\sum_j q_{kj}} \quad (2.9)$$

$$q_{kj} \leftarrow q_{kj} \cdot \frac{\sum_i \frac{x_{ij}}{\sum_{k'} f_{ik'}q_{k'j}} f_{ik}}{\sum_i f_{ik}} \quad (2.10)$$

- IS ダイバージェンス

$$f_{ik} \leftarrow f_{ik} \cdot \sqrt{\frac{\sum_j \frac{x_{ij}}{\left(\sum_{k'} f_{ik'} q_{k'j}\right)^2} q_{kj}}{\sum_j \frac{q_{kj}}{\sum_{k'} f_{ik'} q_{k'j}}}} \quad (2.11)$$

$$q_{kj} \leftarrow q_{kj} \cdot \sqrt{\frac{\sum_i \frac{x_{ij}}{\left(\sum_{k'} f_{ik'} q_{k'j}\right)^2} f_{ik}}{\sum_i \frac{f_{ik}}{\sum_{k'} f_{ik'} q_{k'j}}}} \quad (2.12)$$

なお、反復一回目の  $f_{ik}$  及び  $q_{kj}$  の初期値には正の乱数を与える。これは、NMF が非負値行列を分離対象としているとともに、初期値に 0 を与えるとその要素は全く更新されなくなるからである。

## 2.2 短時間フーリエ変換

NMF を音源分離に適用するためには、短時間フーリエ変換 (short-time Fourier transform: STFT) により音を 2 次元表現する必要がある。STFT とは、一次元の時間信号を二次元の時間周波数信号に変換することである。STFT の処理の概要を Fig. 2.1 に示す。与えられた時間波形を短い時間波形 (フーリエ変換長) に分割し、窓関数を乗ずることで離散フーリエ変換を適用する。この処理を任意の時間ごとに行うことにより、時間と周波数の情報を持つ二次元信号が得られ、周波数分布 (スペクトル) の時間的な変化の表現を実現している。この二次元信号はスペクトログラムと呼ばれ、複素数要素を持つ行列として表される。

## 2.3 NMF の音響信号への適用

音響信号に NMF を適用する場合、時間信号に STFT を適用して得られるスペクトログラム (複素行列) の振幅値 (振幅スペクトログラム) やパワー値 (パワースペクトログラム) を非負値観測行列  $\mathbf{X} \in \mathbb{R}_{\geq 0}^{I \times L}$  として扱うことが一般的である。このとき、 $I$  は周波数ビン数、 $J$  は時間フレーム数にそれぞれ対応する。本論文では、常に  $\mathbf{X}$  を振幅スペクトログラムとして定義する。

振幅スペクトログラム  $\mathbf{X}$  を NMF で分解した様子が Fig. 2.2 である。観測行列  $\mathbf{X}$  は NMF を用いることで基底行列  $\mathbf{F}$  とアクティベーション行列  $\mathbf{Q}$  の行列積で近似されている。このとき  $\mathbf{X}$  は非負の時間周波数行列 (振幅スペクトログラム) であり、 $i = 1, \dots, I$ ,  $j = 1, \dots, J$ , 及び  $k = 1, \dots, K$  はそれぞれ周波数ビン, 時間フレーム, 及び基底行列のインデックスを表

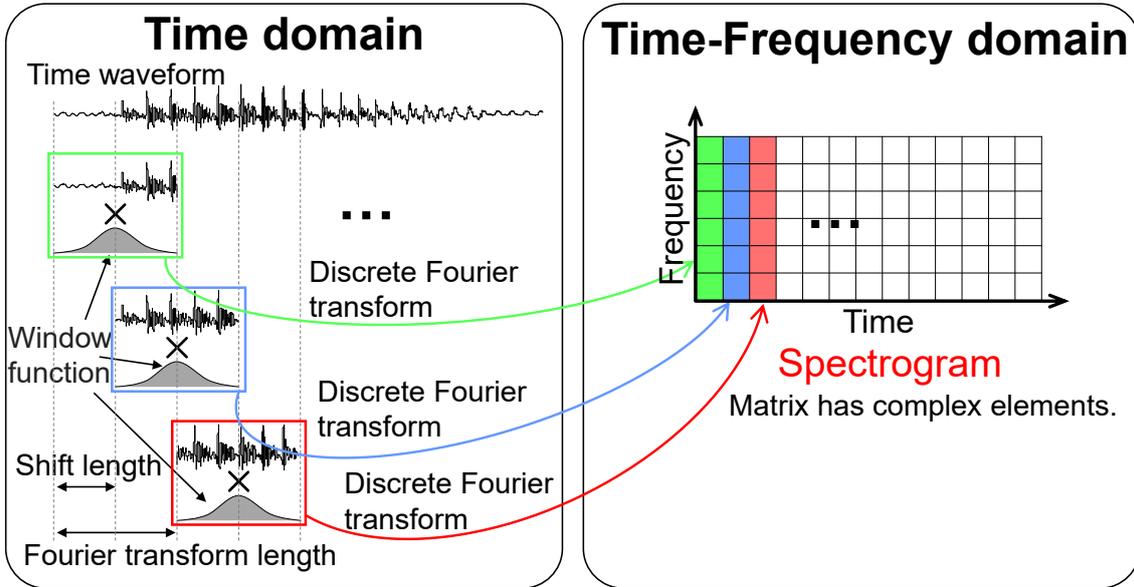


Fig. 2.1: Process of STFT.

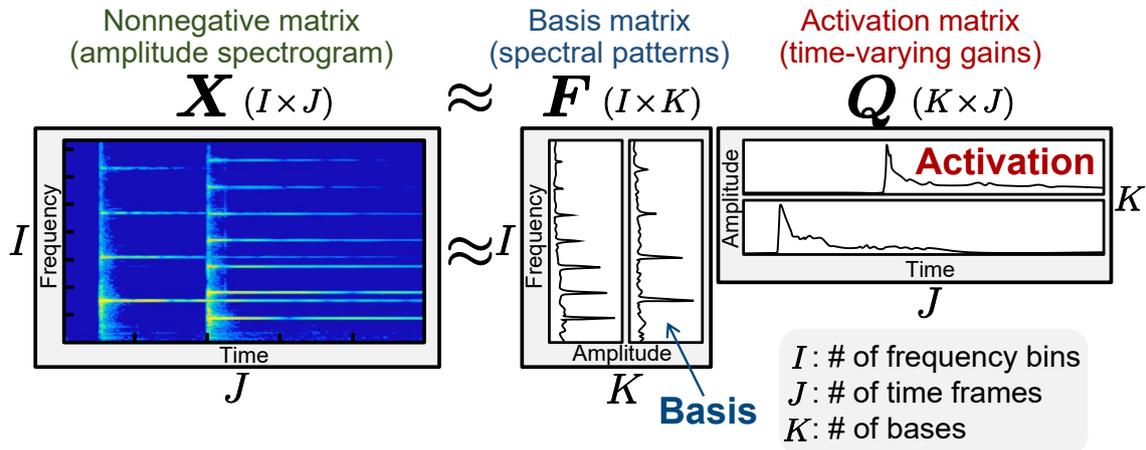


Fig. 2.2: Matrix decomposition by NMF.

す. Fig. 2.2 において基底行列  $F$  は 2 つの基底ベクトルからなっている. すなわち, 基底行列  $F$  では観測行列  $X$  のスペクトルパターンを 2 本の基底ベクトルで近似している. 同様に, アクティベーション行列  $Q$  は  $F$  中の 2 本の基底ベクトルがどの時間にどの程度の強さで生起するかを表す時間的強度変化を行ベクトルとして含んでいる. 従って, Fig. 2.2 では  $K = 2$  として, 2 本の基底ベクトルだけで観測行列  $X$  を表している. 本論文では, 1 章で述べた通り, SNMF に基づく音源分離手法の現実的な応用可能性が高いことを考慮し, 次節で説明する SNMF に主眼を置き, その改良方法について検討する.

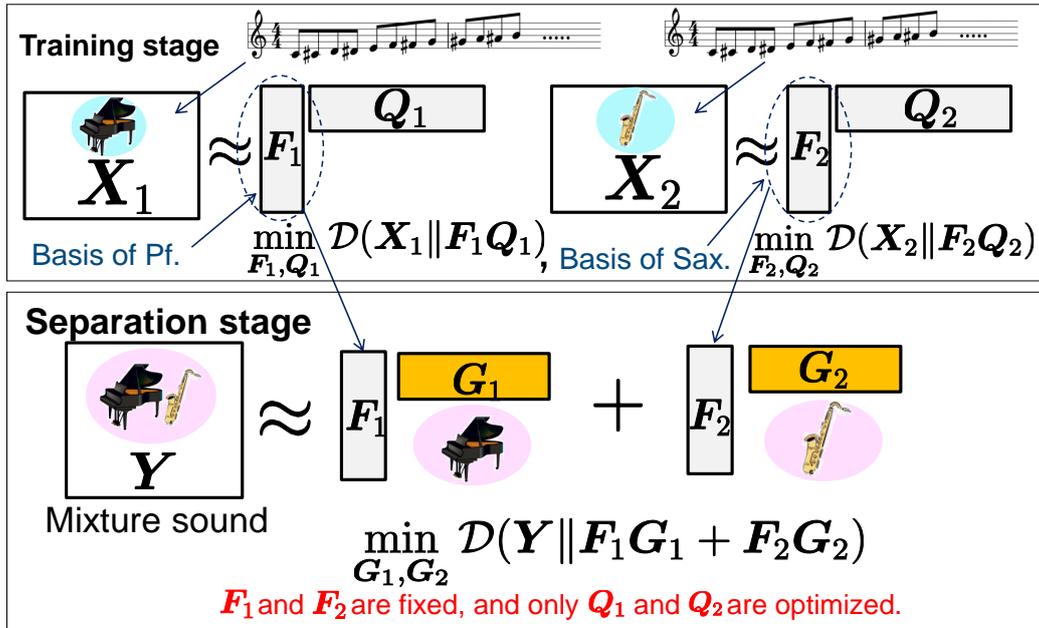


Fig. 2.3: Separation algorithm in full-SNMF.

## 2.4 SNMF による音源分離

NMF を用いた音源分離において、分離対象となる音源（目的音源）の学習用データセットが極僅かしか用意できないような状況においては、省サンプルの学習データから特徴量を学習しそのまま活用する SNMF [2, 3] は比較的高性能な分離精度である。SNMF には全教師ありの SNMF と半教師ありの SNMF がある。SNMF は、学習ステージと分離ステージから構成されている。Fig. 2.3 は全教師ありの SNMF の音源分離アルゴリズムを示している。学習ステージでは、全ての音源の教師情報（ここでは各音源のサンプル信号）の振幅スペクトログラム  $X_1$  及び  $X_2$  を NMF でそれぞれ分解し、各音源の頻出スペクトルパターンを含む教師基底行列  $F_1$  及び  $F_2$  を学習する。分離ステージでは、事前学習した教師基底行列  $F_1$  及び  $F_2$  を用いて、混合信号の振幅スペクトログラム  $Y$  を分解することで、 $F_1 G_1$  及び  $F_2 G_2$  には、それぞれ事前に学習した各音源の成分が現れ、音源が分離される。Fig. 2.4 は半教師ありの SNMF の音源分離アルゴリズムを示している。学習ステージでは、目的音源のサンプル信号の振幅スペクトログラム  $X$  を NMF で分解し、頻出スペクトルパターンを含む教師基底行列  $F$  を学習する。分離ステージでは、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  $Y$  を分解することで、目的音源成分  $FG$  とその他の音源（非目的音源）成分  $HU$  を推定する。本論文では、半教師ありの SNMF のみを議論の対象とするが、提案するアイデア等は一般性を失うことなく全教師ありの SNMF にも応用可能である。以降は半教師ありの SNMF のことを単に「SNMF」と呼ぶ。

SNMF を用いた音源分離では、振幅スペクトルを KL ダイバージェンスに基づく NMF で

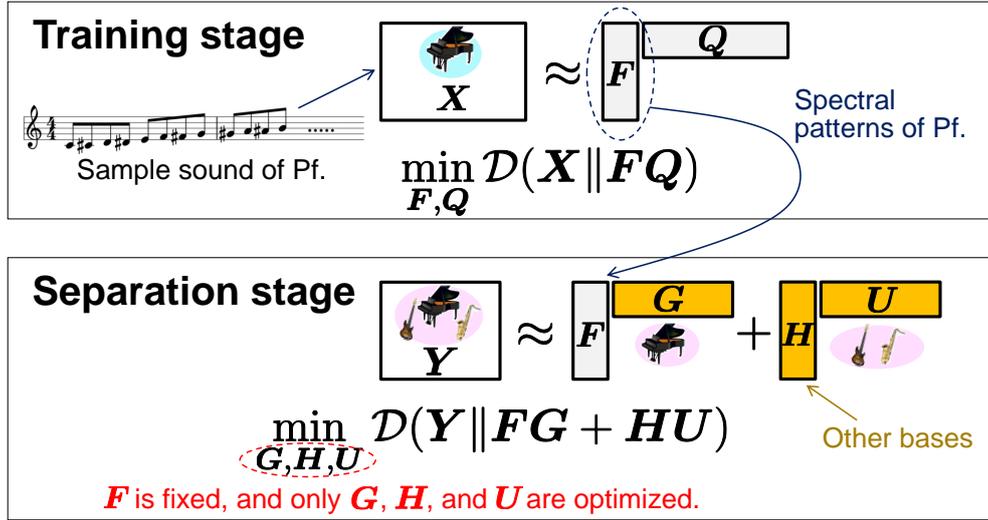


Fig. 2.4: Separation algorithm in semi-SNMF.

分解する手法が一番性能が良いことが実験的に確認されている [3, 14]. よって、一番性能の良い手法を改良するのが妥当であるため、本研究では式 (2.5) に示す KL ダイバージェンスのみを用いる. SNMF では、学習ステージにおいて目的音源の音階を並べた音等のサンプル信号の振幅スペクトログラム  $X$  に NMF を適用することで、教師基底行列  $F$  を事前学習する. このとき、 $F$  は目的音源の頻出スペクトルパターンを  $K$  本の列ベクトル (基底ベクトル) として含んでいる. 分離ステージでは、教師基底行列  $F$  が与えられた下で、混合信号の振幅スペクトログラム  $Y \in \mathbb{R}_{\geq 0}^{I \times J}$  を次の最適化問題により分解する.

$$\begin{aligned} & \min_{G,H,U} \mathcal{D}(Y || FG + HU) \\ & \text{s.t. } g_{kj}, h_{il}, u_{lj} \geq 0 \quad \forall i, j, k, l \end{aligned} \quad (2.13)$$

ここで、 $H \in \mathbb{R}_{\geq 0}^{I \times L}$  及び  $U \in \mathbb{R}_{\geq 0}^{L \times J}$  はそれぞれ非目的音源を表すための基底行列とアクティベーション行列であり、 $h_{il}$  及び  $u_{lj}$  はそれぞれ  $H$  及び  $U$  の要素である. また、 $l = 1, \dots, L$  は非目的音源の基底行列  $H$  に含まれる基底のインデックスを表す. 理想的には、混合信号  $Y$  中の目的音源の成分が  $FG$ 、非目的音源の成分が  $HU$  として分離されるため、音源分離が可能となる.

MM アルゴリズム [13] に基づく変数行列  $G$ ,  $H$ , 及び  $U$  の更新則は式 (2.14)~式 (2.16) のようになる.

$$g_{kj} \leftarrow g_{kj} \cdot \frac{\sum_i \frac{y_{ij}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} h_{il'} u_{l'j}} f_{ik}}{\sum_i f_{ik}} \quad (2.14)$$

$$h_{il} \leftarrow h_{il} \cdot \frac{\sum_j \frac{y_{ij}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} h_{il'} u_{l'j}} u_{lj}}{\sum_j u_{lj}} \quad (2.15)$$

$$u_{lj} \leftarrow u_{lj} \cdot \frac{\sum_i \frac{y_{ij}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} h_{il'} u_{l'j}} h_{il}}{\sum_i h_{il}} \quad (2.16)$$

ここで、 $y_{ij}$  は  $\mathbf{Y}$  の要素である。

SNMF に基づく音源分離では、混合信号中の目的音源と非目的音源が類似したスペクトルを含む場合、分離精度が劣化する問題がある。これは、類似するスペクトルが教師基底行列  $\mathbf{F}$  又は非目的音源の基底行列  $\mathbf{H}$  のいずれを用いても表現できることが原因である。これは、式 (2.13) で示す最適化問題において、コスト関数が  $\mathbf{X}$  と  $\mathbf{FG} + \mathbf{HU}$  間の距離のみで定義されており、このコスト関数には基底行列同士が類似しないような制約が含まれていることが原因である。この場合、目的音源の一部が  $\mathbf{HU}$  に、あるいは非目的音源の一部が  $\mathbf{FG}$  に混入することとなり、音源分離精度が劣化する。

この問題に対処するために、直交化に基づく PSNMF [3] が提案されている。PSNMF は、教師基底行列  $\mathbf{F}$  と非目的音源の基底行列  $\mathbf{H}$  ができるだけ直交化する罰則条件を付与した SNMF であり、目的音源成分  $\mathbf{FG}$  の品質を大幅に向上させる効果がある。本手法では、分離ステージにおいて次の最適化問題を考える。

$$\begin{aligned} \min_{\mathbf{G}, \mathbf{H}, \mathbf{U}} \mathcal{D}(\mathbf{Y} \| \mathbf{FG} + \mathbf{HU}) + \mu \mathcal{P}_{\text{orth}}(\mathbf{F}, \mathbf{H}) \\ \text{s.t. } g_{kj}, h_{il}, u_{lj} \geq 0 \quad \forall i, j, k, l \end{aligned} \quad (2.17)$$

$$\begin{aligned} \mathcal{P}_{\text{orth}}(\mathbf{F}, \mathbf{H}) &= \|\mathbf{F}^T \mathbf{H}\|_{\text{Fr}}^2 \\ &= \sum_{k,l} \left( \sum_i f_{ik} h_{il} \right)^2 \end{aligned} \quad (2.18)$$

ここで、 $\mu$  は罰則項  $\mathcal{P}_{\text{orth}}(\mathbf{F}, \mathbf{H})$  の重み係数である。また、 $\|\mathbf{A}\|_{\text{Fr}}$  は任意の行列  $\mathbf{A} \in \mathbb{R}_{\geq 0}^{I \times J}$  のフロベニウスノルムであり、次式で表される。

$$\|\mathbf{A}\|^2 = \sqrt{\sum_{i,j} a_{ij}^2} \quad (2.19)$$

ここで、 $a_{ij}$  は  $\mathbf{A}$  の要素である。罰則項  $\mathcal{P}_{\text{orth}}(\mathbf{F}, \mathbf{H})$  は、 $\mathbf{F}$  の基底ベクトル ( $K$  本) と  $\mathbf{H}$  の基底ベクトル ( $L$  本) の総組み合わせの内積の二乗和に対応する。従って、 $\mathbf{H}$  を推定する際に

は、「 $\mathbf{Y}$  と  $\mathbf{FG} + \mathbf{HU}$  が近くなること」及び「 $\mathbf{H}$  の基底ベクトルが  $\mathbf{F}$  の基底ベクトルとなるべく直交すること」の2点が考慮され、結果として SNMF よりも音源分離が促進されることが実験的に確認されている [3]. MM アルゴリズム [13] に基づく変数行列  $\mathbf{G}$ ,  $\mathbf{H}$ , 及び  $\mathbf{U}$  の更新則はそれぞれ次のようになる [3].

$$g_{kj} \leftarrow g_{kj} \cdot \frac{\sum_i \frac{y_{ij}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} h_{il'} u_{l'j}} f_{ik}}{\sum_i f_{ik}} \quad (2.20)$$

$$h_{il} \leftarrow h_{il} \cdot \frac{\sum_j \frac{y_{ij}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} h_{il'} u_{l'j}} u_{lj}}{\sum_j u_{lj} + \mu f_{ik} \sum_{i'} f_{i'k} h_{i'l}} \quad (2.21)$$

$$u_{lj} \leftarrow u_{lj} \cdot \frac{\sum_i \frac{y_{ij}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} h_{il'} u_{l'j}} h_{il}}{\sum_i h_{il}} \quad (2.22)$$

## 第 3 章

# コサイン類似度に基づく PSNMF

### 3.1 動機

直交化に基づく PSNMF には、罰則項の効果が直接反映されない本質的な問題がある。これは、NMF が本来持っている基底行列・アクティベーション行列間のスケール不定性が原因である。即ち、 $\mathbf{X} \approx \mathbf{FQ}$  という NMF 分解を考えた際に、基底行列  $\mathbf{F}$  に任意の正の係数  $\alpha$  を乗じて  $\alpha\mathbf{F}$  と変形しても、同時にアクティベーション行列  $\mathbf{Q}$  を  $(1/\alpha)\mathbf{Q}$  とすれば、 $\mathbf{FQ}$  自体の値は変わらないという不定性である。

直交化に基づく PSNMF では、罰則項  $\mathcal{P}_{\text{orth}}(\mathbf{F}, \mathbf{H})$  の値を小さくすることで  $\mathbf{F}$  と  $\mathbf{H}$  の非類似性を高めることを考えるが、実際にはこの罰則項は非常に小さな値の  $\alpha$  を乗じて  $\alpha\mathbf{H}$  とすることで最小化可能である。このとき、 $\mathbf{HU}$  の値が変わらないように、 $\mathbf{U}$  は  $(1/\alpha)\mathbf{U}$  と更新される。従って、この罰則項は直接的に  $\mathbf{F}$  と  $\mathbf{H}$  を直交化しておらず、式 (2.17) の最適化問題は、理論的には、式 (2.13) と同一となる。

この問題を避けるために、文献 [3] では、 $\mathbf{H}$  を更新する度に全基底ベクトルを正規化する（正規化係数の逆数を  $\mathbf{U}$  の行ベクトルに乗じて  $\mathbf{HU}$  は不変とする）ようにプログラムを実装しており、実験的には性能改善が見られたが、この実装方法も根本的な解決とは言えない。同様の現象は、 $L_1$  ノルムに基づくスパース正則化をアクティベーション行列に与える NMF でも確認されており、基底ベクトルの長さを制約条件に与える最適化法が提案されている [15]。PSNMF においても、文献 [15] と同様の制約条件を導入して解くことはできるが、よりシンプルな解決法として、基底ベクトル間の大きさには依存せず、角度にのみ依存するコサイン類似度を新たに罰則項に用いる PSNMF を提案する。

### 3.2 コスト関数

コサイン類似度に基づく PSNMF として、次の最適化問題を考える。

$$\begin{aligned} \min_{\mathbf{G}, \mathbf{H}, \mathbf{U}} \mathcal{D}(\mathbf{Y} \| \mathbf{FG} + \mathbf{HU}) + \mu \mathcal{P}_{\text{cos}}(\mathbf{F}, \mathbf{H}) \\ \text{s.t. } g_{kj}, h_{il}, u_{lj} \geq 0 \quad \forall i, j, k, l \end{aligned} \quad (3.1)$$

$$\mathcal{P}_{\cos}(\mathbf{F}, \mathbf{H}) = \sum_{k,l} \log \frac{\sum_i f_{ik} h_{il}}{\left(\sum_i f_{ik}^2\right)^{\frac{1}{2}} \left(\sum_i h_{il}^2\right)^{\frac{1}{2}}} \quad (3.2)$$

ここで、罰則項  $\mathcal{P}_{\cos}(\mathbf{F}, \mathbf{H})$  は、 $\mathbf{F}$  の基底ベクトルと  $\mathbf{H}$  の基底ベクトルの総組み合わせのコサイン類似度の対数和に対応する。コサイン類似度であるので、基底ベクトルの長さには依らず、角度のみで直交性を測ることができ、NMFにおけるスケールの不定性は影響しない。また、基底ベクトルは全て非負であることから、コサイン類似度の値は最小で0（直交）、最大で1（平行）となる。ここで、単純な「コサイン類似度の和」による定式化では、MM アルゴリズムでの導出は困難であったが、「コサイン類似度の対数和」として罰則項を設計することで、次節に述べる MM アルゴリズムでの導出が可能となる。

### 3.3 反復更新式の導出

式 (3.1) は次のように変形できる。

$$\begin{aligned} & \mathcal{D}(\mathbf{Y} \|\mathbf{F}\mathbf{G} + \mathbf{H}\mathbf{U}) + \mu \mathcal{P}_{\cos}(\mathbf{F}, \mathbf{H}) \\ &= \sum_{i,j} \left[ y_{ij} \log y_{ij} - y_{ij} \log \left( \sum_k f_{ik} g_{kj} + \sum_l h_{il} u_{lj} \right) - y_{ij} + \sum_k f_{ik} g_{kj} + \sum_l h_{il} u_{lj} \right] \\ & \quad + \sum_{k,l} \log \frac{\sum_i f_{ik} h_{il}}{\left(\sum_i f_{ik}^2\right)^{\frac{1}{2}} \left(\sum_i h_{il}^2\right)^{\frac{1}{2}}} \end{aligned} \quad (3.3)$$

$$\begin{aligned} &= \sum_{i,j} \left[ y_{ij} \log y_{ij} - y_{ij} \log \left( \sum_k f_{ik} g_{kj} + \sum_l h_{il} u_{lj} \right) - y_{ij} + \sum_k f_{ik} g_{kj} + \sum_l h_{il} u_{lj} \right] \\ & \quad + \mu \sum_{k,l} \left[ \log \sum_i f_{ik} h_{il} - \frac{1}{2} \log \sum_i f_{ik}^2 - \frac{1}{2} \log \sum_i h_{il}^2 \right] \end{aligned} \quad (3.4)$$

式 (3.4) のコスト関数を最小化する変数  $g_{kj}$ ,  $h_{il}$  及び  $u_{lj}$  を求めることが目的であるが、この関数を各変数で偏微分することは困難である。その理由は、第二項、第六項及び第八項が対数の内部に変数の総和を含んでいるからである。このようにコスト関数の微分を解析的に求めることが困難である場合、解析的に微分が求められる上限関数（コスト関数と1点でのみ接し、それ以外の点ではコスト関数を上回る関数）を設計して、コスト関数の代わりに上限関数の最小点を求めることを繰り返すことで、コスト関数の最小点を求める最適化手法が MM アルゴリズム [13] である。本論文においても、従来の NMF の最適化手法と同様に、MM アルゴリズムに基づく最適化アルゴリズム（反復更新式）を導出する。

まず、コスト関数 (3.4) を  $\mathcal{J}$  とし、次のように変形する。

$$\begin{aligned} \mathcal{J} = \sum_{i,j} \left[ y_{ij} \log y_{ij} - y_{ij} \log \left( \gamma_{ij1} \frac{\sum_k f_{ik} g_{kj}}{\gamma_{ij1}} + \gamma_{ij2} \frac{\sum_l h_{il} u_{lj}}{\gamma_{ij2}} \right) - y_{ij} + \sum_k f_{ik} g_{kj} \right. \\ \left. + \sum_l h_{il} u_{lj} \right] + \mu \sum_{k,l} \left[ \log \sum_i f_{ik} h_{il} - \frac{1}{2} \log \sum_i f_{ik}^2 - \frac{1}{2} \log \sum_i h_{il}^2 \right] \end{aligned} \quad (3.5)$$

ここで、 $\gamma_{ij1} > 0$  及び  $\gamma_{ij2} > 0$  は  $\gamma_{ij1} + \gamma_{ij2} = 1$  を満たす補助変数である。式 (3.4) の第二項は変数 ( $g_{kj}$ ,  $h_{il}$  及び  $u_{lj}$ ) の総和を含む負対数関数となっており、上限関数を設計する必要がある。負対数関数 ( $-\log$ ) は、変数に対して凸関数であるため、次式のようにイェンセンの不等式 (A.1) を適用することで上限関数を設計できる。

$$\begin{aligned} & -\log \left( \gamma_{ij1} \frac{\sum_k f_{ik} g_{kj}}{\gamma_{ij1}} + \gamma_{ij2} \frac{\sum_l h_{il} u_{lj}}{\gamma_{ij2}} \right) \\ & \leq -\gamma_{ij1} \log \frac{\sum_k f_{ik} g_{kj}}{\gamma_{ij1}} - \gamma_{ij2} \log \frac{\sum_l h_{il} u_{lj}}{\gamma_{ij2}} \\ & = -\gamma_{ij1} \log \sum_k f_{ik} g_{kj} + \gamma_{ij1} \log \gamma_{ij1} - \gamma_{ij2} \log \sum_l h_{il} u_{lj} + \gamma_{ij2} \log \gamma_{ij2} \\ & = -\gamma_{ij1} \log \sum_k \alpha_{ijk} \frac{f_{ik} g_{kj}}{\alpha_{ijk}} + \gamma_{ij1} \log \gamma_{ij1} - \gamma_{ij2} \log \sum_l \beta_{ijl} \frac{h_{il} u_{lj}}{\beta_{ijl}} + \gamma_{ij2} \log \gamma_{ij2} \end{aligned} \quad (3.6)$$

ここで  $\alpha_{ijk} > 0$  及び  $\beta_{ijl} > 0$  は  $\sum_k \alpha_{ijk} = 1$  及び  $\sum_l \beta_{ijl} = 1$  を満たす補助変数である。式 (3.6) の等号成立条件は次式で与えられる。

$$\gamma_{ij1} = \frac{\sum_k f_{ik} g_{kj}}{\sum_k f_{ik} g_{kj} + \sum_l h_{il} u_{lj}} \quad (3.7)$$

$$\gamma_{ij2} = \frac{\sum_l h_{il} u_{lj}}{\sum_k f_{ik} g_{kj} + \sum_l h_{il} u_{lj}} \quad (3.8)$$

式 (3.6) の第一項及び第三項は、依然として変数 ( $g_{kj}$ ,  $h_{il}$  及び  $u_{lj}$ ) の総和を含む負対数関数となっているため、次のように再度イェンセンの不等式を適用する。

$$-\gamma_{ij1} \log \sum_k \alpha_{ijk} \frac{f_{ik} g_{kj}}{\alpha_{ijk}} \leq -\gamma_{ij1} \sum_k \alpha_{ijk} \log \frac{f_{ik} g_{kj}}{\alpha_{ijk}} \quad (3.9)$$

$$-\gamma_{ij2} \log \sum_l \beta_{ijl} \frac{h_{il} u_{lj}}{\beta_{ijl}} \leq -\gamma_{ij2} \sum_l \beta_{ijl} \log \frac{h_{il} u_{lj}}{\beta_{ijl}} \quad (3.10)$$

式 (3.9) 及び (3.10) の等号成立条件はそれぞれ次式で与えられる.

$$\alpha_{ijk} = \frac{f_{ik}g_{kj}}{\sum_{k'} f_{ik'}g_{k'j}} \quad (3.11)$$

$$\beta_{ijl} = \frac{h_{il}u_{lj}}{\sum_{l'} h_{il'}u_{l'j}} \quad (3.12)$$

また, 式 (3.4) の第六項も変数 ( $h_{il}$ ) の総和を含む正対数関数となっており, 上限関数を設計する必要がある. 正対数関数 ( $+\log$ ) は, 変数に対して凹関数であるため, 次式のように接線不等式 (A.2) を適用することで上限関数を設計できる.

$$\log \sum_i f_{ik}g_{kj} \leq \frac{1}{\zeta_{kli}} \left( \sum_i f_{ik}g_{kj} - \zeta_{kli} \right) + \log \zeta_{kli} \quad (3.13)$$

ここで,  $\zeta_{kli} > 0$  は  $\sum_i \zeta_{kli} = 1$  を満たす補助変数である. 式 (3.13) の等号成立条件は次式で与えられる.

$$\zeta_{kli} = \sum_i f_{ik}h_{il} \quad (3.14)$$

さらに, 式 (3.4) の第八項も変数 ( $h_{il}$ ) の総和を含む負対数関数となっており, 上限関数を設計する必要がある. 次式のようにイェンセンの不等式を適用することで上限関数を設計できる.

$$\begin{aligned} -\log \sum_i h_{il}^2 &= -\log \sum_i \delta_{kli} \frac{h_{il}^2}{\delta_{kli}} \\ &\leq -\sum_i \delta_{kli} \log \frac{h_{il}^2}{\delta_{kli}} \\ &= -\sum_i \delta_{kli} (2 \log h_{il} - \log \delta_{kli}) \end{aligned} \quad (3.15)$$

ここで,  $\delta_{kli} > 0$  は  $\sum_i \delta_{kli} = 1$  を満たす補助変数である. 式 (3.15) の等号成立条件は次式で与えられる.

$$\delta_{kli} = \frac{h_{il}^2}{\sum_{i'} h_{i'l}^2} \quad (3.16)$$

式 (3.4) に式 (3.6), (3.9), (3.13), 及び (3.15) を適用すると,  $\mathcal{J} \leq \mathcal{J}^+$  を満たす最終的な上

関数  $\mathcal{J}^+$  が次式のように設計できる.

$$\begin{aligned} \mathcal{J} &\leq \mathcal{J}^+ \\ &= \sum_{i,j} \left[ y_{ij} \log y_{ij} - y_{ij} \left( \gamma_{ij1} \sum_k \alpha_{ijk} \log \frac{f_{ik} g_{kj}}{\alpha_{ijk}} + \gamma_{ij2} \sum_l \beta_{ijl} \log \frac{h_{il} u_{lj}}{\beta_{ijl}} - \gamma_{ij1} \log \gamma_{ij1} \right. \right. \\ &\quad \left. \left. - \gamma_{ij2} \log \gamma_{ij2} \right) - y_{ij} + \sum_k f_{ik} g_{kj} + \sum_l h_{il} u_{lj} \right] + \mu \sum_{k,l} \left[ \frac{1}{\zeta_{kli}} \left( \sum_i f_{ik} h_{il} - \zeta_{kli} \right) \right. \\ &\quad \left. + \log \zeta_{kli} - \frac{1}{2} \log \sum_i f_{ik}^2 - \frac{1}{2} \sum_i \delta_{kli} (2 \log h_{il} - \log \delta_{kli}) \right] \end{aligned} \quad (3.17)$$

$\mathbf{H}$  の更新式は, 式 (3.17) の右辺を  $h_{il}$  で偏微分し 0 とおいた次式 ( $\partial \mathcal{J}^+ / \partial h_{il} = 0$ ) から得られる.

$$\begin{aligned} \sum_j \left[ -y_{ij} \gamma_{ij2} \beta_{ijl} \cdot \frac{1}{h_{il} u_{lj}} u_{lj} + u_{lj} \right] + \mu \sum_k \left[ \frac{1}{\zeta_{kli}} f_{ik} - \delta_{kli} \cdot \frac{1}{h_{il}} \right] &= 0 \\ - \sum_j y_{ij} \gamma_{ij2} \beta_{ijl} \frac{1}{h_{il}} + \sum_j u_{lj} + \mu \sum_k \frac{1}{\zeta_{kli}} f_{ik} - \mu \sum_k \delta_{kli} \frac{1}{h_{il}} &= 0 \end{aligned} \quad (3.18)$$

式 (3.18) を  $\frac{1}{h_{il}}$  について整理すると, 次式となる.

$$\frac{1}{h_{il}} \left[ \sum_j y_{ij} \gamma_{ij2} \beta_{ijl} + \mu \sum_k \delta_{kli} \right] = \sum_j u_{lj} + \mu \sum_k \frac{f_{ik}}{\zeta_{kli}} \quad (3.19)$$

式 (3.19) より,  $h_{il}$  について解くと以下の式に変形できる.

$$h_{il} = \frac{\sum_j y_{ij} \gamma_{ij2} \beta_{ijl} + \mu \sum_k \delta_{kli}}{\sum_j u_{lj} + \mu \sum_k \frac{f_{ik}}{\zeta_{kli}}} \quad (3.20)$$

さらに補助変数の等号成立条件式 (3.7), (3.8), (3.11), (3.12), (3.14), 及び (3.16) を上式に代入すると, 以下に示す式が得られる.

$$h_{il} = \frac{\frac{\sum_j y_{ij} \tilde{h}_{il} u_{lj}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} \tilde{h}_{il'} u_{l'j}} + \mu \sum_k \frac{\tilde{h}_{il}^2}{\sum_{i'} \tilde{h}_{i'l}}}{\sum_j u_{lj} + \mu \sum_k \frac{f_{ik}}{\sum_{i'} f_{i'k} \tilde{h}_{i'l}}} \quad (3.21)$$

ここで,  $\tilde{h}_{il}$  は等号成立条件に含まれていた, 更新前の変数  $h_{il}$  を表す. 式 (3.21) の分子の第二項において, 基底数  $k$  の総和を取っているが第二項の中に  $k$  の要素は含まれていない. よっ

て  $k$  の総和は第二項全体を  $K$  倍することとなり，次式で表される．

$$h_{il} = \frac{\sum_j y_{ij} \tilde{h}_{il} u_{lj}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} \tilde{h}_{il'} u_{l'j}} + \mu K \frac{\tilde{h}_{il}^2}{\sum_{i'} \tilde{h}_{i'l}^2} \quad (3.22)$$

$$\sum_j u_{lj} + \mu \sum_k \frac{f_{ik}}{\sum_{i'} f_{i'k} \tilde{h}_{i'l}}$$

さらに，式 (3.22) の分子の全ての項に含まれている共通因数  $\tilde{h}_{il}$  をくくりだすことで最終的な更新式の形は次式となる．

$$h_{il} = \tilde{h}_{il} \cdot \frac{\sum_j \frac{y_{ij}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} \tilde{h}_{il'} u_{l'j}} u_{lj} + \mu K \frac{\tilde{h}_{il}}{\sum_{i'} \tilde{h}_{i'l}^2}}{\sum_j u_{lj} + \mu \sum_k \frac{f_{ik}}{\sum_{i'} f_{i'k} \tilde{h}_{i'l}}} \quad (3.23)$$

コサイン類似度に基づく PSNMF の罰則項には  $\mathbf{G}$  及び  $\mathbf{U}$  の要素が含まれていないため，式 (3.23) のように更新式を求めようと式 (3.10) の右辺を  $g_{kj}$  及び  $u_{lj}$  で偏微分し 0 とおくと，得られる式では罰則項が含まれない．そのため  $\mathbf{G}$  及び  $\mathbf{U}$  の更新則は罰則項を付与する前の式 (2.20) と式 (2.22) と同様である．

## 第 4 章

# 音楽信号の分離実験

### 4.1 実験条件

提案手法の有効性を確認するために、音楽信号の音源分離性能を直交化に基づく PSNMF (従来手法) とコサイン類似度に基づく PSNMF (提案手法) に関して比較した。実験では、データセットとして songKitamura [10, 16] の YMH MU-1000 の音源を用いた。音源にはオーボエ (Ob.), トランペット (Tp.), ホルン (Hr.), フルート (Fl.), ヴァイオリン (Vn.), クラリネット (Cl.), ピアノ (Pf.), ハープシコード (Hp.), トロンボーン (Tr.), ファゴット (Fg.), チェロ (Vc.) の 11 種類の、学習ステージ用の 2 オクターブ上昇音階 (Fig. 4.1) と分離ステージ用のメロディパート (Fig. 4.2) を用いた。11 種の楽器から 2 つを選択し、それらのメロディパートを等パワーで混合し混合信号とした。従って、合計 90 曲での評価となった。評価指標には、信号対歪み比 (source-to-distortion ratio: SDR) [17] を用いた。

重み係数  $\mu$  は対数的に 72 段階に分割した値を使用した。各行列の初期値は区間 (0, 1) の一様乱数とした。短時間フーリエ変換の窓長とシフト幅はそれぞれ 92.9 ms 及び 46.4 ms とした。目的音源と非目的音源の基底ベクトル数はそれぞれ  $K = 27$  及び  $L = 50$  とした。

### 4.2 実験結果

Fig. 4.3 から Fig. 4.92 に曲ごとの分離結果の例を示す。提案手法のほうが良い SDR 値を示した例として、Figs. 4.12, 4.18 及び 4.87 等が挙げられる。特に、Fig. 4.87 では従来手法よりおよそ 3 dB 高い値が得られた。例として示した図はある重み係数  $\mu$  において従来手法より高い SDR 値が観測された。しかし、特定の重み係数  $\mu$  で SDR 値が最大になるといった法則性は見られなかった。楽器別に SDR 値を着目すると、Vc. 及び Vn. を目的音としたとき、16 曲中 15 曲で提案手法が従来手法の分離精度に比べ勝っていた。反対に、提案手法に比べ低い SDR 値を観測したデータは 90 曲中 31 曲存在した。例として Figs. 4.4, 4.9 及び 4.22 等が挙げられる。これらの図では、提案手法はほとんど分離精度は向上することがなかった。特に Cl. と Fl. を目的音にしたときそれぞれ組み合わせが 8 個ずつある中で Cl. は 7 曲、Fl. ではすべての音源分離において提案手法は従来手法より低い値を示した。これより、Cl. 及び Fl. は目的音とする場合、従来手法による音源分離のほうが良いと思われるが、その理由は掴めない。

Ob. or Hr. or Tp.

Fl. or Cl. or Vn.

Hp. or Pf.

Fg. or Tb. or Vc.

Fig. 4.1: Two-octave ascending notes.

Ob. or Hr. or Tp.

Fl. or Cl. or Vn.

Hp. or Pf.

Fg. or Tb. or Vc.

Fig. 4.2: Four melody parts.

かった。さらに、提案手法と従来手法を比較すると、提案手法のグラフはすべての図を通して重み係数の最適値がピーキーであることが分かる。これは、提案手法のコサイン類似度罰則がスケールの不定性の影響を受けないことが原因であると考えられる。例として、Fig. 4.75 を挙げる。従来手法では SDR が最大値を示した後、さらに  $\mu$  の値を増加させると SDR 値は緩やかに低下する傾向がみられる。しかし、提案手法では、SDR 値が最大値を迎えた後さらに  $\mu$  の値を増加させると急激に SDR 値は低くなった。

Table 4.1 に、各曲の最大 SDR 値を全 90 曲に関して求めた際の平均値と中央値を示す。平均値及び中央値ともに従来手法に比べ提案手法が SDR 値が高いことが分かる。前述した通り提案手法では、ある  $\mu$  の時に従来手法より高い SDR 値が観測され、過半数のデータでそのような傾向が見られた。これらより、提案手法は従来手法より分離精度が高いことが示された。

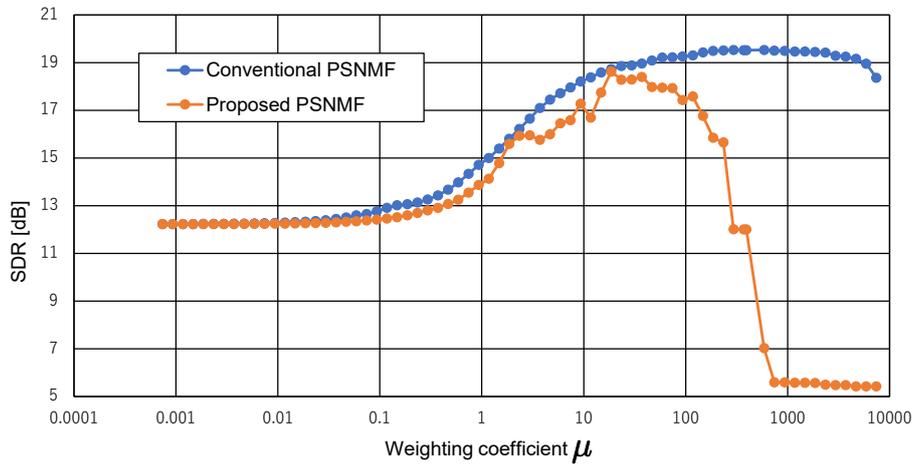


Fig. 4.3: SDR values of conventional and proposed PSNMFs for Cl.&Fg. mixtures.

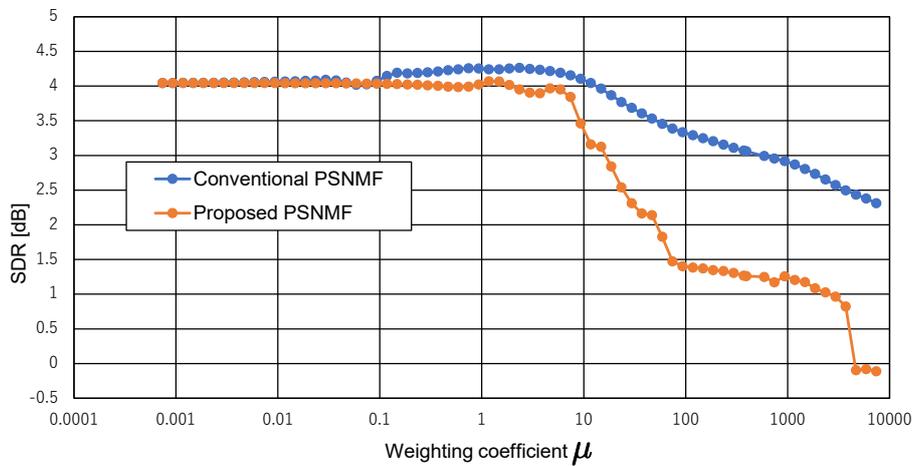


Fig. 4.4: SDR values of conventional and proposed PSNMFs for Cl.&Hp. mixtures.

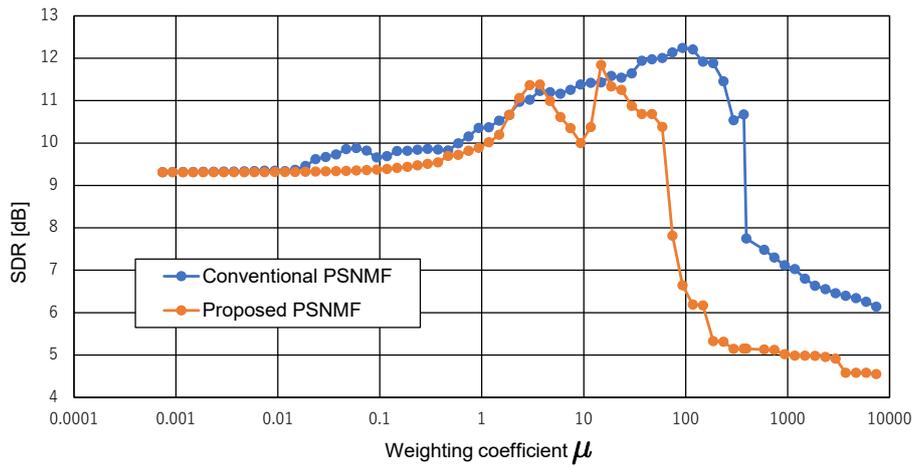


Fig. 4.5: SDR values of conventional and proposed PSNMFs for Cl.&Hr. mixtures.

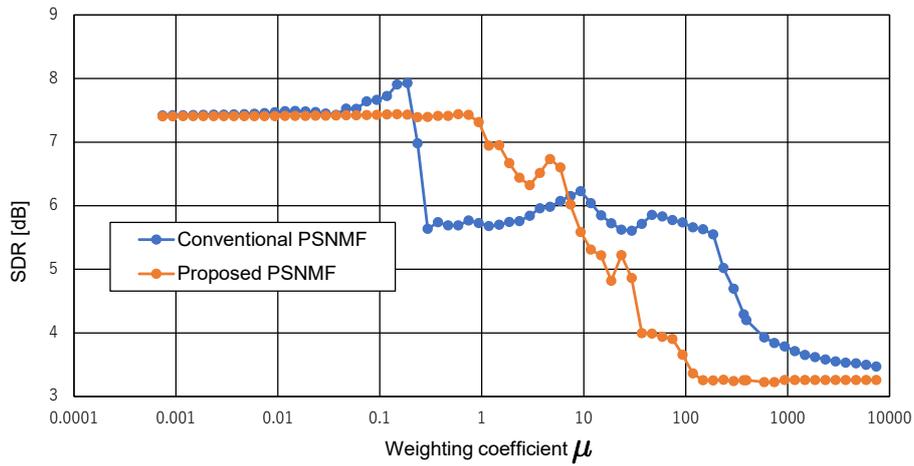


Fig. 4.6: SDR values of conventional and proposed PSNMFs for Cl.&Ob. mixtures.

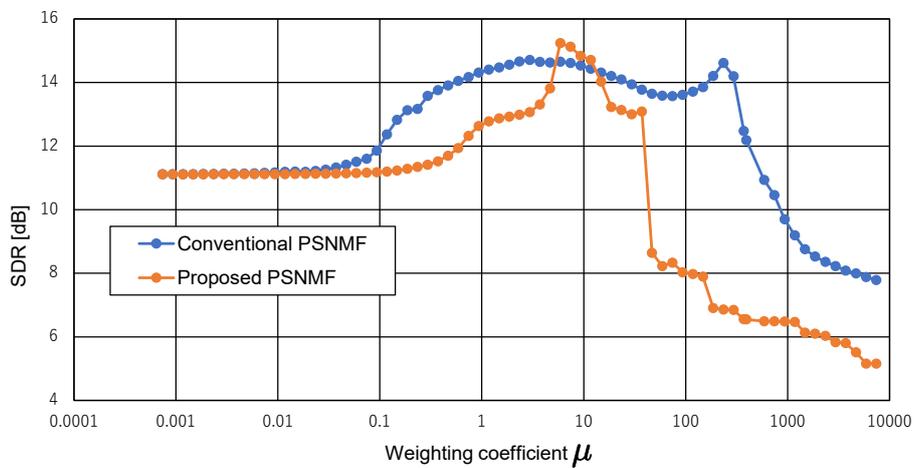


Fig. 4.7: SDR values of conventional and proposed PSNMFs for Cl.&Pf. mixtures.

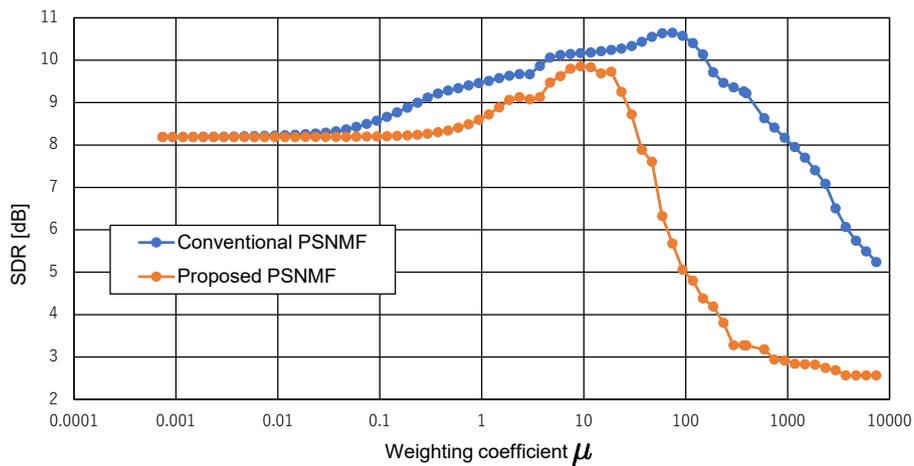


Fig. 4.8: SDR values of conventional and proposed PSNMFs for Cl.&Tb. mixtures.

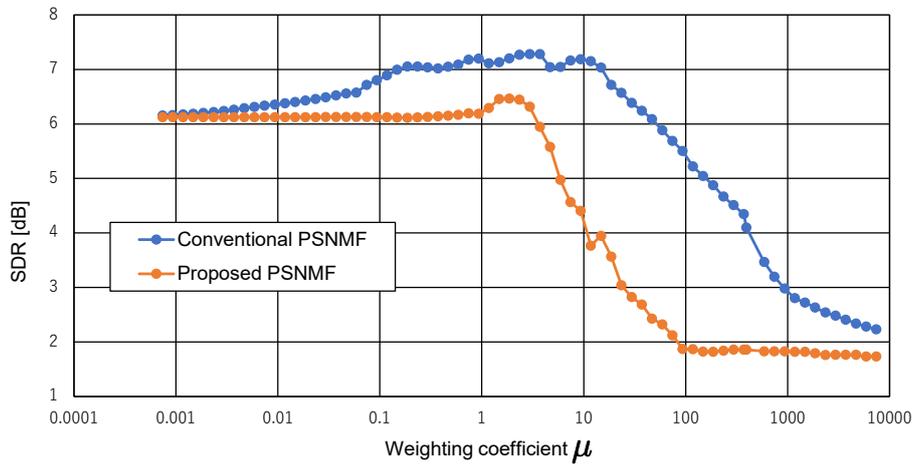


Fig. 4.9: SDR values of conventional and proposed PSNMFs for Cl.&Tp. mixtures.

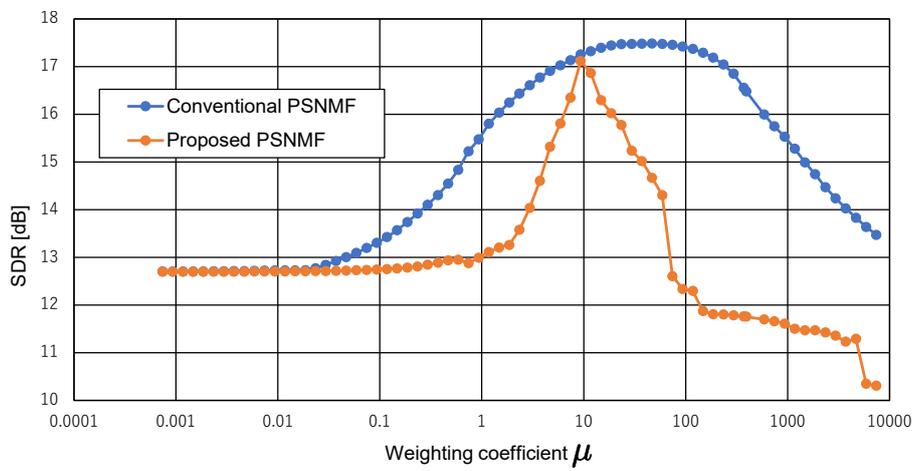


Fig. 4.10: SDR values of conventional and proposed PSNMFs for Cl.&Vc. mixtures.

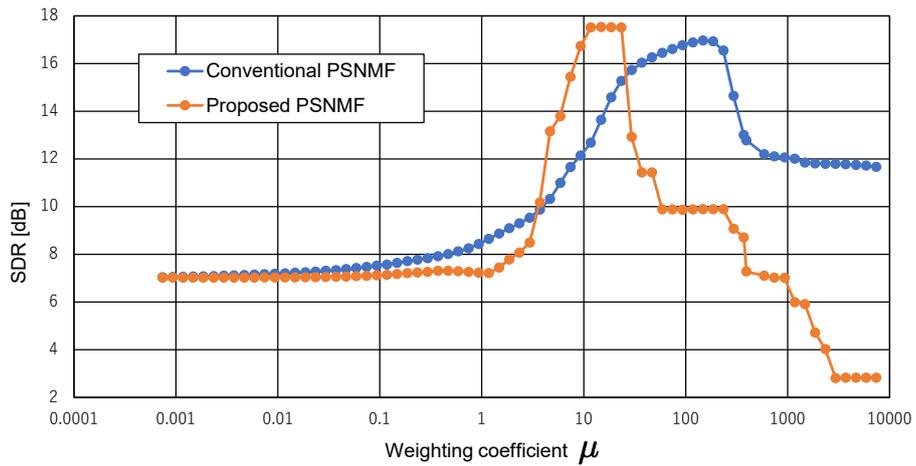


Fig. 4.11: SDR values of conventional and proposed PSNMFs for Fg.&Cl. mixtures.

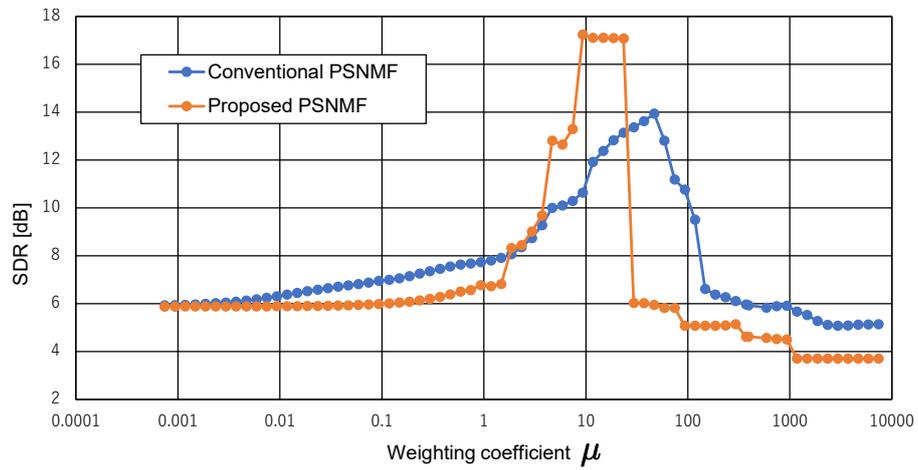


Fig. 4.12: SDR values of conventional and proposed PSNMFs for Fg.&Fl. mixtures.

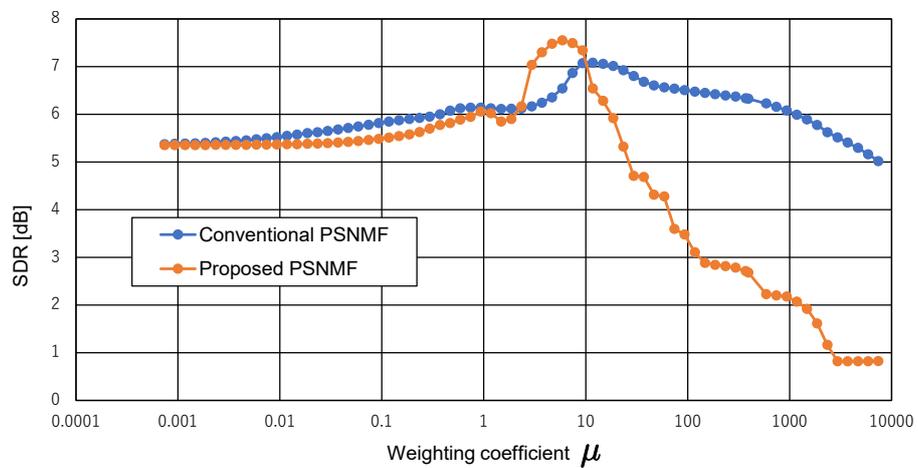


Fig. 4.13: SDR values of conventional and proposed PSNMFs for Fg.&Hp. mixtures.

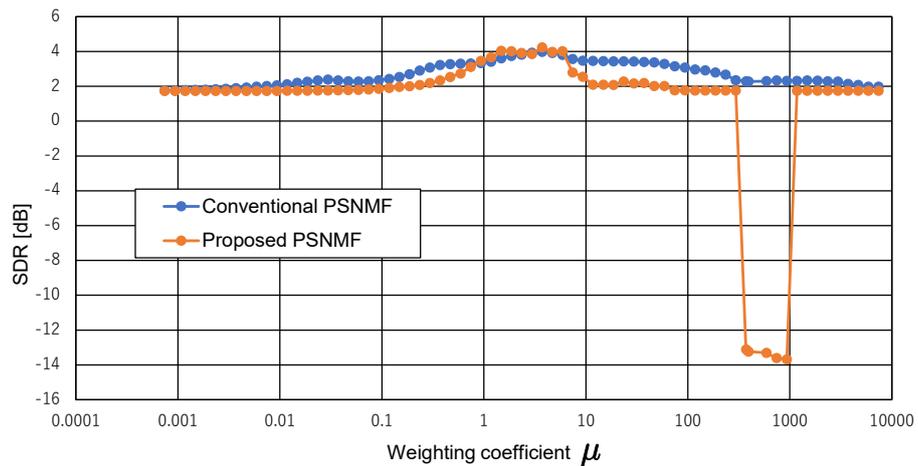


Fig. 4.14: SDR values of conventional and proposed PSNMFs for Fg.&Hr. mixtures.

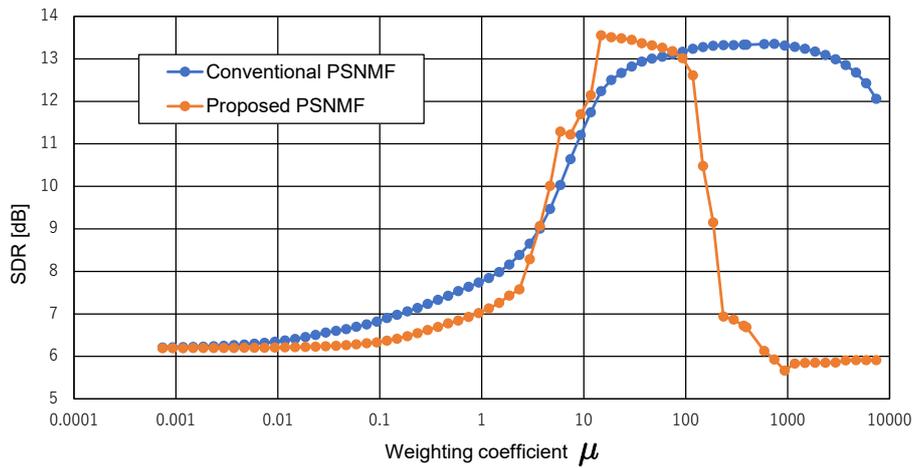


Fig. 4.15: SDR values of conventional and proposed PSNMFs for Fig.&Ob. mixtures.

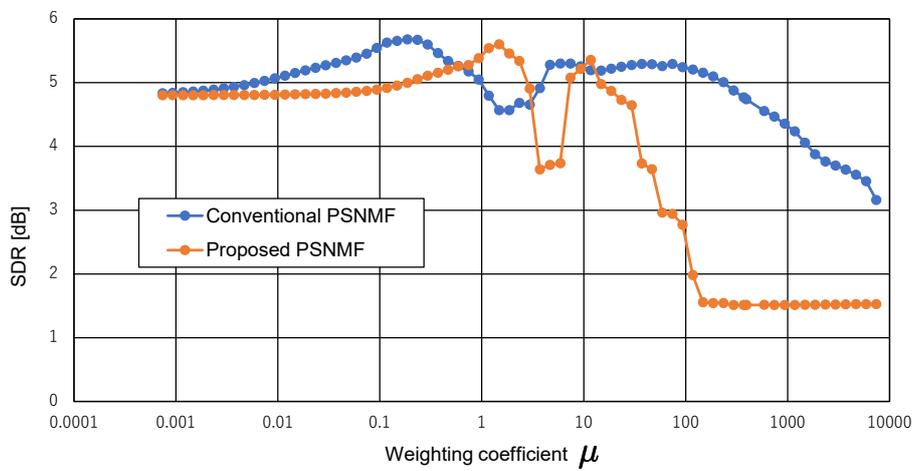


Fig. 4.16: SDR values of conventional and proposed PSNMFs for Fig.&Pf. mixtures.

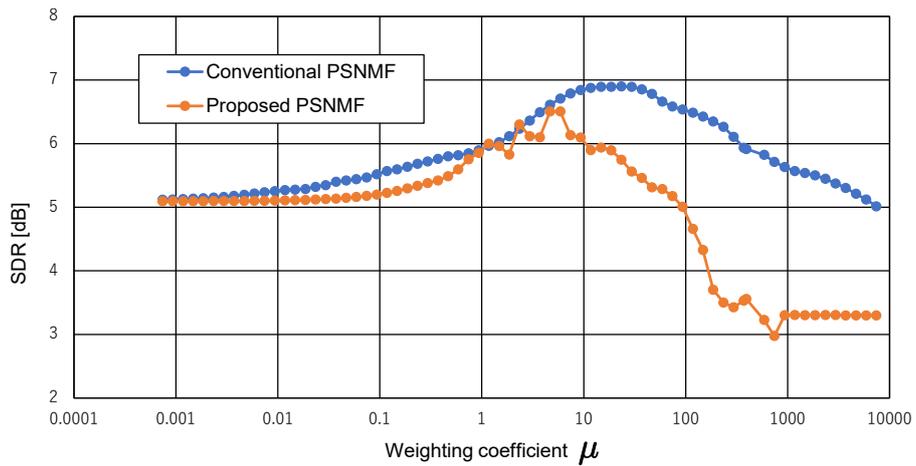


Fig. 4.17: SDR values of conventional and proposed PSNMFs for Fig.&Tp. mixtures.

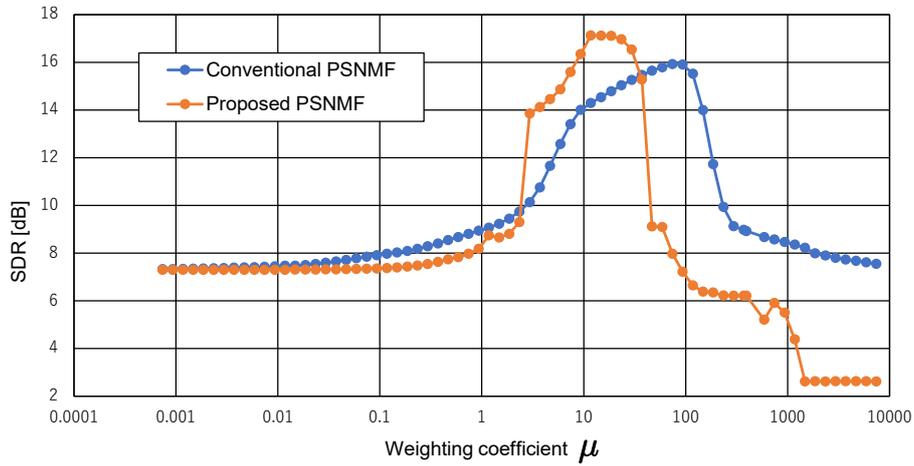


Fig. 4.18: SDR values of conventional and proposed PSNMFs for Fg.&Vn. mixtures.

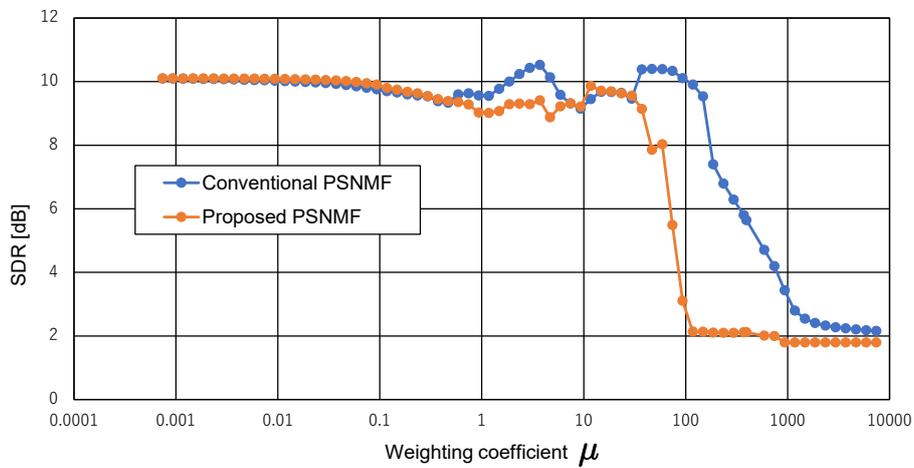


Fig. 4.19: SDR values of conventional and proposed PSNMFs for Fl.&Fg. mixtures.

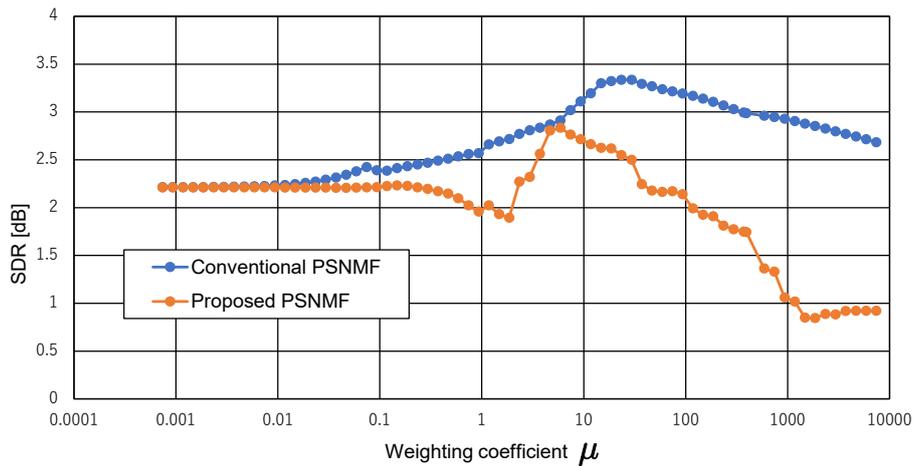


Fig. 4.20: SDR values of conventional and proposed PSNMFs for Fl.&Hp. mixtures.

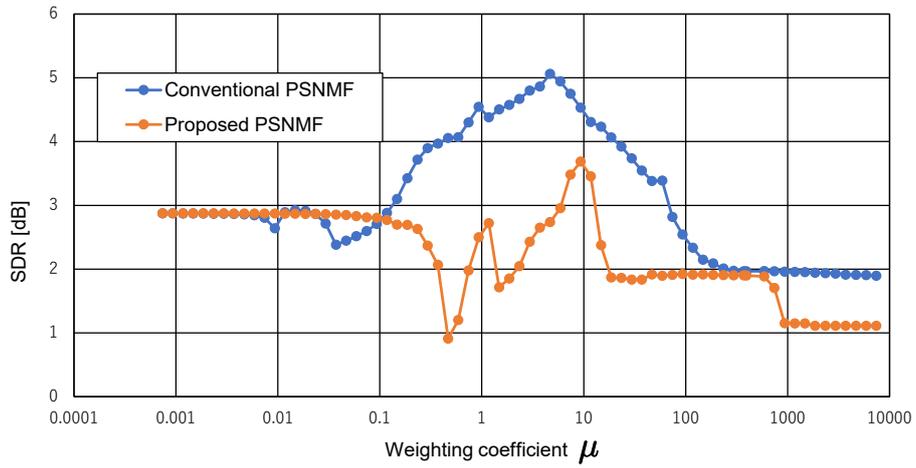


Fig. 4.21: SDR values of conventional and proposed PSNMFs for Fl.&Hr. mixtures.

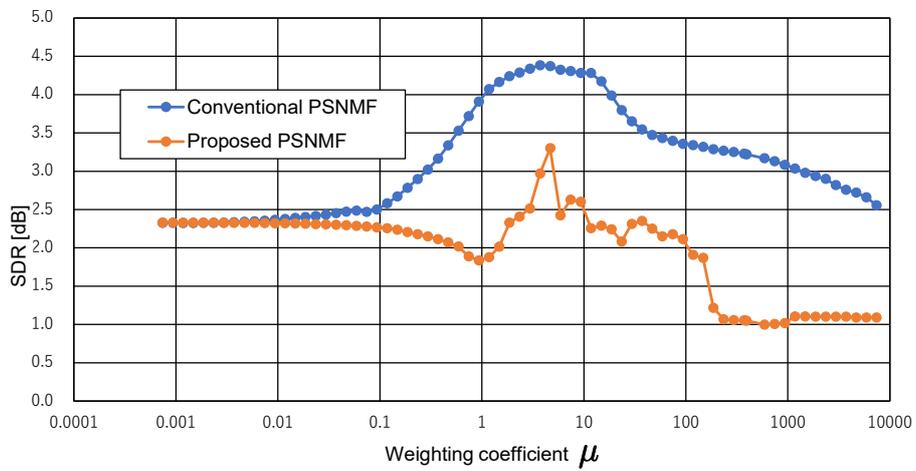


Fig. 4.22: SDR values of conventional and proposed PSNMFs for Fl.&Ob. mixtures.

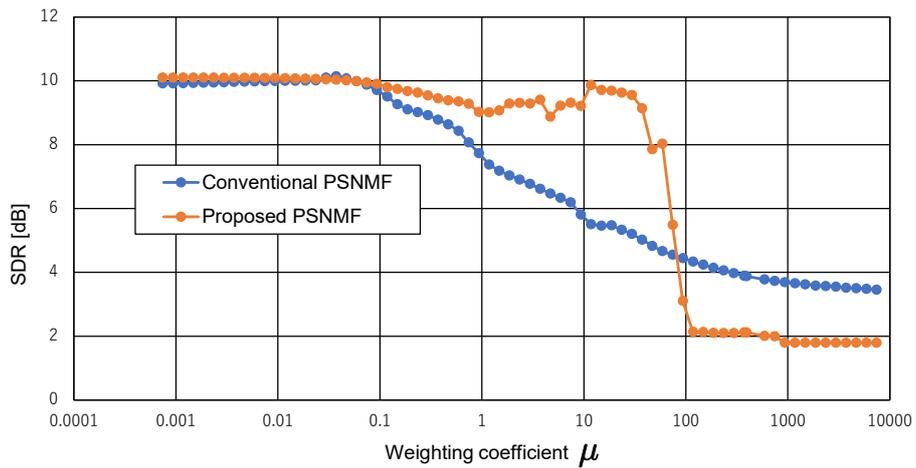


Fig. 4.23: SDR values of conventional and proposed PSNMFs for Fl.&Pf. mixtures.

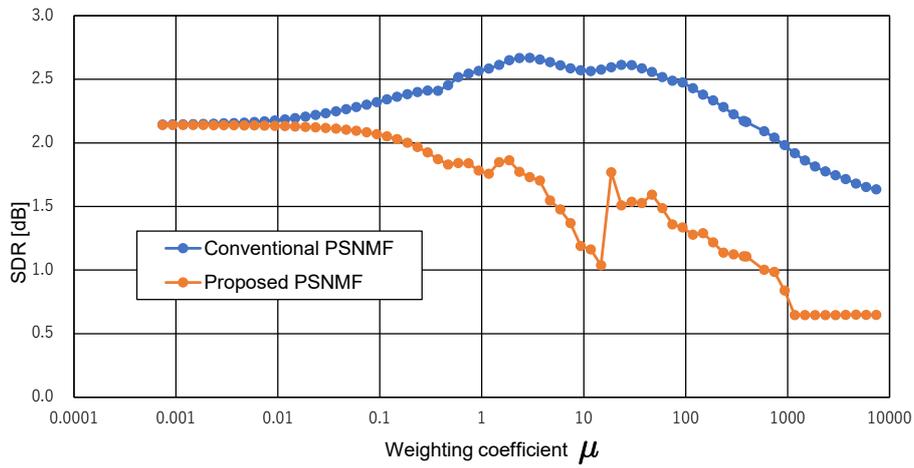


Fig. 4.24: SDR values of conventional and proposed PSNMFs for Fl.&Tb. mixtures.

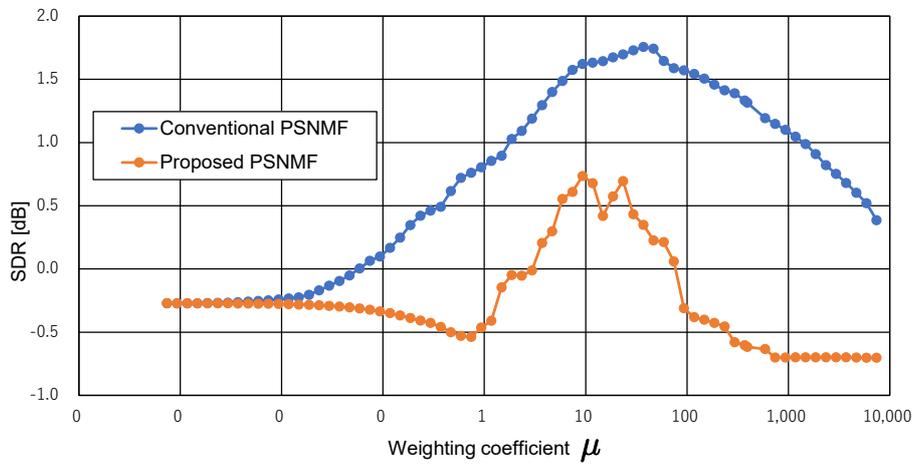


Fig. 4.25: SDR values of conventional and proposed PSNMFs for Fl.&Tp. mixtures.

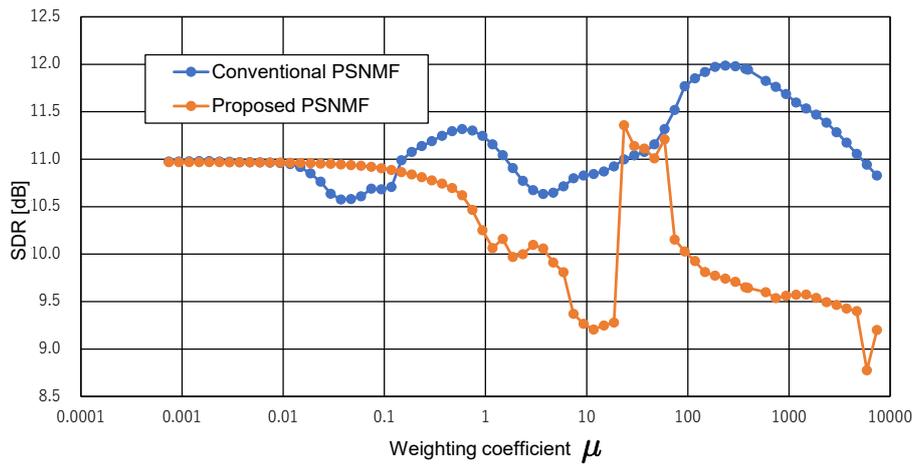


Fig. 4.26: SDR values of conventional and proposed PSNMFs for Fl.&Vc. mixtures.

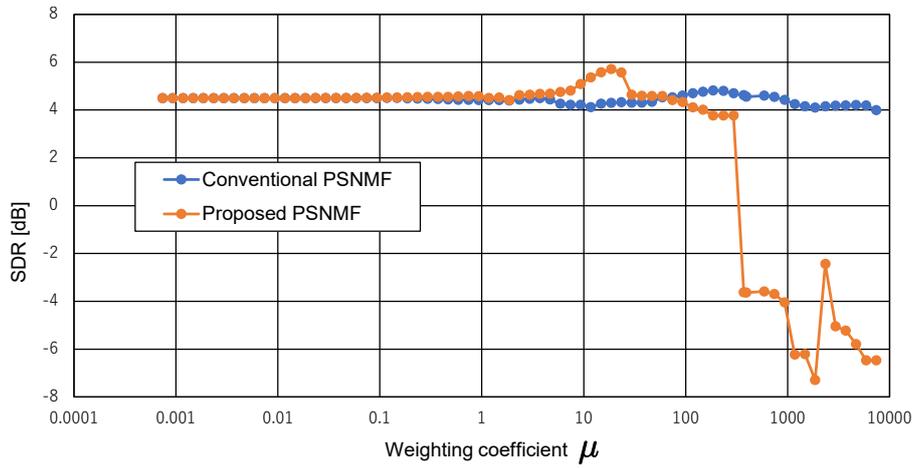


Fig. 4.27: SDR values of conventional and proposed PSNMFs for Hp.&Cl. mixtures.

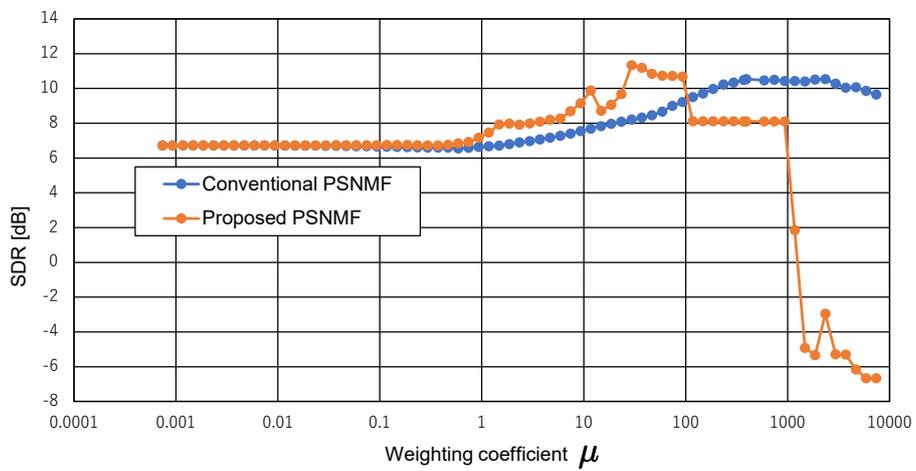


Fig. 4.28: SDR values of conventional and proposed PSNMFs for Hp.&Fg. mixtures.

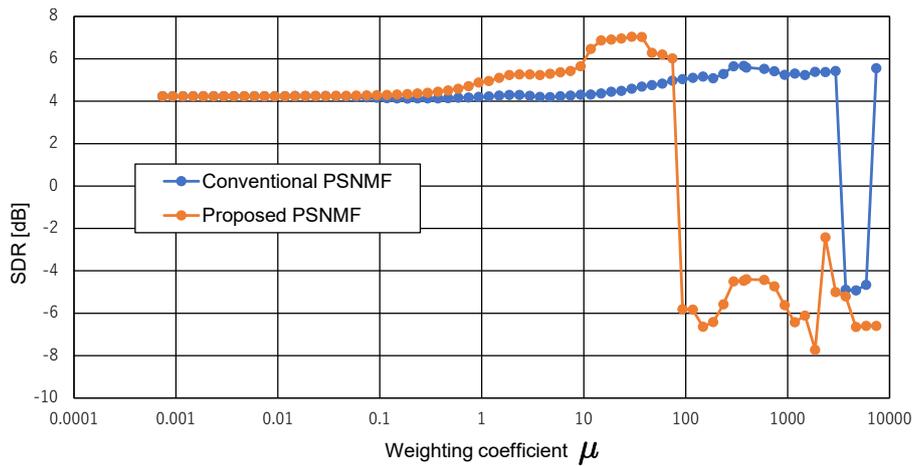


Fig. 4.29: SDR values of conventional and proposed PSNMFs for Hp.&Fl. mixtures.

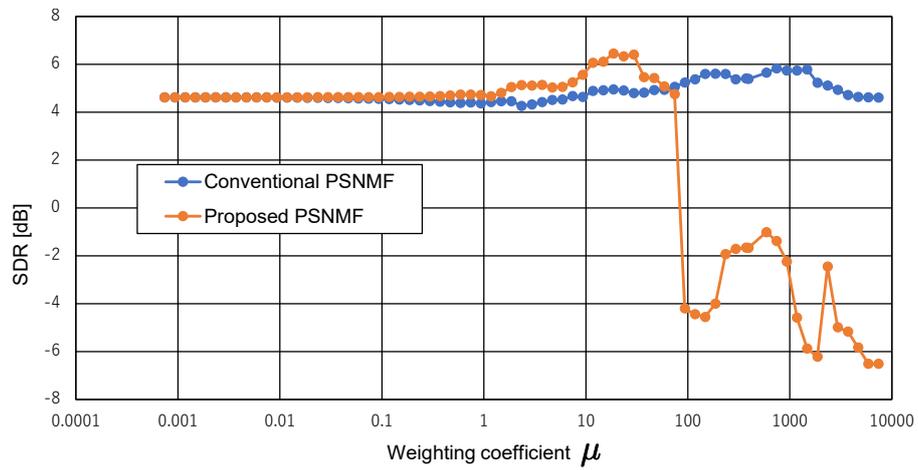


Fig. 4.30: SDR values of conventional and proposed PSNMFs for Hp.&amp;Hr. mixtures.

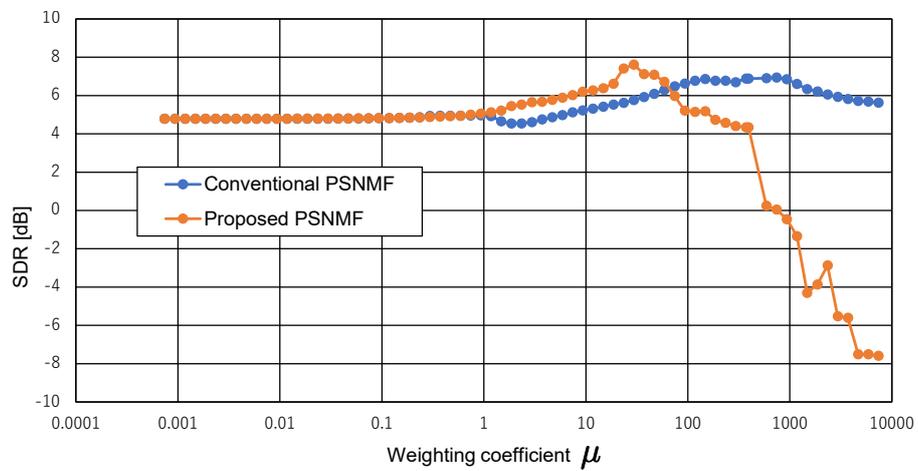


Fig. 4.31: SDR values of conventional and proposed PSNMFs for Hp.&amp;Ob. mixtures.

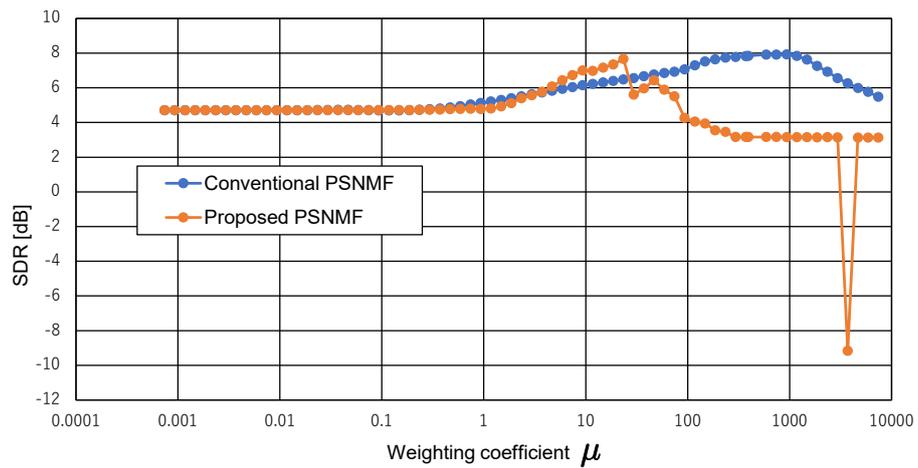


Fig. 4.32: SDR values of conventional and proposed PSNMFs for Hp.&amp;Tb. mixtures.

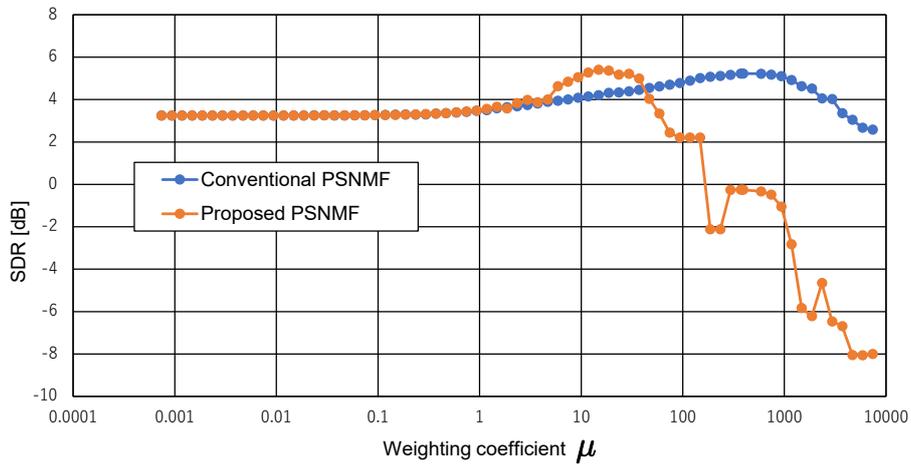


Fig. 4.33: SDR values of conventional and proposed PSNMFs for Hp.&Tp. mixtures.

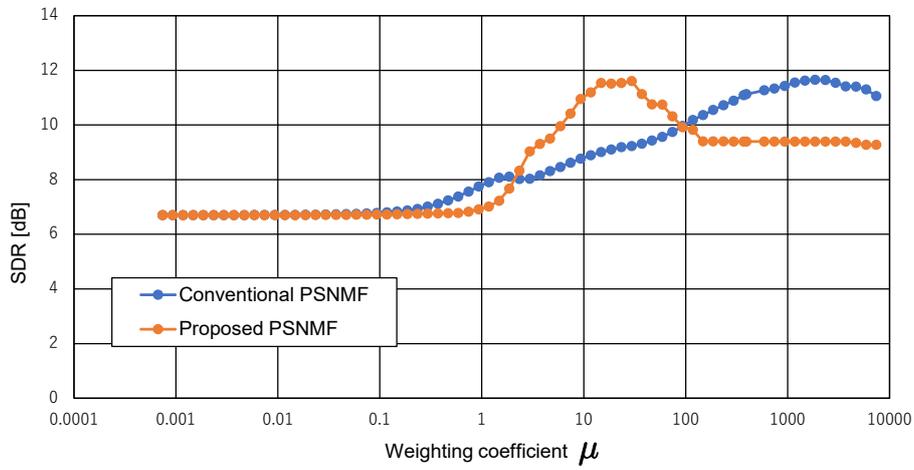


Fig. 4.34: SDR values of conventional and proposed PSNMFs for Hp.&Vc. mixtures.

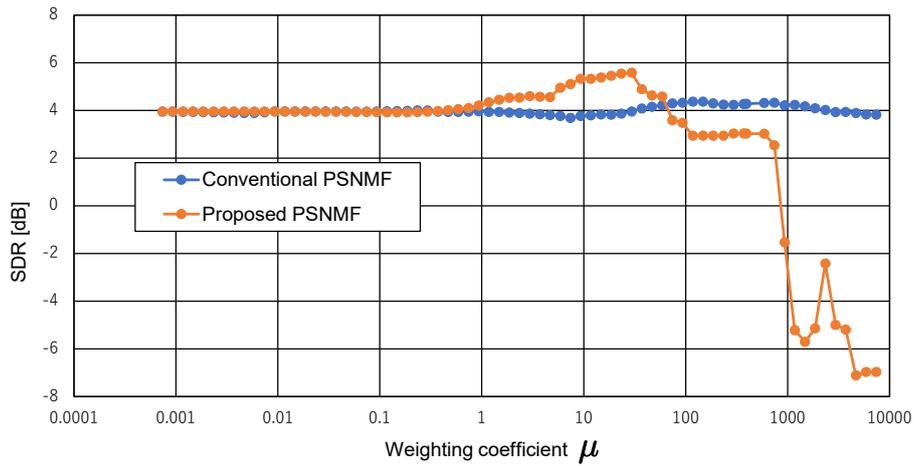


Fig. 4.35: SDR values of conventional and proposed PSNMFs for Hp.&Vn. mixtures.

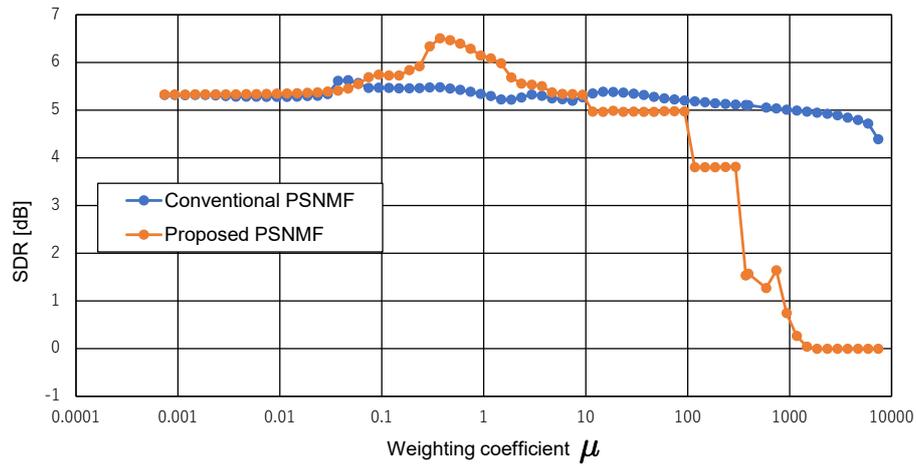


Fig. 4.36: SDR values of conventional and proposed PSNMFs for Hr.&Cl. mixtures.

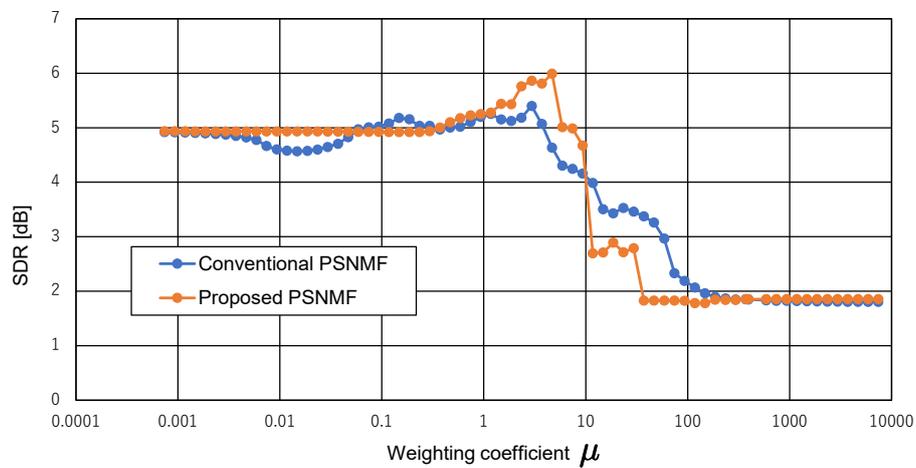


Fig. 4.37: SDR values of conventional and proposed PSNMFs for Hr.&Fg. mixtures.

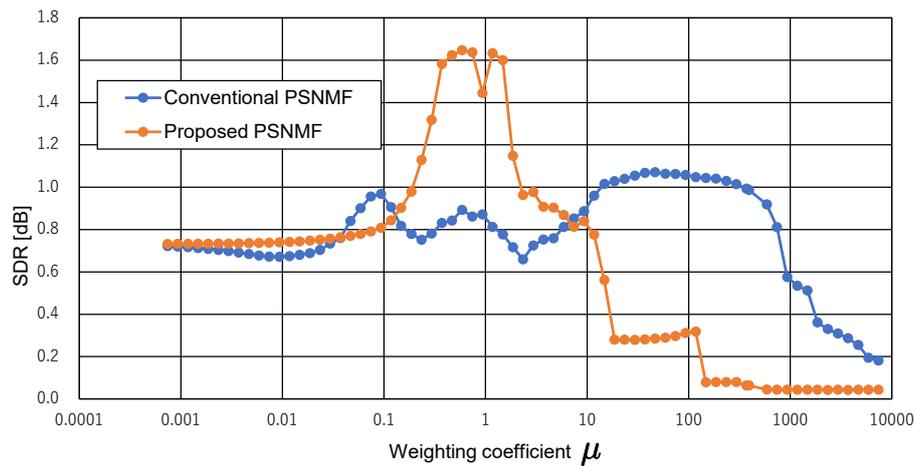


Fig. 4.38: SDR values of conventional and proposed PSNMFs for Hr.&Fl. mixtures.

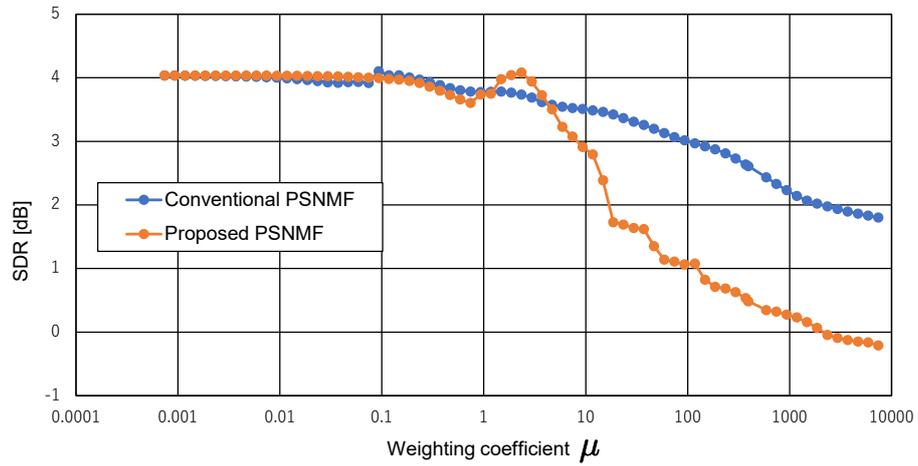


Fig. 4.39: SDR values of conventional and proposed PSNMFs for Hr.&amp;Hp. mixtures.

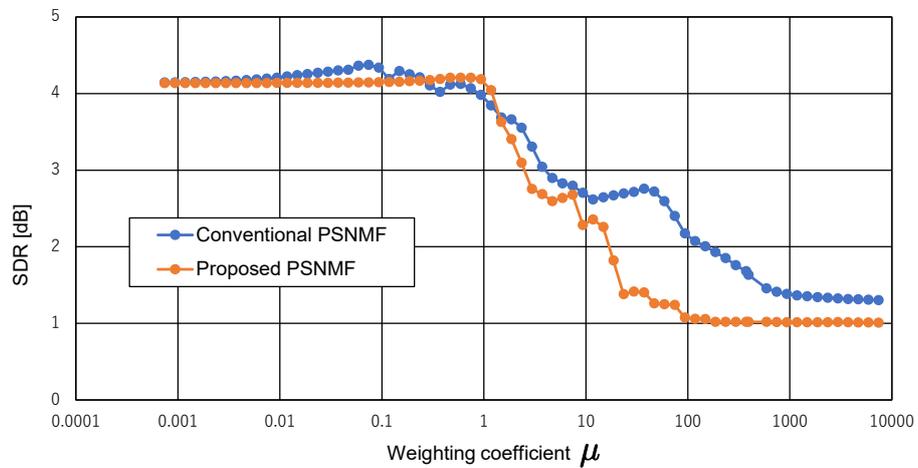


Fig. 4.40: SDR values of conventional and proposed PSNMFs for Hr.&amp;Pf. mixtures.

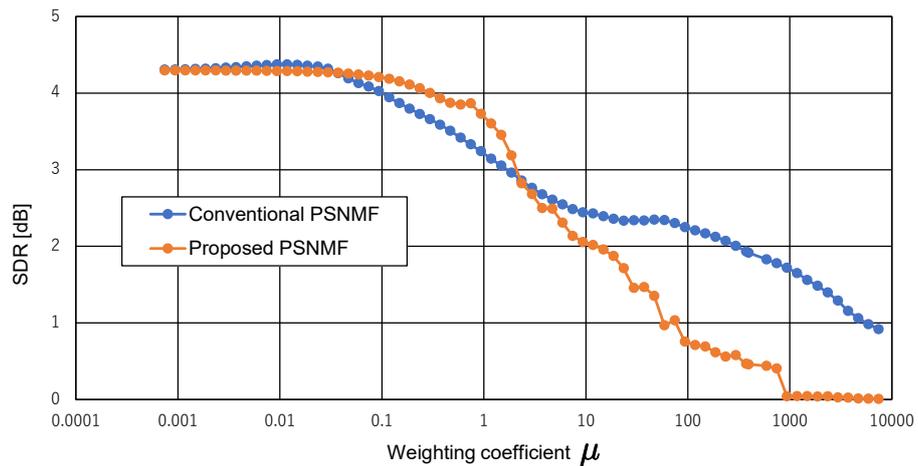


Fig. 4.41: SDR values of conventional and proposed PSNMFs for Hr.&amp;Tb. mixtures.

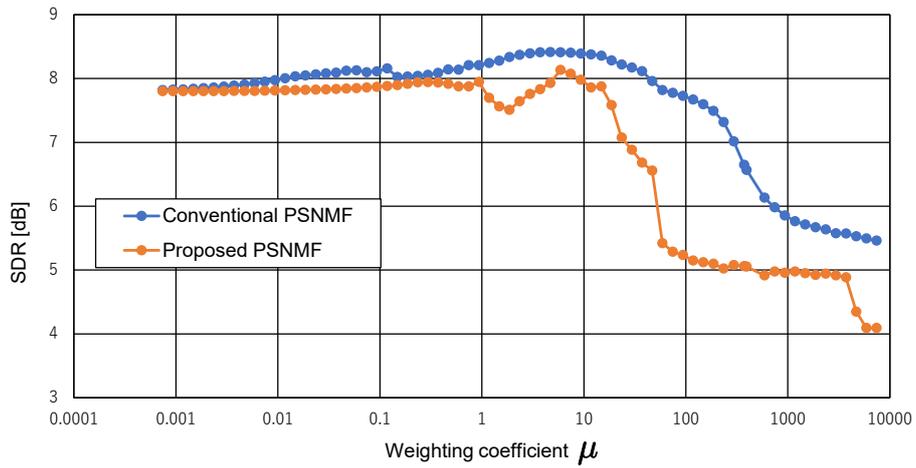


Fig. 4.42: SDR values of conventional and proposed PSNMFs for Hr.&Vc. mixtures.

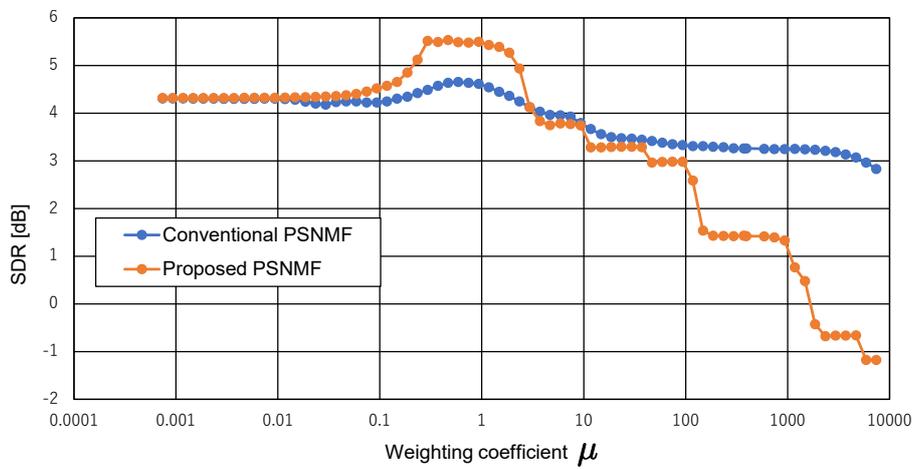


Fig. 4.43: SDR values of conventional and proposed PSNMFs for Hr.&Vn. mixtures.

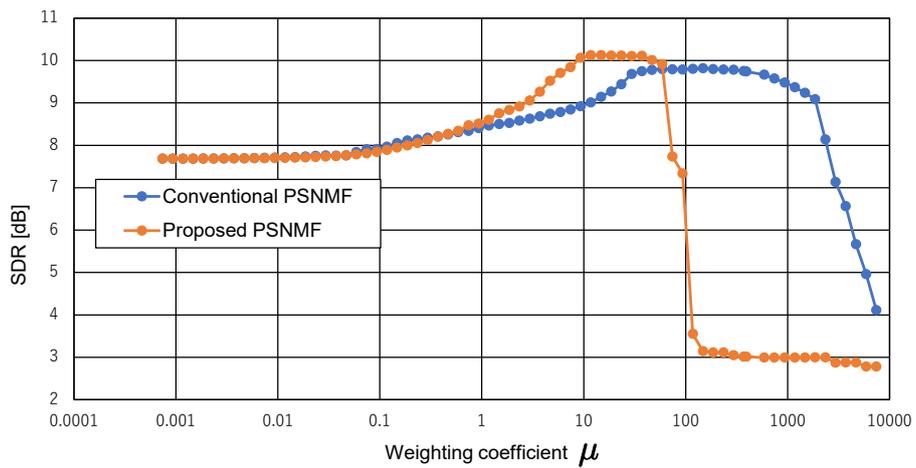


Fig. 4.44: SDR values of conventional and proposed PSNMFs for Ob.&Cl. mixtures.

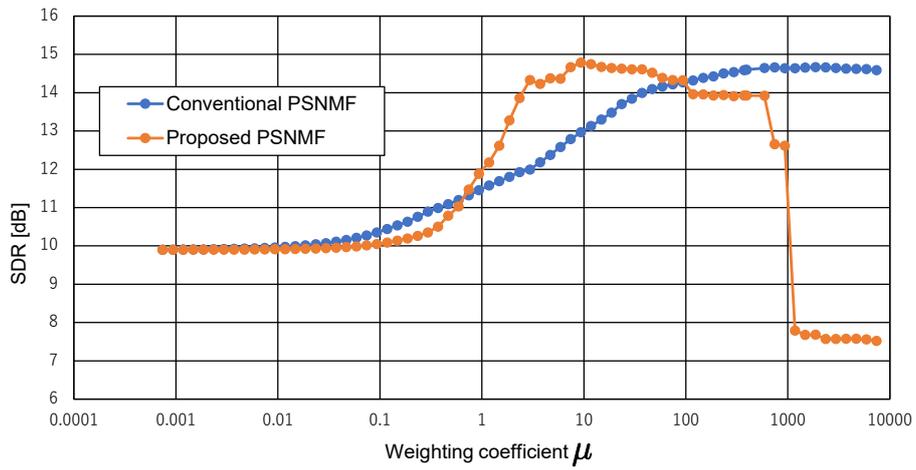


Fig. 4.45: SDR values of conventional and proposed PSNMFs for Ob.&Fg. mixtures.

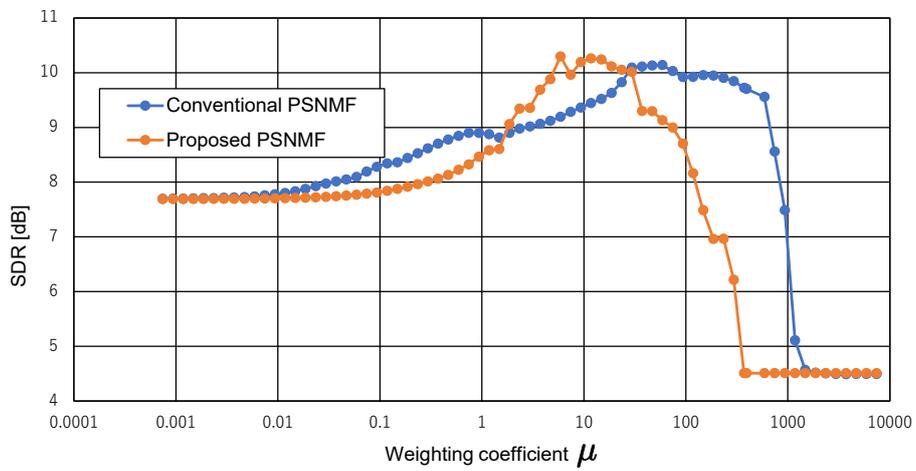


Fig. 4.46: SDR values of conventional and proposed PSNMFs for Ob.&Fl. mixtures.

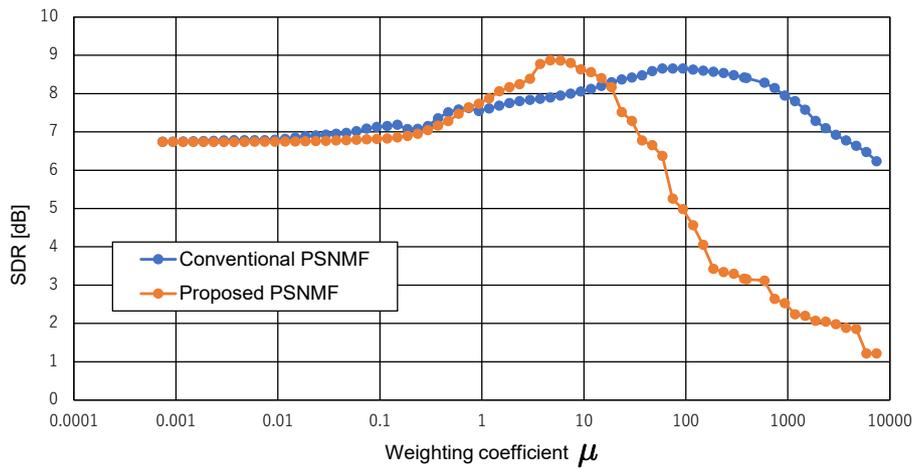


Fig. 4.47: SDR values of conventional and proposed PSNMFs for Ob.&Hp. mixtures.

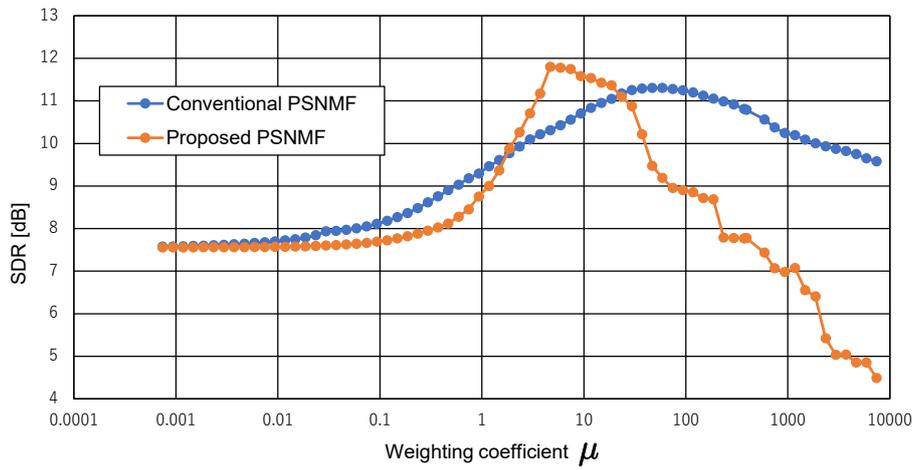


Fig. 4.48: SDR values of conventional and proposed PSNMFs for Ob.&amp;Pf. mixtures.

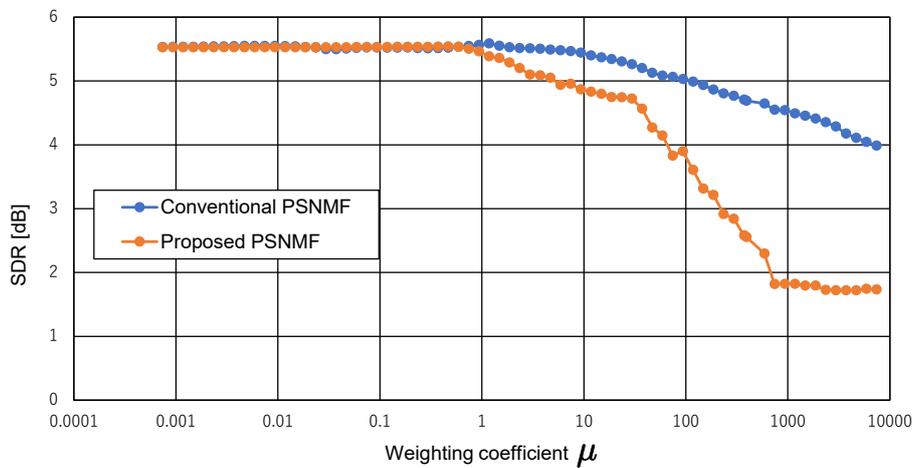


Fig. 4.49: SDR values of conventional and proposed PSNMFs for Ob.&amp;Tb. mixtures.

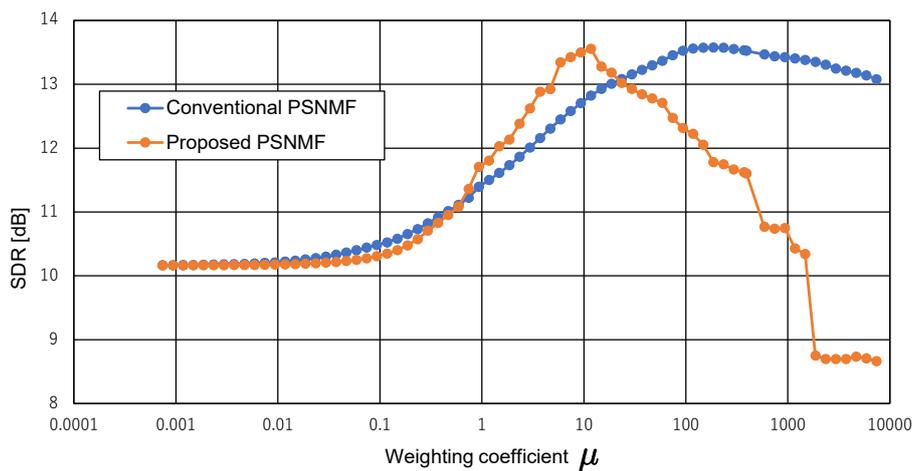


Fig. 4.50: SDR values of conventional and proposed PSNMFs for Ob.&amp;Vc. mixtures.

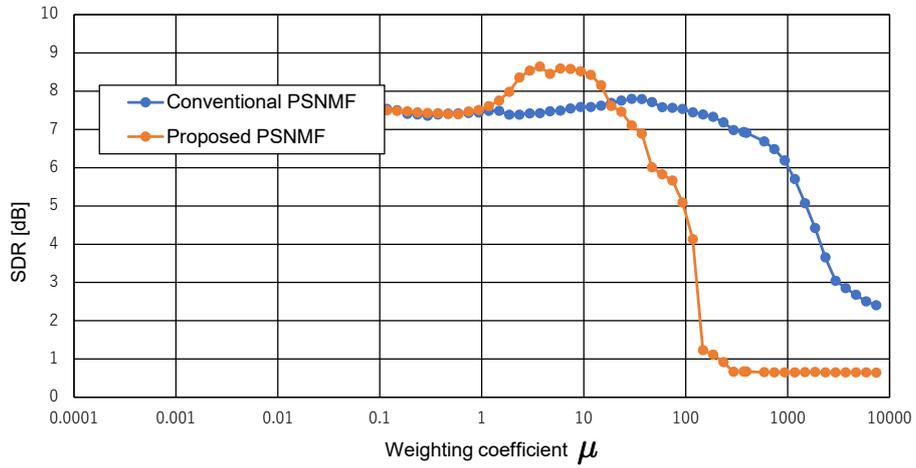


Fig. 4.51: SDR values of conventional and proposed PSNMFs for Ob.&Vn. mixtures.

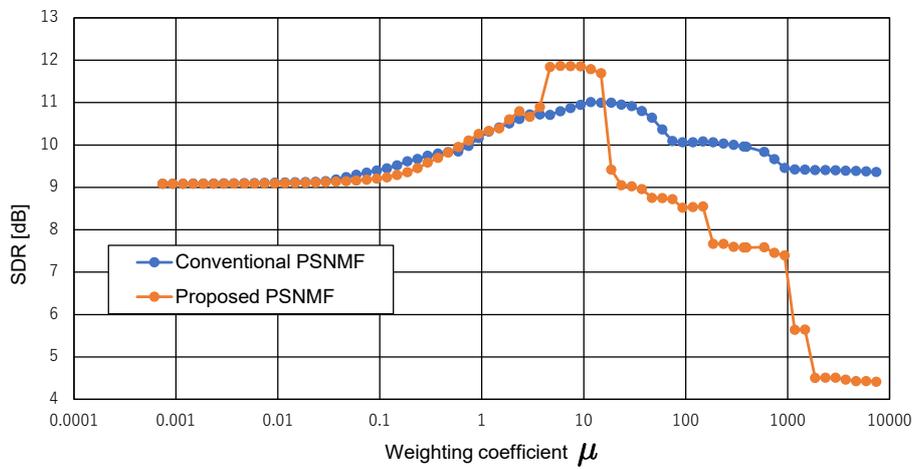


Fig. 4.52: SDR values of conventional and proposed PSNMFs for Pf.&Cl. mixtures.

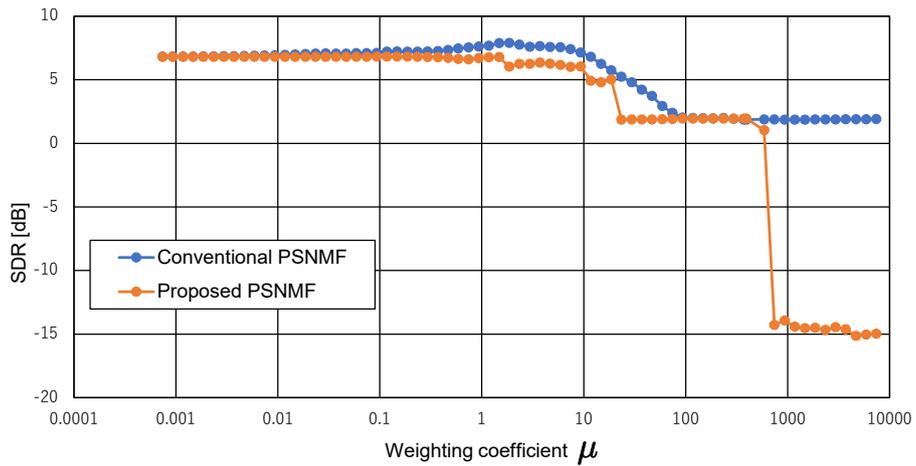


Fig. 4.53: SDR values of conventional and proposed PSNMFs for Pf.&Fg. mixtures.

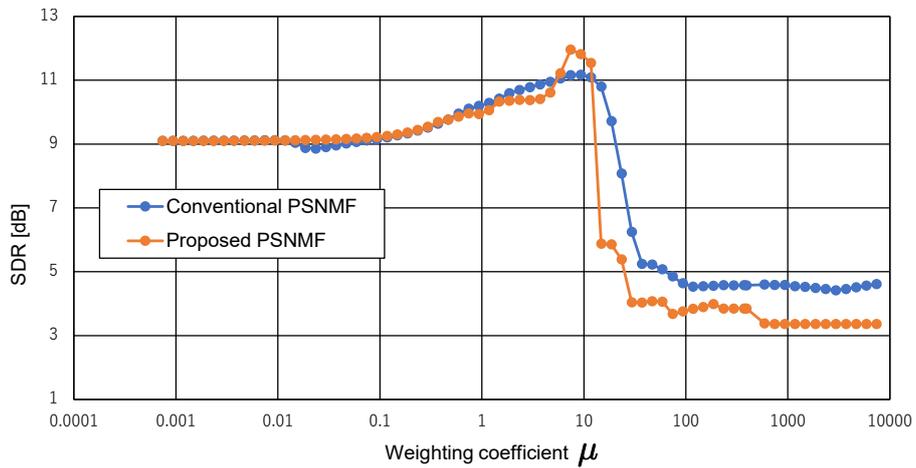


Fig. 4.54: SDR values of conventional and proposed PSNMFs for Pf.&amp;Fl. mixtures.

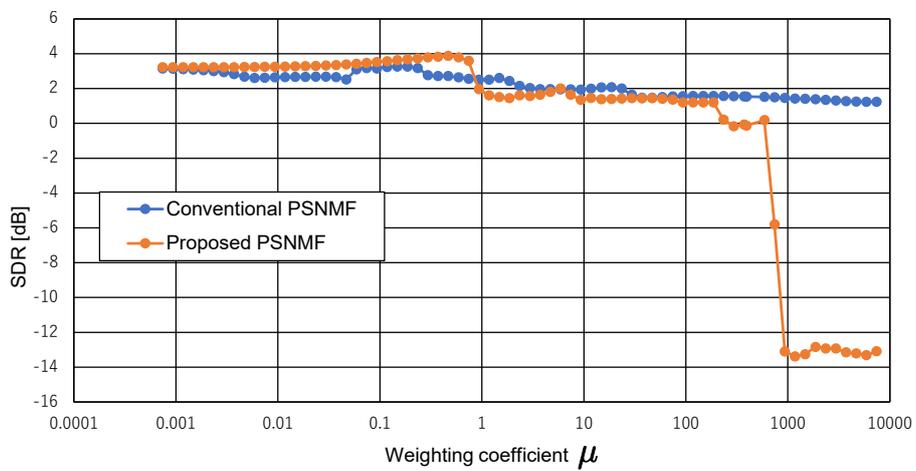


Fig. 4.55: SDR values of conventional and proposed PSNMFs for Pf.&amp;Hr. mixtures.

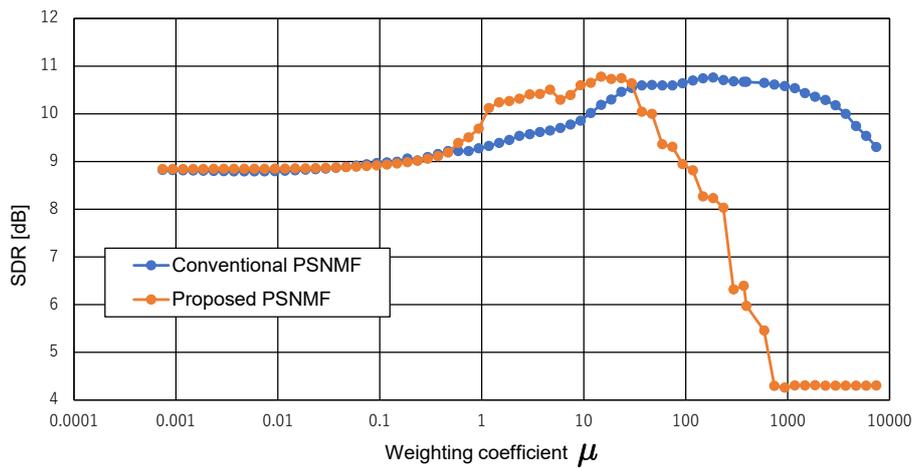


Fig. 4.56: SDR values of conventional and proposed PSNMFs for Pf.&amp;Ob. mixtures.

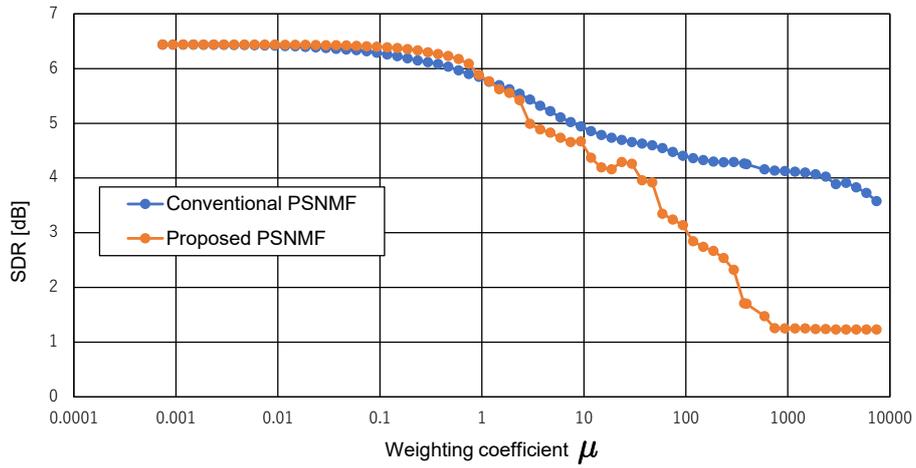


Fig. 4.57: SDR values of conventional and proposed PSNMFs for Pf.&Tb. mixtures.

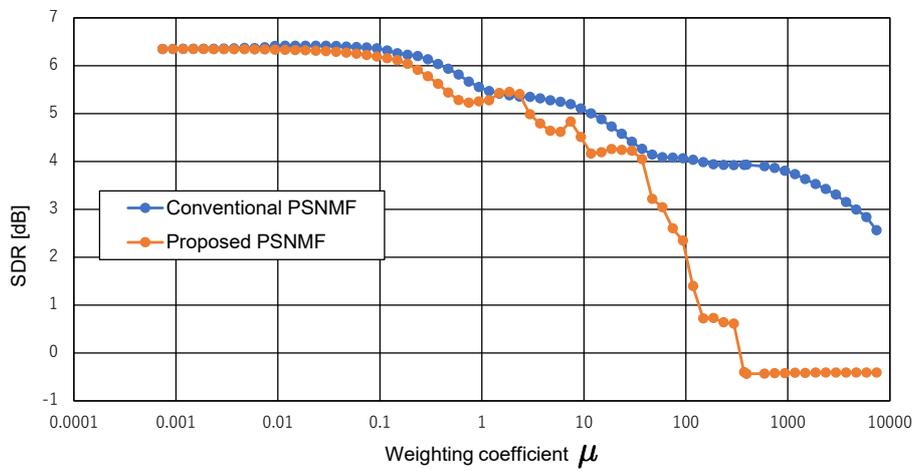


Fig. 4.58: SDR values of conventional and proposed PSNMFs for Pf.&Tp. mixtures.

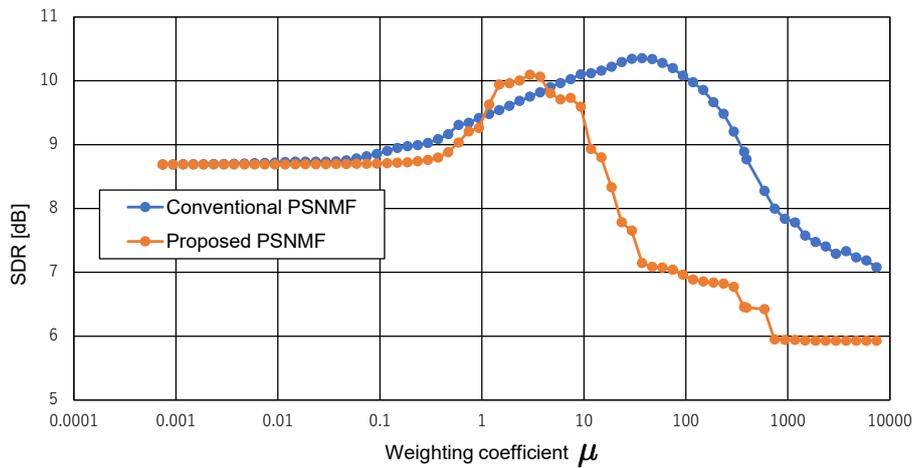


Fig. 4.59: SDR values of conventional and proposed PSNMFs for Pf.&Vc. mixtures.

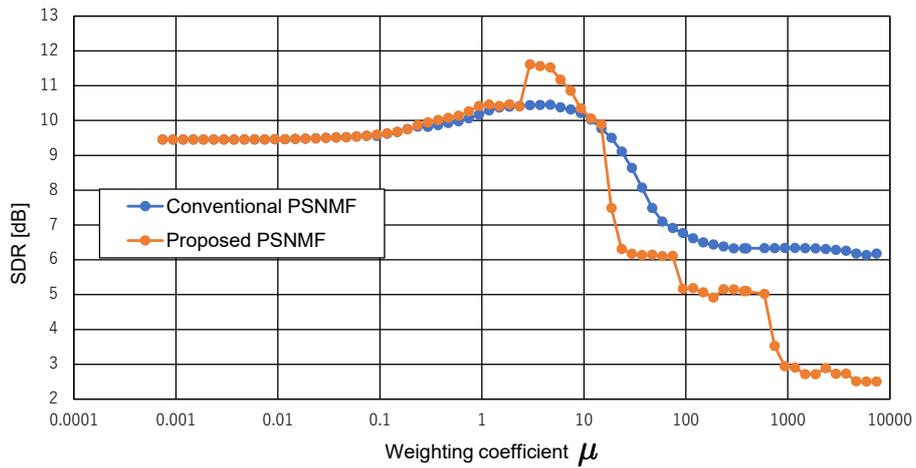


Fig. 4.60: SDR values of conventional and proposed PSNMFs for Pf.&amp;Vn. mixtures.

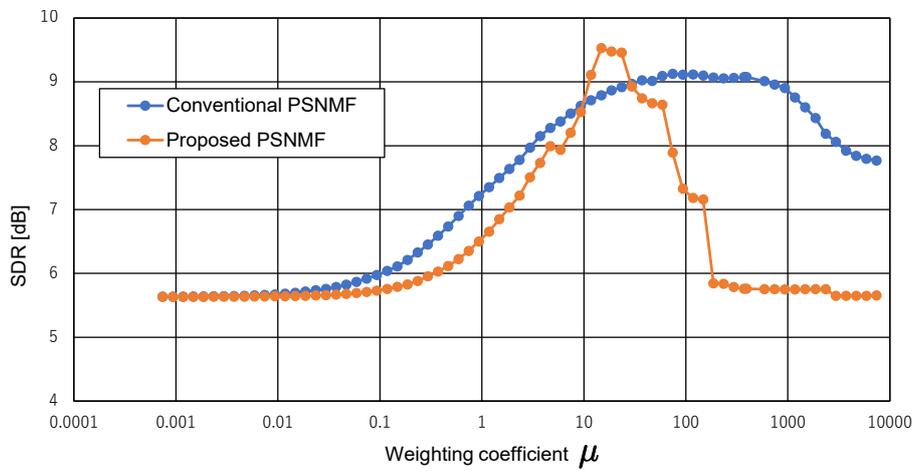


Fig. 4.61: SDR values of conventional and proposed PSNMFs for Tb.&amp;Cl. mixtures.

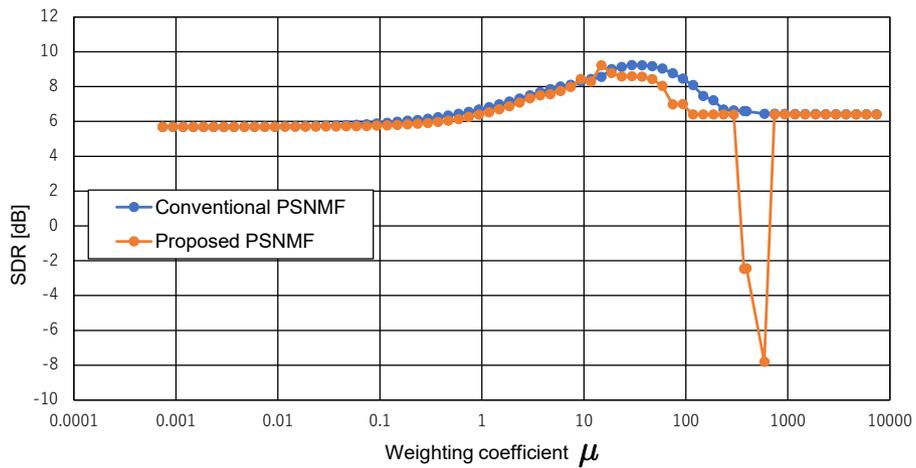


Fig. 4.62: SDR values of conventional and proposed PSNMFs for Tb.&amp;Fl. mixtures.

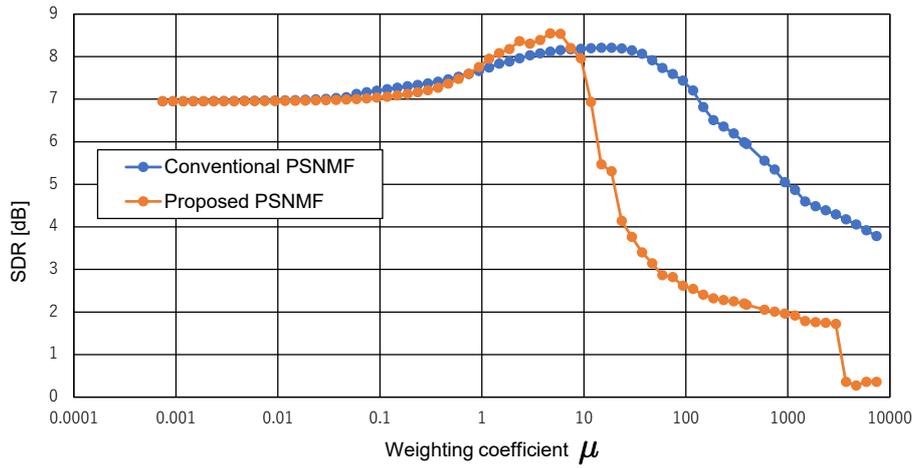


Fig. 4.63: SDR values of conventional and proposed PSNMFs for Tb.&Hp. mixtures.

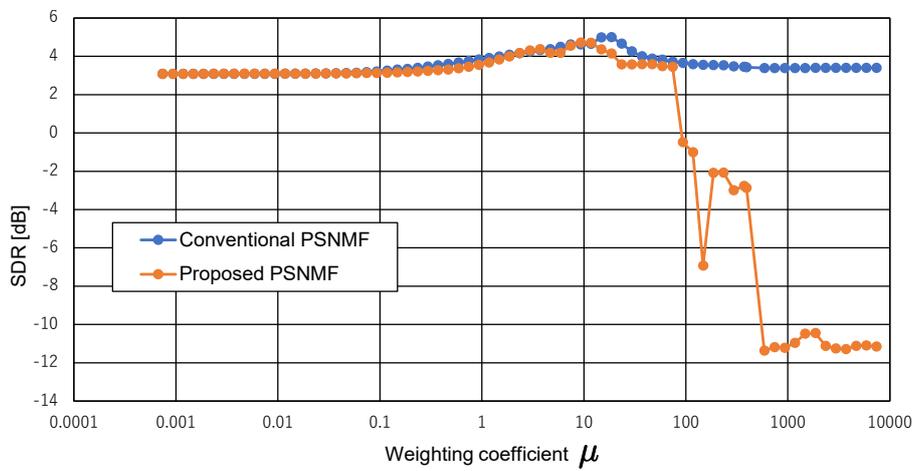


Fig. 4.64: SDR values of conventional and proposed PSNMFs for Tb.&Hr. mixtures.

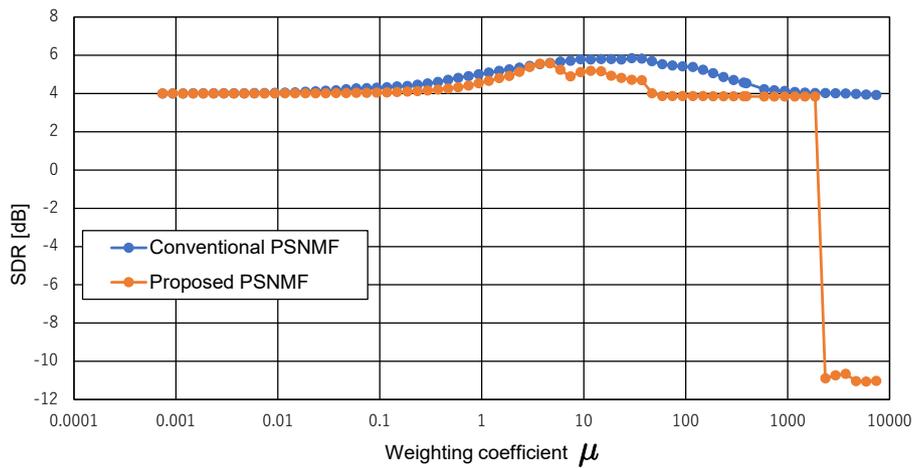


Fig. 4.65: SDR values of conventional and proposed PSNMFs for Tb.&Ob. mixtures.

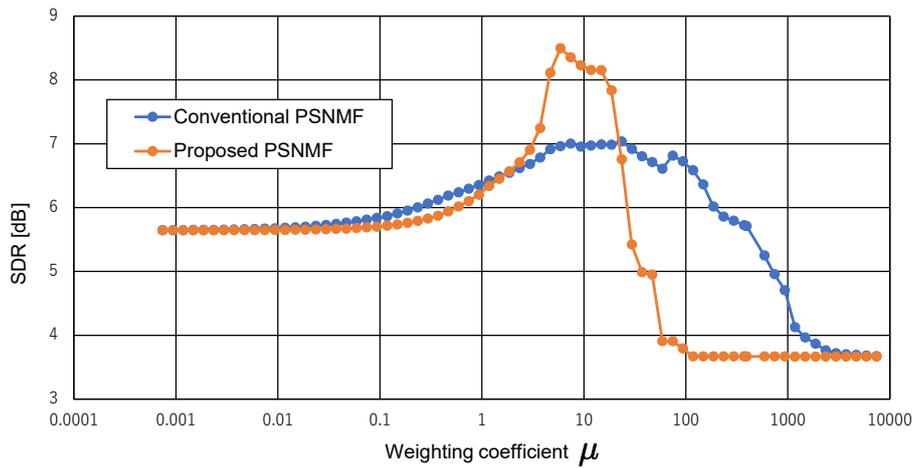


Fig. 4.66: SDR values of conventional and proposed PSNMFs for Tb.&Pf. mixtures.

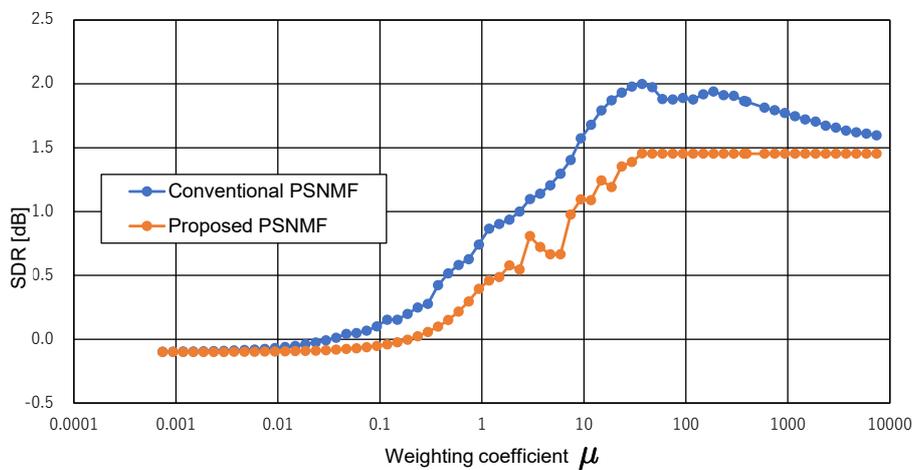


Fig. 4.67: SDR values of conventional and proposed PSNMFs for Tb.&Tp. mixtures.

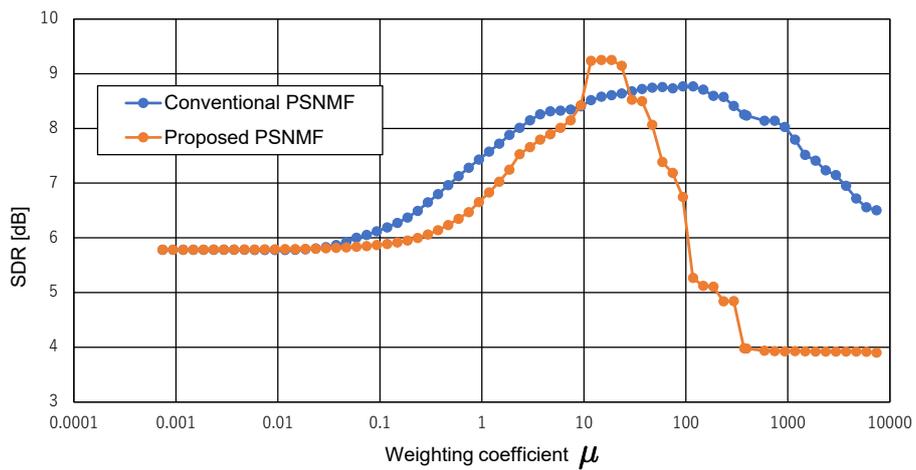


Fig. 4.68: SDR values of conventional and proposed PSNMFs for Tb.&Vn. mixtures.

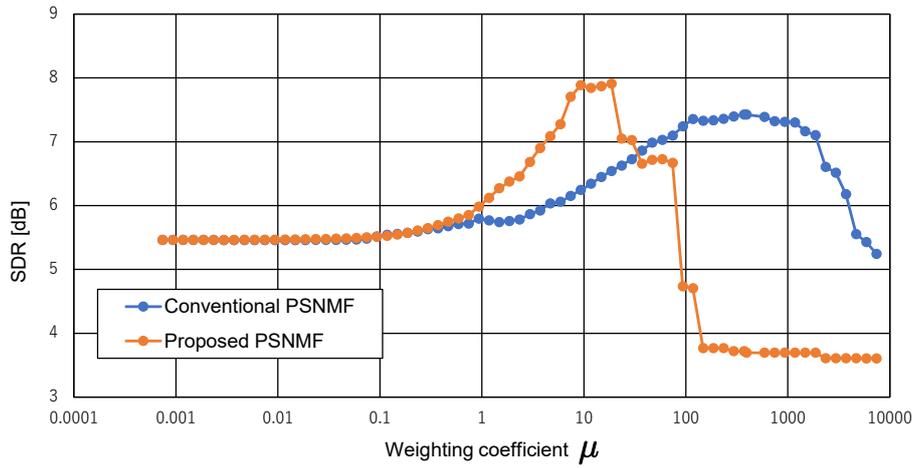


Fig. 4.69: SDR values of conventional and proposed PSNMFs for Tp.&Cl. mixtures.

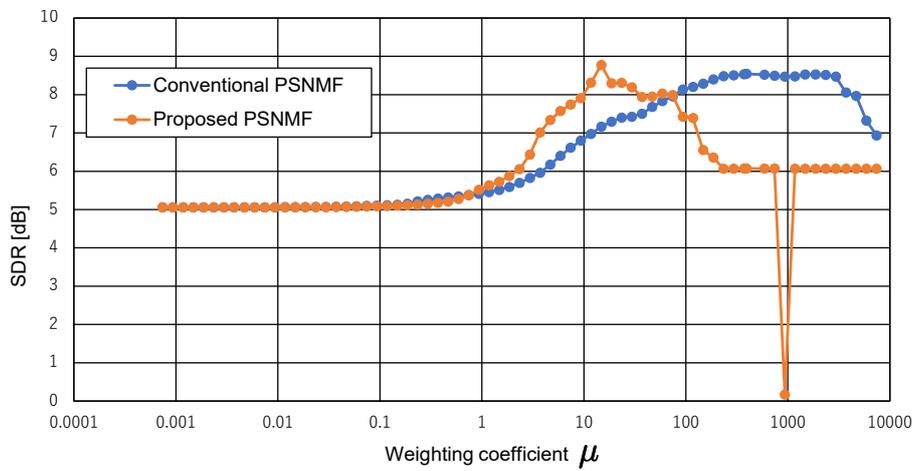


Fig. 4.70: SDR values of conventional and proposed PSNMFs for Tp.&Fg. mixtures.

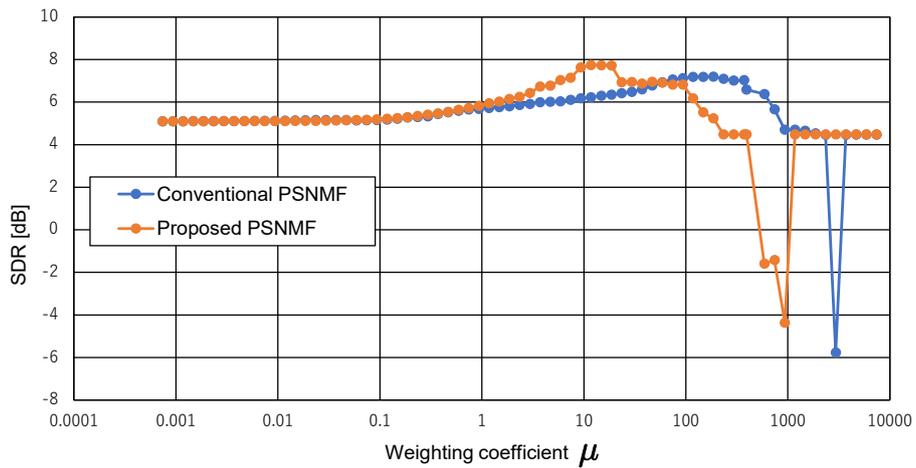


Fig. 4.71: SDR values of conventional and proposed PSNMFs for Tp.&Fl. mixtures.

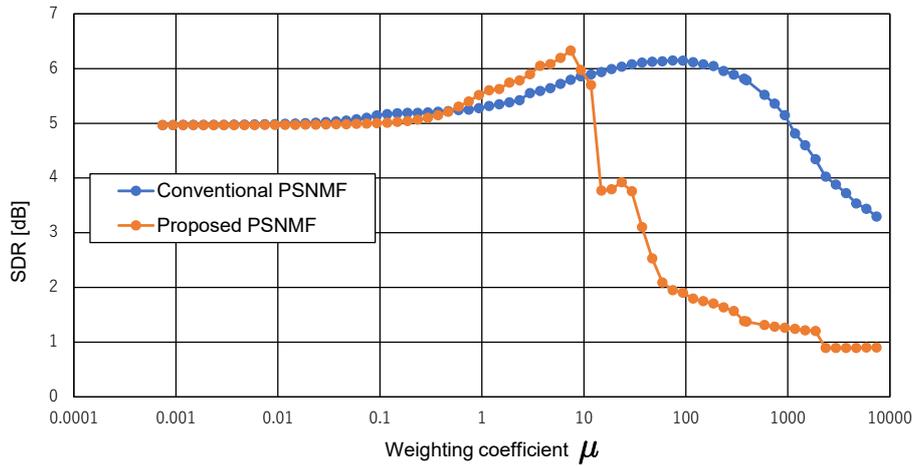


Fig. 4.72: SDR values of conventional and proposed PSNMFs for Tp.&Hp. mixtures.

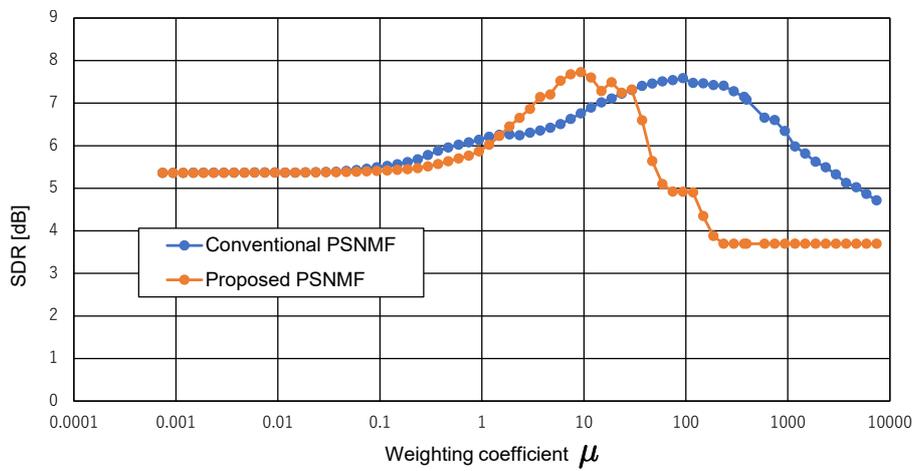


Fig. 4.73: SDR values of conventional and proposed PSNMFs for Tp.&Pf. mixtures.

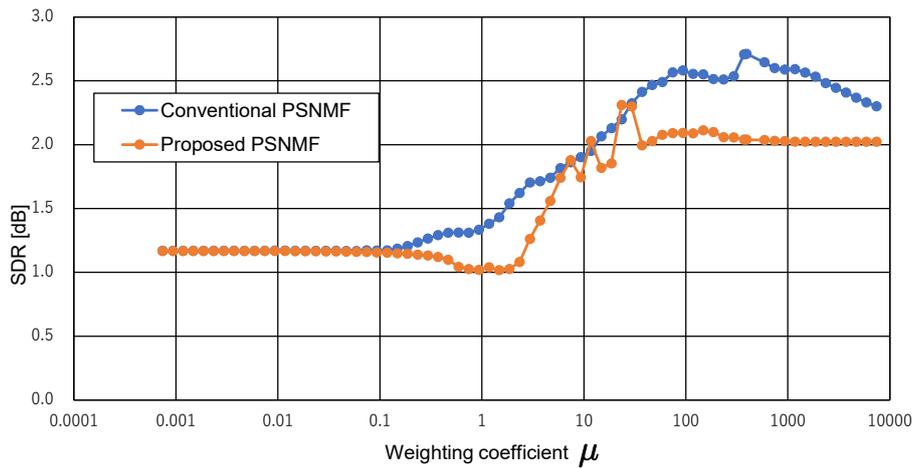


Fig. 4.74: SDR values of conventional and proposed PSNMFs for Tp.&Tb. mixtures.

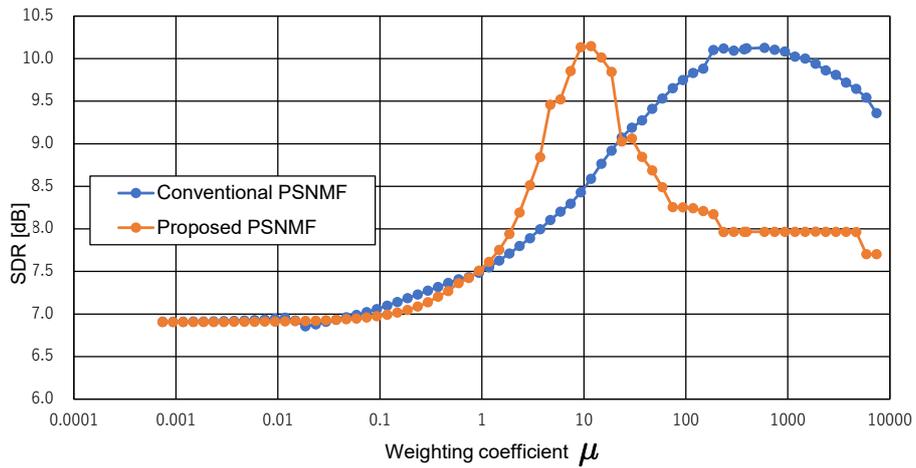


Fig. 4.75: SDR values of conventional and proposed PSNMFs for Tp.&Vc. mixtures.

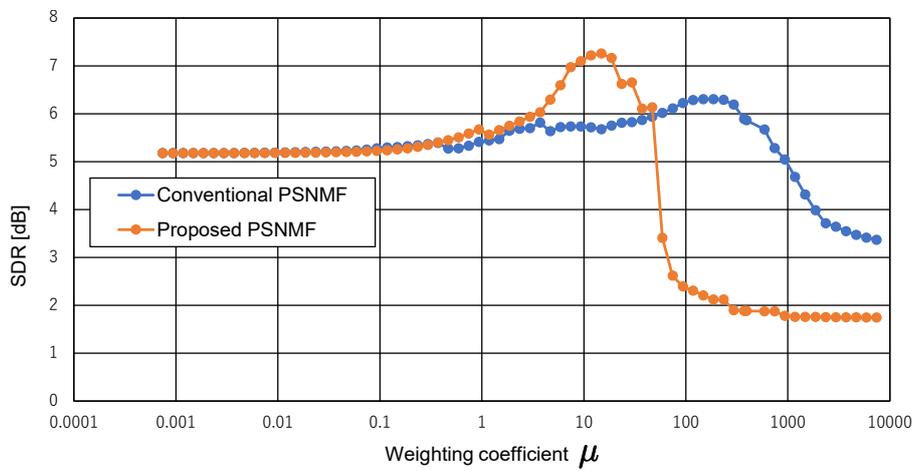


Fig. 4.76: SDR values of conventional and proposed PSNMFs for Tp.&Vn. mixtures.

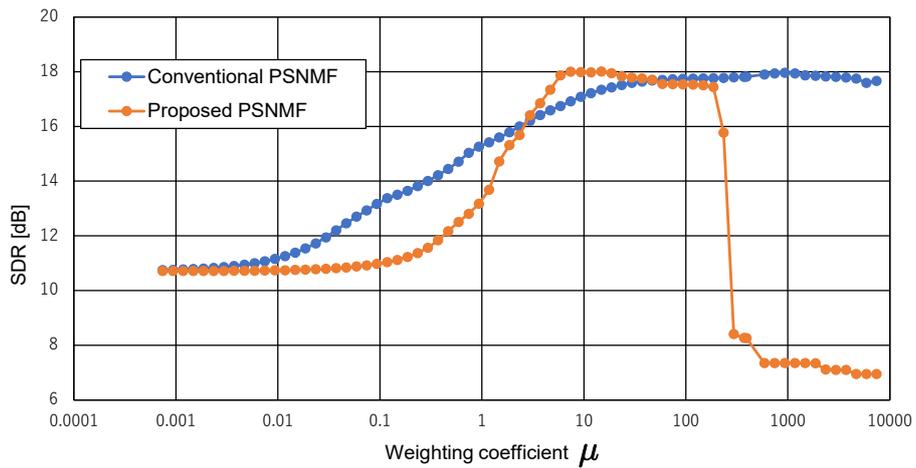


Fig. 4.77: SDR values of conventional and proposed PSNMFs for Vc.&Cl. mixtures.

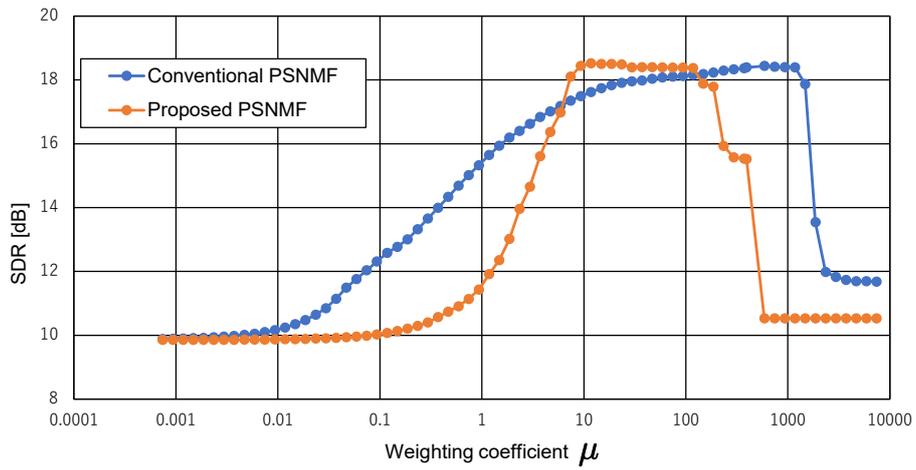


Fig. 4.78: SDR values of conventional and proposed PSNMFs for Vc.&Fl. mixtures.

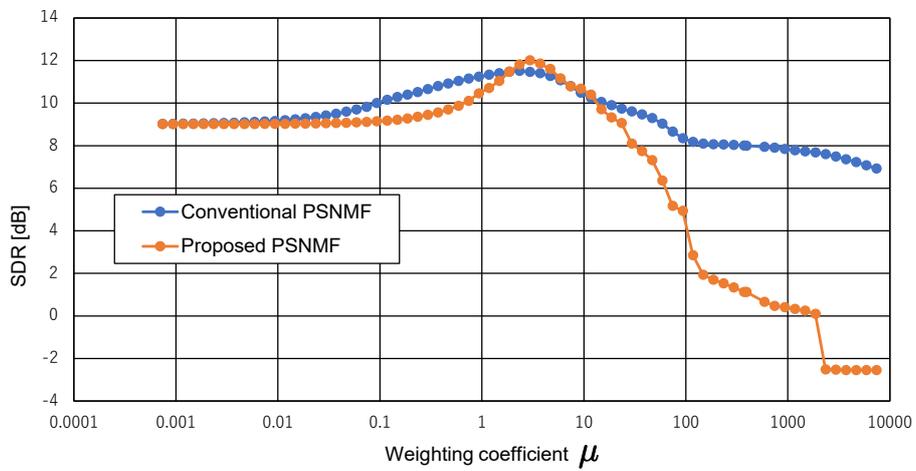


Fig. 4.79: SDR values of conventional and proposed PSNMFs for Vc.&Hp. mixtures.

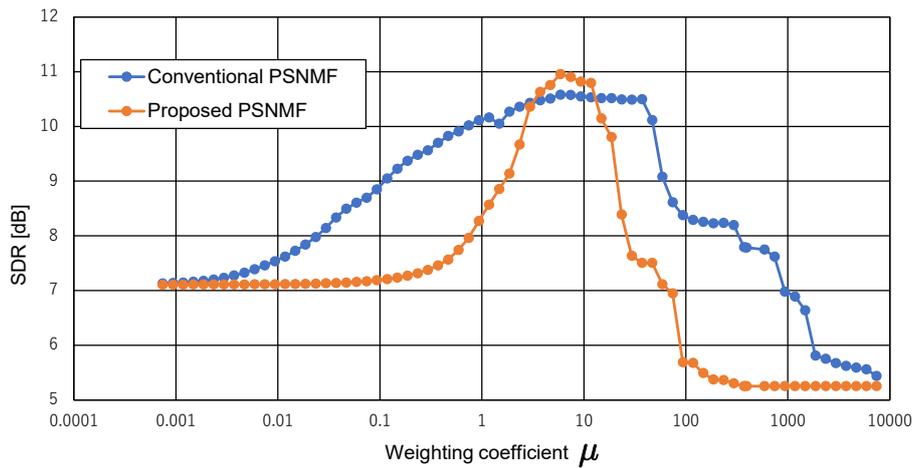


Fig. 4.80: SDR values of conventional and proposed PSNMFs for Vc.&Hr. mixtures.

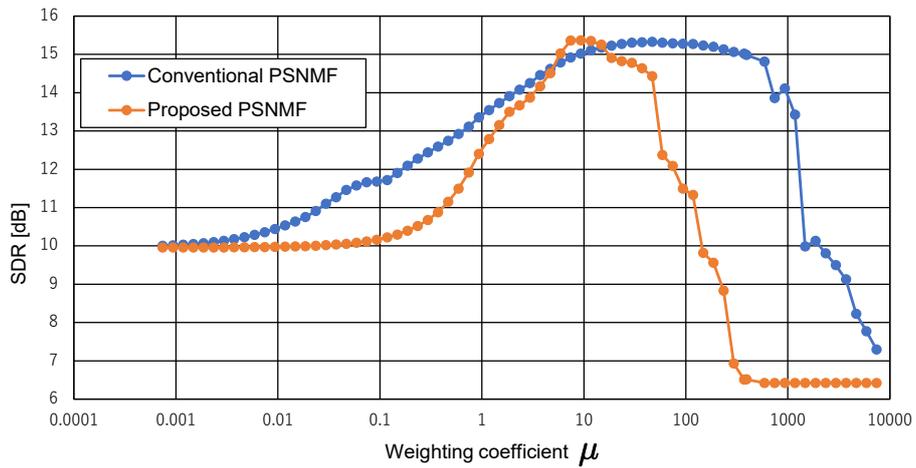


Fig. 4.81: SDR values of conventional and proposed PSNMFs for Vc.&Ob. mixtures.

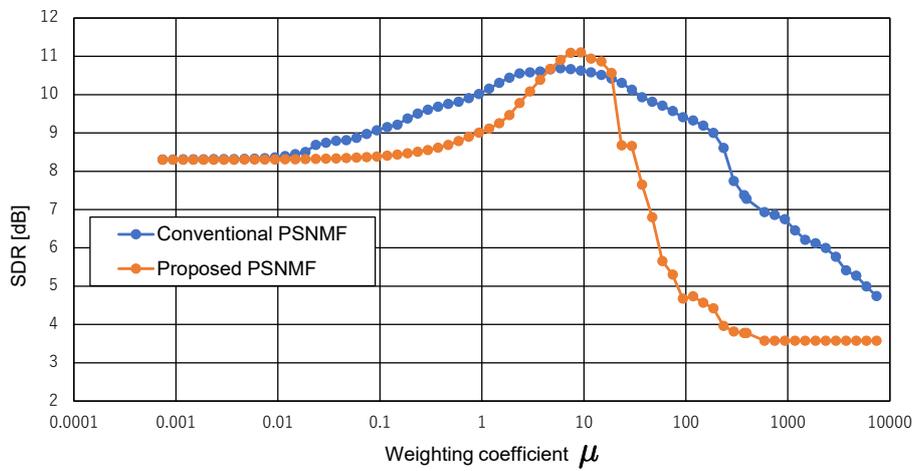


Fig. 4.82: SDR values of conventional and proposed PSNMFs for Vc.&Pf. mixtures.

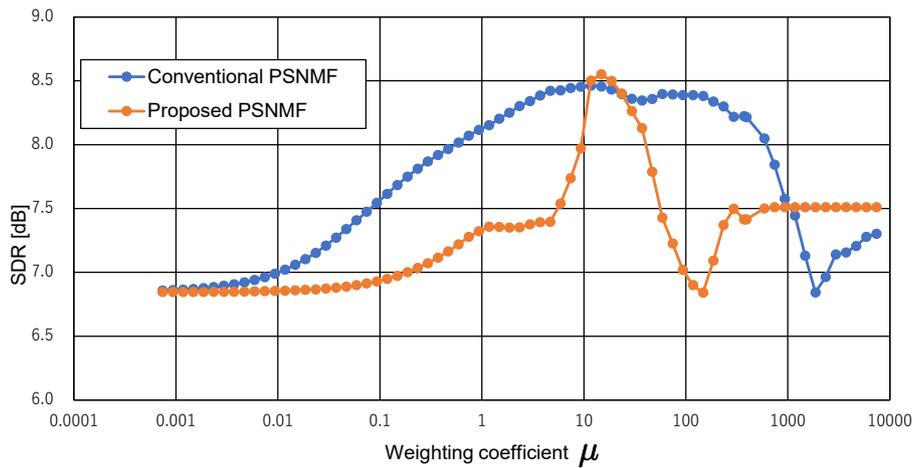


Fig. 4.83: SDR values of conventional and proposed PSNMFs for Vc.&Tb. mixtures.

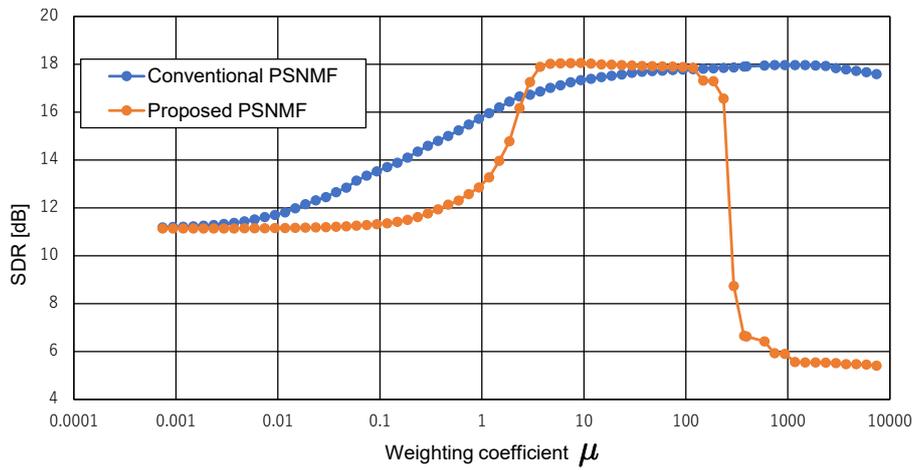


Fig. 4.84: SDR values of conventional and proposed PSNMFs for Vc.&Vn. mixtures.

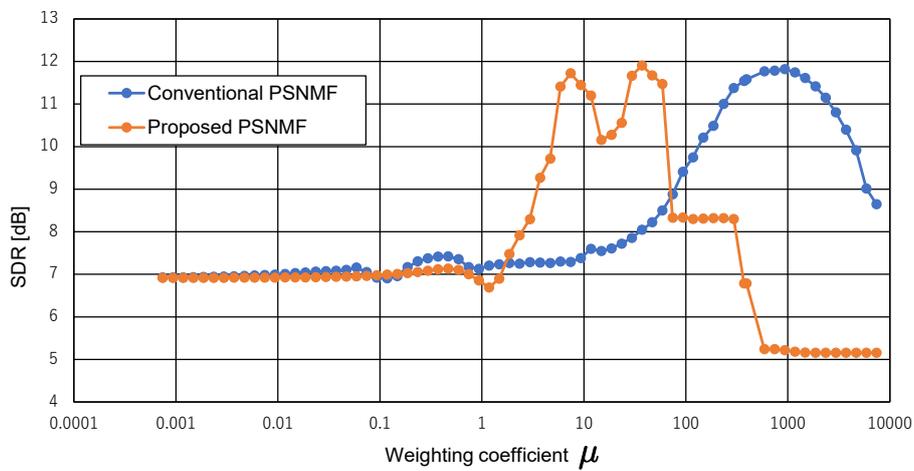


Fig. 4.85: SDR values of conventional and proposed PSNMFs for Vn.&Fg. mixtures.

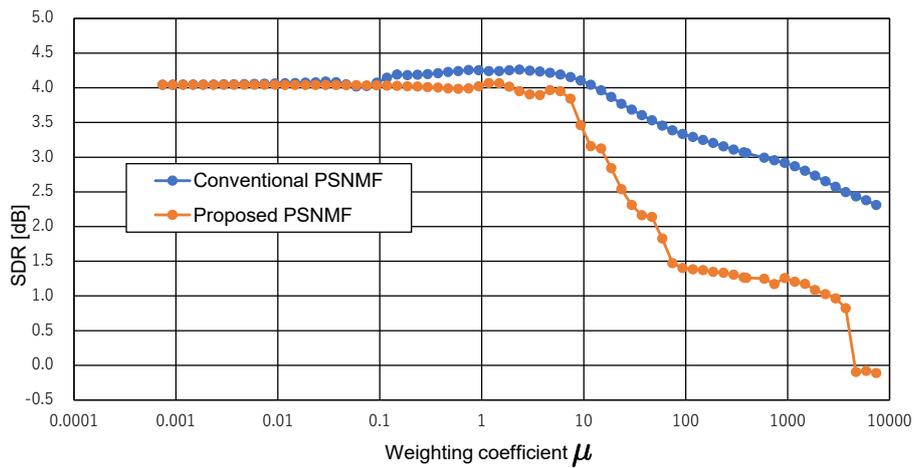


Fig. 4.86: SDR values of conventional and proposed PSNMFs for Vn.&Hp. mixtures.

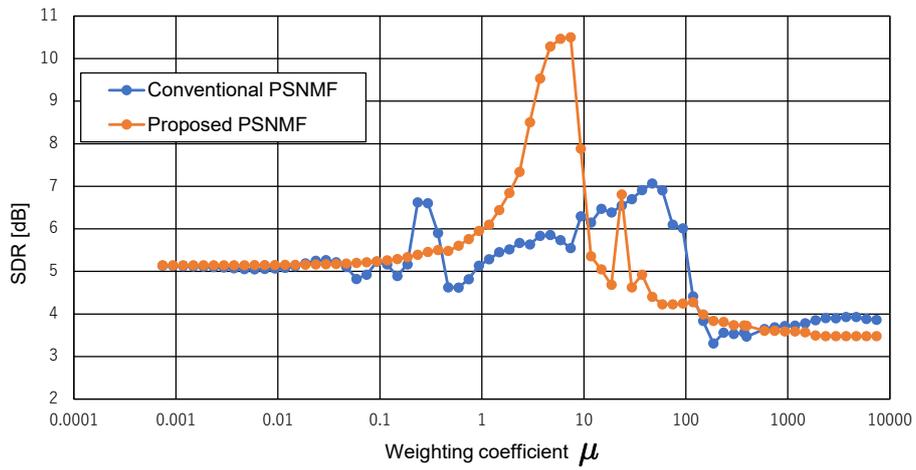


Fig. 4.87: SDR values of conventional and proposed PSNMFs for Vn.&Hr. mixtures.

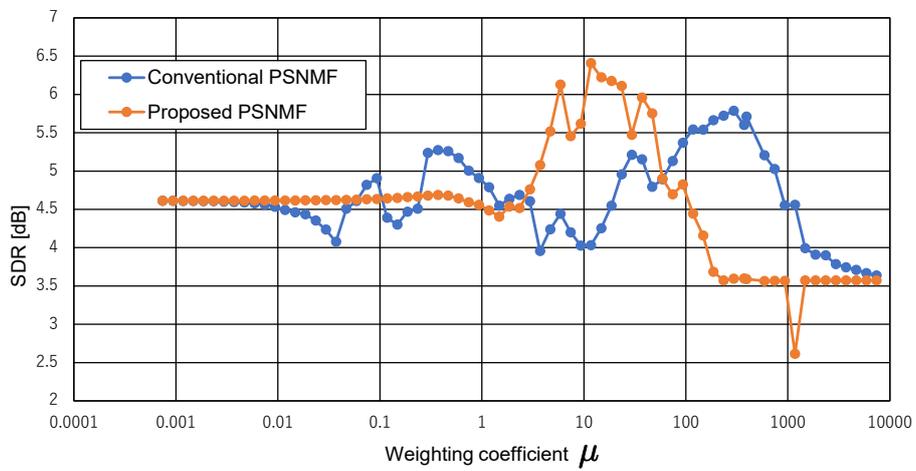


Fig. 4.88: SDR values of conventional and proposed PSNMFs for Vn.&Ob. mixtures.

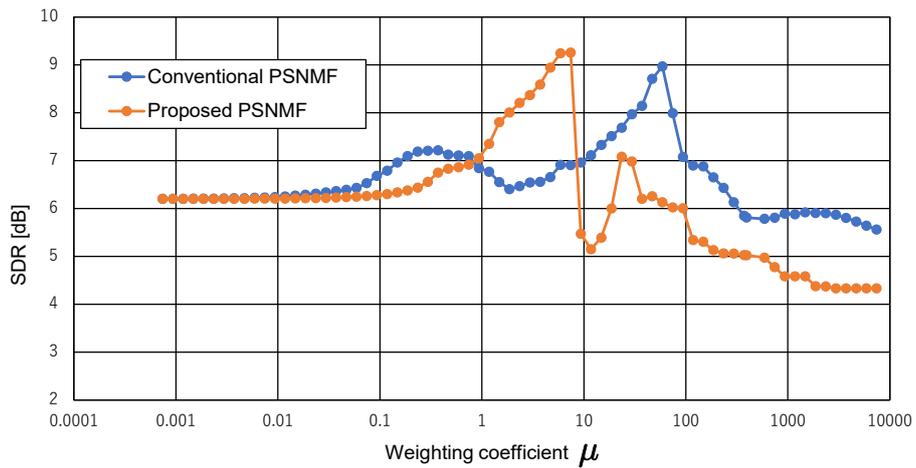


Fig. 4.89: SDR values of conventional and proposed PSNMFs for Vn.&Pf. mixtures.

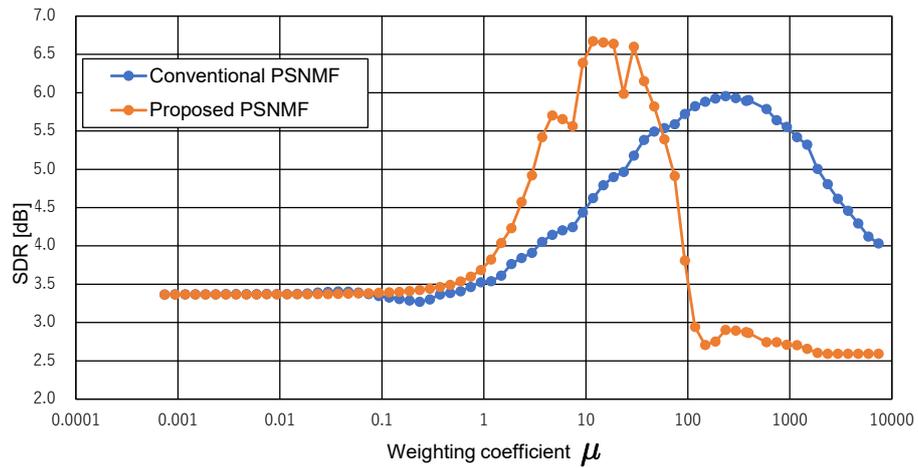


Fig. 4.90: SDR values of conventional and proposed PSNMFs for Vn.&Tb. mixtures.

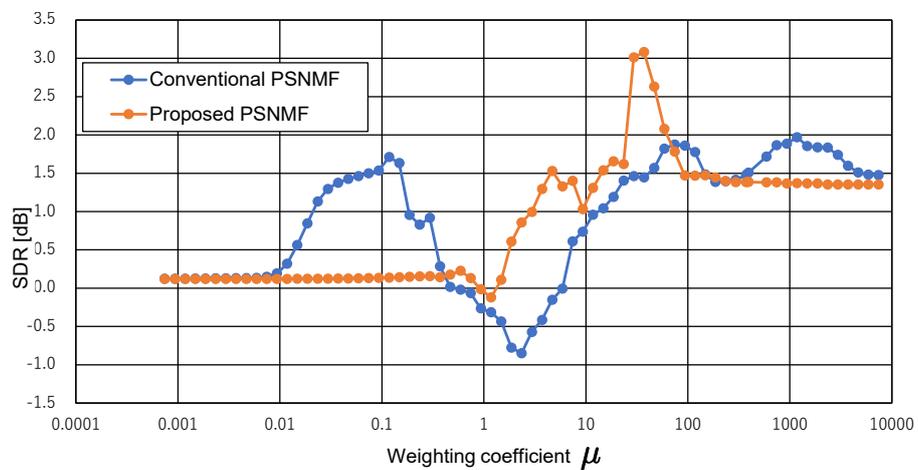


Fig. 4.91: SDR values of conventional and proposed PSNMFs for Vn.&Tp. mixtures.

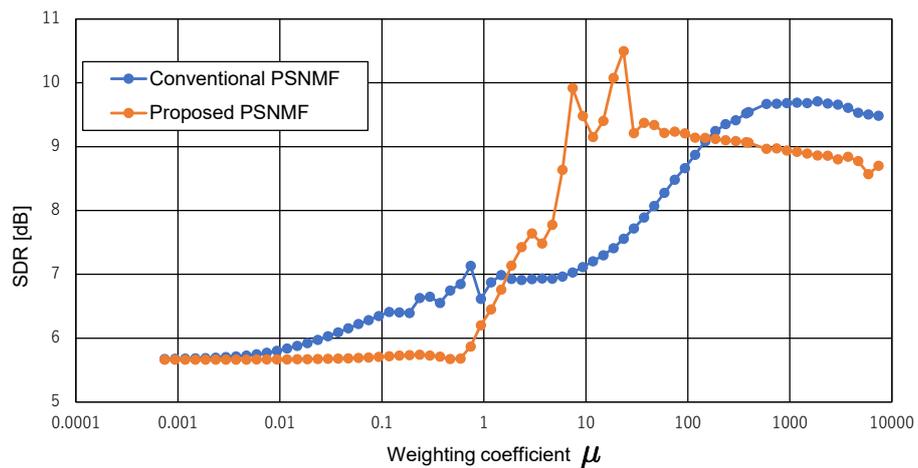


Fig. 4.92: SDR values of conventional and proposed PSNMFs for Vn.&Vc. mixtures.

Table 4.1: Average and median SDRs for each method

Method	Average	Median
SNMF	6.09 dB	5.73 dB
Conventional PSNMF	8.51 dB	7.91 dB
Proposed PSNMF	8.73 dB	8.22 dB

## 第 5 章

# 結言

本稿では, SNMF において, スケールの不定性の影響を受けない罰則項としてコサイン類似度を用いたアルゴリズムを新たに提案し性能調査を行った. 特に, 既存の手法である直交化罰則項を用いた PSNMF と比較した. 実験結果より, 多くの混合信号において, 提案手法が既存手法の分離精度を上回った. これにより SNMF にコサイン類似度をコスト関数として付与する手法は音源分離に有用であることが確認できた. 特に,  $V_c$  及び  $V_n$  を目的音としたとき, 前章で述べた通り多くの実験結果で提案手法が従来手法の分離精度に比べ勝っていた. よって弦楽器のメロディーを混合音から抽出する際は, コサイン類似度を用いたほうが良いといえる. しかし, たとえ弦楽器を目的音をした場合であっても, ピーク値を過ぎた提案手法の SDR 値は従来手法より低い. したがって, 今後提案手法を実用的にするためには, 楽器ごとに SDR 値が最大となる重み係数の傾向を掴む必要がある.

# 謝辞

本論文は、香川高等専門学校電気情報工学科北村研究室にて行われた研究に基づくものです。

まず、本研究を進めるにあたり、ご多忙のところ熱心にご指導くださいました指導教員の北村大地助教に心より感謝申し上げます。北村大地助教には、論文執筆や研究に関する議論など、細部にわたるまで丁寧にご指導いただきました。

本論の副査である太良尾浩生准教授には、論文の構成や記述に関して大変有益な助言を頂き、大変お世話になりました。ここに厚く御礼申し上げます。

また、北村研究室同期の渡辺瑠伊氏・大藪宗一郎氏・中野将生氏には大変お世話になりました。

最後になりますが、現在に至るまで私の学生生活を金銭的に支え、暖かく見守って下さった両親には感謝の念に堪えません。これまで本当にありがとうございました。

## 参考文献

- [1] H. Kameoka, M. Nakano, K. Ochiai, Y. Imoto, K. Kashino, and S. Sagayama, “Constrained and regularized variants of non-negative matrix factorization incorporating music-specific constraints,” *Proc. International Conference on Acoustics, Speech and Signal Processing*, pp. 5365–5368. 2012,
- [2] P. Smaragdis, B. Raj, and M. Shashanka, “Supervised and semi-supervised separation of sounds from single-channel mixtures,” *Proc. Independent Component Analysis and Signal Separation*, pp. 414–421, 2007.
- [3] D. Kitamura, H. Saruwatari, K. Yagi, K. Shikano, Y. Takahashi, and K. Kondo, “Music signal separation based on supervised nonnegative matrix factorization with orthogonality and maximum-divergence penalties,” *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol. E97-A, no. 5, pp. 1113–1118, 2014.
- [4] E. M. Grais, M. U. Sen, and H. Erdogan, “Deep neural networks for signal channel source separation,” *Proc. International Conference on Independent Component Analysis and Signal Separation*, pp. 3734–3738. 2014.
- [5] S. Uhilch, F. Giron, and Y. Mitsufuji, “Deep neural network based instrument extraction from music,” *Proc. International Conference on Independent Component Analysis and Signal Separation*, pp. 2135–2139. 2015.
- [6] D. D. Lee, and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [7] D. D. Lee, and H. S. Seung, “Algorithms for non-negative matrix factorization,” *Proc. Neural Information Processing Systems*, pp. 556–562, 2000.
- [8] T. Virtanen, “Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1066–1074, 2007.
- [9] A. Ozerov, C. Févotte, and M. Charbit, “Factorial scaled hidden Markov model for polyphonic audio representation and source separation,” *Proc. Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 121–124, 2009.
- [10] D. Kitamura, H. Saruwatari, H. Kameoka, Y. Takahashi, K. Kondo, and S. Naka-

- mura, “Multichannel signal separation combining directional clustering and nonnegative matrix factorization with spectrogram restoration,” *IEEE/ACM Trans. Audio, Speech, and Language Processing*, vol. 23, no. 4, pp. 654–669, 2015.
- [11] 蓮本恭介, 後藤正幸, 雲居玄道, “非負値行列因子分解による顧客購買パターン抽出と顧客生涯価値予想,” 第 80 回全国大会講演論文集, pp. 151–152, no. 1, 2018.
- [12] C. Fèvotte, N. Bertin, and J.L. Durrieu, “Nonnegative matrix factorization with Itakura–Saito divergence: With application to music analysis,” *Neural Computation*, vol. 21, no. 3, pp. 793–830, 2009.
- [13] D. R. Hunter and K. Lange, “Quantile regression via an MM algorithm,” *Journal of computational and Graphical Statistics.*, vol. 9, no. 1, pp. 60–77, 2000.
- [14] D. FitzGerald, M. Cranitch, and E. Coyle, “On the use of the beta divergence for musical source separation,” *Proc. Irish Signals and Systems Conference*, 2009.
- [15] J. Le Roux, F. Weninger, and J. R. Hershey, “Sparse NMF - half-baked or well done?,” *Mitsubishi Electric Research Lab. Technical Report*, TR2015-023, 2015.
- [16] D. Kitamura, “Open dataset: songKitamura” <http://d-kitamura.net/dataset.html>.
- [17] E. Vincent, R. Gribonval, and C. Fèvotte, “Performance measurement in blind audio source separation,” *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.

# 発表文献一覧

## 国内学会

1. 岩瀬佑太, 北村大地, “コサイン類似度罰則条件付き半教師あり非負値行列因子分解,” 日本音響学会 2020 年春季研究発表会講演論文集, 2-P-39, 2020 (to appear).

## 付録 A

# 補助関数法で利用される不等式

### A.1 Jensen の不等式

**補題 A.1.** (Jensen の不等式)  $\alpha_i > 0$  を,  $\sum_i \alpha_i = 1$  を満たす補助変数とする. 関数  $f(x)$  が凸関数であるとき,  $x_i$  ( $i = 1, \dots, I$ ) に対して以下の不等式が成立する.

$$f\left(\sum_{i=1}^I \alpha_i x_i\right) \leq \sum_{i=1}^I \alpha_i f(x_i) \quad (\text{A.1})$$

$f(x)$  が狭義凸関数であるとき, 不等式中の等号が成立するための条件は  $x_1 = \dots = x_i = \dots = x_I$  である.

### A.2 接線不等式

**補題 A.2.** (接線不等式)  $f(x)$  が凹関数であるとき, 以下の不等式が成立する.

$$f(x) \leq f'(\bar{x})(x - \bar{x}) + f(\bar{x}) \quad (\text{A.2})$$

不等式中の等号が成立するための条件は  $x = \bar{x}$  である.